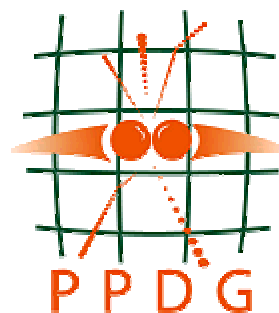


**Particle Physics Data Grid:
From Fabric to Physics**

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

July-September 2004

31 October 2004



1 Project Overview	2	4.7.4 Infrastructure and other activities	14
1.1 Highlights	2	4.8 PHENIX	14
1.2 Papers and Documents.....	2	4.9 Condor.....	15
2 Summary of Work in Focus Areas and the Common Project.....	2	4.10 Globus – ANL	15
2.1 Data Management.....	3	4.10.1 Coordination and Support.....	15
2.2 Job Management.....	3	4.10.2 Globus Toolkit 2.x Updates, Bug Fixes and Open Issues	15
2.3 Production Grids.....	3	4.10.3 Globus Toolkit 3.2.....	15
2.4 Data Analysis Working Group	3	4.10.4 GridFTP / XIO.....	15
2.5 Monitoring.....	4	4.10.5 Monitoring and MDS work	15
2.6 AAA	4	4.11 SRM.....	16
2.7 PPDG Common Project.....	4	4.12 SRB	17
3 Collaborations	5	4.13 IEPM, Network Performance Monitoring	18
3.1 Trillium and Grid3.....	5	4.13.1 Bandwidth/Throughput measurement (IEPM-BW).....	18
3.2 Global Grid Forum and PNPA Research Group.....	5	4.13.2 Lightweight Bandwidth Estimation .	18
3.3 Joint Technical Board (JTB).....	5	4.13.3 Bandwidth performance anomalous events.....	18
3.4 Open Science Grid.....	5	4.13.4 Traceroute Analysis and Visualization	19
4 Single Team Reports	5	4.13.5 PingER.....	19
4.1 ATLAS	5	4.13.6 SC2004 Bandwidth Challenge.....	19
4.2 BaBar.....	6	4.13.7 Proposals and Representation	19
4.3 CMS	6	5 Appendix	20
4.4 D0.....	10	5.1 List of participants.....	20
4.5 JLab	10	5.2 Meetings	22
4.6 BNL RCF/ACF.....	10	5.3 Related Publications	23
4.7 STAR.....	11		
4.7.1 Monitoring.....	11		
4.7.2 Job Management.....	11		
4.7.3 Data Management.....	13		

1 Project Overview

1.1 Highlights

The ATLAS experiment began its production running on the grid for ATLAS Data Challenge 2, as shown in the figure below. The Grid3 environment in the U.S. is providing about 30% of the global ATLAS resources for DC2. An important aspect of the PPDG effort is utilization of the Chimera/Pegasus software from GriPhyN for job planning.

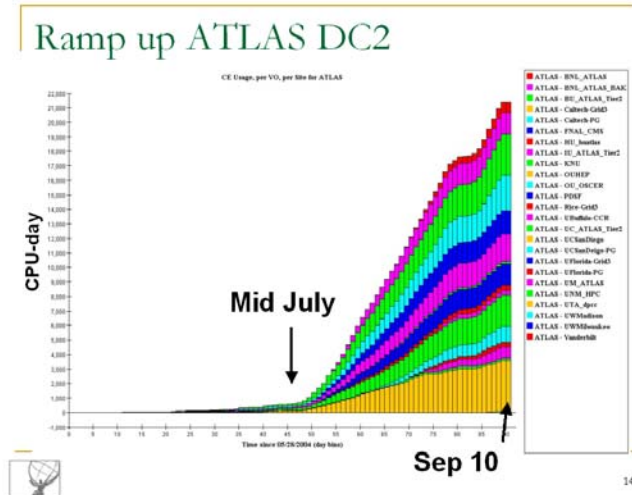
BaBar is using the “zones” feature of SRB, in one case to provide metadata and snapshots from the conditions database (used in data analysis), and also testing two zones with a view towards federating the SRB MCAT installations at SLAC and IN2P3 in the future.

CMS finished its main DC04 production in the previous quarter and the CMS facilities participating in Grid3 are now being utilized by ATLAS for its production computing. One important aspect of the Caltech effort is the Clarens Rendezvous service, which is now a candidate for the main Discovery Service for Open Science Grid as well as being discussed within the PPDG analysis group (CS-11) as one of the interfaces necessary to standardize for interactive data analysis.

BNL has set up the VOMRS service for managing VO membership for STAR and PHENIX as well as deploying SRM as a public interface to the mass storage system for RHIC and ATLAS.

STAR has been developing grid-level scheduling algorithms based on monitoring information from different sites & clusters. This is being integrated with SUMS to provide for grid-based data analysis.

The SRM group at LBNL hosted a face to face meeting of the SRM working group where details of SRM V2.1.1 were discussed along with planning for SRM V3. Adjacent to this meeting was a focus meeting on the Replica Registration Service in the context of an overall replica management service. There is also effort going into the support of testing the DRM deployed via VDT.



1.2 Papers and Documents

Members of the teams participating in PPDG made 55 presentations at CHEP'04. Links to these presentations are posted at <http://www.ppdg.net/docs/chep04-presentations.html>.

2 Summary of Work in Focus Areas and the Common Project

In addition to the work described below many of the people participating in PPDG attended the CHEP conference and made presentations (see above) covering nearly all aspects of work in PPDG.

2.1 Data Management

SRM related work this quarter included deployment of Berkeley_SRM in multiple sites, coordinating and running two collaboration meetings for Replica Registration Service (RRS) and Storage Resource management (SRM), and the writing of draft functional specification for the RRS 1.0 and the SRM v3.0. The RRS fits into a broader Replica Management Service (RMS) that also includes a Replica Copy Service (RCS) and a Replica Selection Service (RSS). These activities are described in detail in section 4.11.

Work on the XIO component of the new GridFTP server continued. A MODE E driver enables Open/Close/Read/Write (OCRW) access to the GridFTP data channel. There is now also a queuing driver to allow multiple outstanding writes. See section 4.10.4 for more detailed description of this work.

In addition to the SRB-related work for BaBar mentioned above, there has been significant progress in the integration of a GridFTP driver into the SRB and an SRB interface into GridFTP using Globus Toolkit V3 (see section 4.12).

JLab has implemented an SRM V2.1.1 interface to the Jasmine storage system and client tools, for use at JLab and at remote sites for collaborators to transfer data to/from JLab. An installation of grid middleware and SRM client tools at CMU is being tested (see section 4.5).

2.2 Job Management.

Work continued on enhancing the STAR Unified Meta-Scheduler (SUMS) as a front end for users submitting data analysis and simulation work for execution on the grid (described in section 4.7). In addition, several job description language specifications are being compared in collaboration with Tech-X Corp which should lead to an updated specification for a U-JDL (section 4.7). PHENIX is using SUMS and participating with STAR in this work.

JLab is evolving their job submission tool (Auger) to support grid access via GSI and removing the LSF dependencies (See section 4.5).

The Condor team has worked on developing Condor-C as part of the overall strategy to run VO-specific services at sites and maintaining a VO-specific queue of work at each site. This is planned for inclusion in VDT 1.3 (see <http://www.opensciencegrid.org/events/meetings/boston0904/docs/vdt-roy.ppt>).

2.3 Production Grids

The Grid3 environment¹ is currently being used as part of the ATLAS Data Challenge 2 production, providing about 30% of the global resource usage by ATLAS.

D0 is expanding the number of sites running JIM for the Monte Carlo production in addition to supporting the ongoing Monte Carlo running. JIM is also being deployed at the D0 reconstruction farm at Fermilab that will enable a fully unified grid job submission to be used for the full D0 reprocessing project, scheduled to begin in January 2005.

The BNL RACF has deployed SRM as a production grid interface to the storage systems for RHIC and ATLAS.

2.4 Data Analysis Working Group

In addition to the analysis projects of the experiments, the main activities of the CS-11 group have been interacting with European activities & documents from ARDA and EGEE.gLite, discussing and understanding the Clarens Rendezvous Service as a discovery service, and individual preparation of various demos for Supercomputing.

The Clarens Rendezvous service has been proposed as a top-level discovery service for the Open Science Grid. An example service open for others to use is at <http://discover.gridservice.info/>.

¹ www.ivdgl.org/grid3

The ATLAS Distributed Analysis system (ADA) has been interfaced to the various job submission services used by ATLAS worldwide (LCG, Nordugrid, Grid3) as well as the DIAL (Distributed Interactive Analysis of Large datasets) service running at BNL. As the ATLAS participation in ARDA it has been interfaced with gLite and will be available for trial usage following the gLite release in December.

The Grid Analysis Environment (GAE) development and deployment for USCMS now includes a dozen types of services and several more in progress, and includes services deployed at 20 sites worldwide. This is described in a detailed CHEP presentation.²

2.5 Monitoring

MonALISA has been extended in functionality to include; discovery of network topology, an API for applications to report user-defined information, and fabric monitoring. In addition there have been improvements to the repository and GUI (see section 4.3).

There is work at FNAL to interface between the SAM Monitoring and Information Server with MonALISA.

There is work at BNL developing custom MonALISA modules to collect information from batch systems to help with grid-level scheduling decisions. This work is coordinated with the SUMS effort at BNL (see section 4.7).

Effort by Globus has ported the MDS from OGSi to WSRF in addition to adding a trigger service, an archiver and a web user interface (see section 4.10).

The network performance monitoring cooperative effort of the IEPM project at SLAC is described in section 4.13.

2.6 AAA

Some members of PPDG are participating in the Open Science Grid Security Technical Group. The major focus of this effort was to develop a draft Incident Response document³. This effort is in close cooperation with the EGEE and LCG security group work.

It has been a long standing desire of many users of the DOEGrids CA to have a command-line interface to requesting and retrieving X509 certificates. A combination of efforts from ESnet, NERSC, Fusion Grid and PPDG has resulted in a set of shell scripts that satisfy this goal. They will be available soon in a contributed area of the DOEGrids web site.

2.7 PPDG Common Project

Starting this quarter (July 2004) PPDG has organized its effort as described in the January 2004 SciDAC proposal⁴ with all teams contributing to the PPDG-OSG Common Project with the major goal of connecting resources at 5 DOE labs to a shared grid infrastructure for use by all the PPDG applications. This effort is also carried out as a contribution to the Open Science Grid (OSG). This section summarizes many of the activities within this new common project.

Several members of PPDG are participating in the OSG Blueprint activity, which is defining guiding principles and technical roadmap for OSG, as well as contributing to defining the capabilities to include in OSG version 0 for Spring 2005.

The most significant addition to the Grid3 environment in the transition to OSG is storage as a fully functional grid service. The SRM interface has been chosen to be the storage interface for OSG (and PPDG). The LBNL SDM group is supporting their implementation of SRM for distribution in VDT and use by OSG sites. Fermilab, BNL and JLab have deployed SRM for their storage systems. STAR has been

² <http://indico.cern.ch/contributionDisplay.py?contribId=182&sessionId=9&confId=0>

³ http://computing.fnal.gov/cgi-bin/docdb/osg_public/ShowDocument?docid=19&version=1

⁴ http://www.ppdg.net/docs/Proposals/scidac04_ppdg-final.pdf

using a VO-specific SRM at NERSC and BNL and NERSC is working on integrating SRM with the HPSS system and expect to deploy it as an interface for all NERSC users in the near future.

Both Fermilab and BNL have participated in data transfer tests between CERN and the Tier-1 sites for USATLAS and USCMS, as part of the overall LCG program. This involved transferring TBs of data from the Data Challenges at rates in the 45-70 MB/sec range.

Resource accounting is one of the PPDG goals and work to define accounting details for storage systems (space and I/O) was started this quarter. Part of this included discussion of accounting information at the SRM collaboration meeting at Berkeley.

The Clarens Rendezvous service is being contributed as the top level discovery service for OSG-V0. Discussions comparing the Clarens and Globus MDS interfaces (WSDL) are in progress.

3 Collaborations

3.1 Trillium and Grid3

PPDG continues to collaborate closely with iVDGL, GriPhyN and the various partners of Grid3. While the Open Science Grid governance is forming the Trillium (PPDG, GriPhyN, iVDGL) Steering Committees are serving as the interim decision making body for OSG.

3.2 Global Grid Forum and PNPA Research Group

A workshop that had been planned at GGF12 with joint sponsorship from several GGF groups including the PNPA research group was cancelled due to conflicts with the PGM research group chairs who were the primary hosts for the workshop. The CHEP conference the week after GGF12 served as the primary forum for discussing grid issues for the high energy and nuclear physics community with about 65 presentations in two parallel tracks on distributed computing issues and experiences.

3.3 Joint Technical Board (JTB)

No meetings this quarter.

3.4 Open Science Grid

As mentioned above in the Common Project subsection, PPDG has a strong commitment to participation in the Open Science Grid Consortium⁵ and having the computing facilities at the DOE labs participating in PPDG to also be part of the OSG.

Many members of PPDG are participating in the OSG Blueprint activity, which is holding monthly face-to-face meetings of several days each. The PPDG common project is making contributions to the OSG blueprint. Other OSG groups and activities to which PPDG is contributing include Security, Storage, Deployment, and Governance.

A general workshop for Open Science Grid was held at Harvard in September where much progress was made working out organizational details of OSG as well as attracting participation from communities beyond Trillium.

4 Single Team Reports

4.1 ATLAS

Effort report from Jerry Gieraltowski of ATLAS on Production

⁵ www.opensciencegrid.org

Description:

My efforts during this interval were devoted solely towards the successful operation of ATLAS Data Challenge 2. My particular area of focus was the application of the Chimera/Pegasus software as deployed in the VDT to create ATLAS client software which could be used to run DC2 applications. The most difficult tasks involved in this effort were the analysis and resolution of site problems which contributed to job failures, and the analysis and resolution of production problems directly linked to high usage of resources in distributed systems. I constructed, tested, and analyzed a number of distinct enhancements to the client software used in DC2 which were intended to resolve resource usage problems.

4.2 BaBar

Effort report from Wilko Kroeger of BaBar on Data Management

Description:

My main effort was to setup an SRB zone that provides the BaBar conditions snapshots. Besides adding the files to the SRB, meta data about the snapshots was added. This meta data is used to find and select snapshots. Tools were created to load snapshots into the SRB and to download a snapshot of interest.

A new SRB version (3.2.1) has been installed and tested for two test SRB zones and the conditions SRB zone. A strange problem that caused intermittent data transfer failures was discovered. The problem was found but not completely understood but a workaround was devised. The SRB for the BaBar root files has not yet been upgraded, but this will be done shortly.

Some effort was spent to support the data distribution of BaBar root-files from SLAC to IN2P3 using the SRB. We are transferring constantly root files from SLAC to IN2P3. The transfer works very smoothly, but little work is still needed to register root-files in SRB, and fix little problems. I spent some time to automate the procedures.

4.3 CMS

Effort report from Michael Thomas of CMS on Analysis

Description:*Rendezvous Service*

I implemented a dynamic rendezvous service for Clarens based on MonALISA. Clarens now publishes service descriptions to MonALISA via the UDP-based ApMon interface. The ML GUI can be used to browse the list of Clarens servers and services. New services appear within 1-2 seconds from the time that a new Clarens server starts up. I also added methods to the rendezvous service that would allow it to query service information from MonALISA. This query can be done by either proxying SOAP requests to a MonALISA server (slightly slow) or by having JClarens start up as a client on the Jini network, just like any MonALISA client. This enables Clarens clients to obtain the very latest up-to-date service information that is available.

This was a key step in the evolution of the GAE as it finally allows us to have a P2P based system for service publication and discovery.

I Presented this rendezvous implementation at a CS11 meeting and sent around some sample code showing how to use Java to access this public service.

JClarens

I presented a paper on JClarens and the GAE at ICWS in San Diego at the start of July. The paper was well received, and many people were interested in our merging of P2P and Web Services in a dynamic Service Oriented Architecture.

While at ICWS I learned about many of the latest Web Service research areas such as dynamic composite services and semantic web services.

Another version of JClarens was released in September. This long awaited release has many installer improvements, bug fixes, and a few new features. SOAP access is now officially supported and tested thanks to a lot of hard work by NUST student Tahir Azim. The dynamic rendezvous service is now the default. This new version can be downloaded from http://sourceforge.net/project/showfiles.php?group_id=53073&package_id=114981

Clarens + BOSS

I was finally able to successfully run a sample ORCA application remotely from a Clarens web service. This is a big step in the GAE as it allows us to run a popular CMS analysis program in a grid environment.

Worked with Jordan Carlson (Caltech undergraduate) on improvements to the BOSS/Clarens integration work. In particular, Jordan had some interesting ideas on how to get Clarens and BOSS to work well with a Condor pool that does not have a shared filesystem. Claudio Grande, the maintainer of BOSS, agreed to accept Jordan's changes into future BOSS releases. We plan to use the results of Jordan's work as part of the upcoming SC2004 demonstration in November.

Met with PPDG team at CHEP in Interlaken. Fun was had by all even though nobody wanted to try out the Alpine Horn.

Effort report from Suresh Man Singh of CMS on Production

Description:

In this quarter, I was basically engaged in properly maintaining Caltech Grid3 production clusters (Caltech-PG and Caltech-Grid3) for the smooth operation of Atlas Data Challenge (DC2) productions. Caltech-PG has become one of the main CMS Grid3 sites to run large number of Atlas jobs, whereas Caltech-Grid3 is continuously being used for Grid Exerciser jobs. Both clusters were upgraded to linux 2.4.26 kernel and openssh-3.8p1-1. Later on they had been upgraded to Grid3 V2.1 (Grid3 software cache). New DOEGrids host certificates have been installed on both clusters. As part of the new Grid3 cache, condor version 6.6.5 has been configured as batch scheduler by providing jobmanger-condor Globus GRAM bridge. Recently, EDG gridmap file making VOMS software have been upgraded on both clusters. Cleaning of runtime disk scratch area and deleting stale jobs from queue are being routinely done.

Due to hard drive failure, Caltech-DGT cluster had been rebuilt from scratch. Currently, it's taken offline for forensic analysis of a security incidence.

Also, a six nodes test cluster has been built using ROCKS 3.2.0 to see how well cluster perform with Red Hat Enterprise Linux. Very soon this cluster will be built with to be released ROCKS 3.3.0. Various Grid and CMS related software will be tested on it.

Effort report from Iosif Legrand of CMS on Instrumentation & Monitoring

Description:

I continue to develop the MonALISA system (<http://monalisa.caltech.edu>) and to provide support for the users.

There are a set of new modules available in the MonALISA framework:

- 1) We developed a network topology discovery service implemented as a set of monalisa agents. This service can be used by any community and provides a dynamic graph for the WAN connectivity and its performance. The system is using a tracepath agent for each site to all the others and combines the information to generate the global topology graph. Dedicated services were developed to identify when the interfaces (IPs) which are on the same router. The agents collect the connectivity tree and the delay on each segment. This information is used to analyze the connectivity and to spot when asymmetric routing is found or when a better connection is possible.

We developed a specialized GUI interface to present the global topology. It allows viewing the routers, networks and Autonomous Systems (AS). The GUI allows selecting any domain and provides several layouts for the connectivity graph

- 2) Support for monitoring applications. This new package named “ApMon” provides APIs in C, C++, Java, Perl, Python which are easy to use inside any application and report user defined monitoring information into MonALISA services. ApMon provides dynamic configuration and it can report to multiple destinations. These APIs are currently used in the CMS-ORCA, ARDA project at CERN and Clarens project. It is also used by a NetFlow module which is used to collect data from a CISCO net Flow unit and to report the traffic information into the monitoring service
- 3) We developed a monitoring module to collect information from the LEMON system which is used at CERN for fabric monitoring. The future plans are to develop monalisa alarm triggers based on this information and to help in operating large facilities.
- 4) We developed a lightweight end user Agent (LISA – Localhost Information Service Agent) which is based on the java web start technology and provides complete host monitoring (cpu, memory, IO, paging, disk io ...), integrates end to end network performance measurements (e.g. iperf) and reports this information into monitoring services. It has a user friendly GUI to present all the measured values. LISA is using the discovery mechanism in MonALISA and can be used to dynamically connect a user application (like VRVS client or an analysis program) to the most appropriate service. This dynamic load balancing is based on the attributes published by the service and the quality of the connection with different sites which is continuously monitored.

Several improvements were done in the service itself, the repository system and GUI :

- Provide a real-time “Status” we page for all the GRID3 sites which shows the services with problems at different sites and the monitoring information that it is not available.
- A “Statistic” page shows all the major activities and the way resources are used for the entire Grid3 system
- GUI allows plotting summary information for different time intervals for large clusters or functional units. This includes mean, integral and the probability distributions.
- MonALISA service distribution includes now MySQL, Postgres, McKoi and a memory table. The user can easily select the storage option which more appropriate.
- Dynamic allocation of clients to proxy services based on connectivity and load of the services. This provides clients with the best connection in an automatic way, and does a global load balancing for clients and repositories.

Effort report from Conrad Steenberg of CMS on Analysis

Description:

Collaboration

The Clarens discovery service was proposed as an official component of the Open Science Grid consortium (OSG) and efforts at providing functionality required by the OSG community is ongoing via the PPDG CS-11, OSG Blueprint and PPDG Common projects. A test server was set up under the name <https://discover.gridservice.info> to act as a development implementation.

The CMS Monte Carlo Processing (MCPS) service project at Fermilab is continuing with expanded workflow management capabilities being implemented by the authors.

A second release of the Clarens server and clients is being prepared for release as part of the CMS Distributed Production Environment (DPE). The DPE offers an automated user-installable collection of software packages that is used for data production and analysis at CMS sites in the US.

A new version of the BOSS job submission and monitoring framework was released by the authors at INFN containing Clarens functionality in the default distribution.

A project was started at the Caltech Center for Advanced Computing Research (CACR) to use Clarens as the basis for a Astronomy Grid portal on the TeraGrid.

A joint demonstration for the Supercomputing 2004 conference is being prepared with the Fermilab CDF experiment to show an integrated, high-performance distributed analysis system based on Clarens web services, PROOF and ROOT analysis code.

Technical development

During the past three months the Clarens server was adapted to use a site-authorization service (e.g. SAZ) as an additional security measure. This allows site administrators to quickly and easily administer the users who have access to site resources in a central place. This is particularly important in cases where machines have been compromised and access to a large number of machines must be blocked to prevent attackers to use compromised credentials to access other machines at a site.

The site authorization functionality is also offered as a web service, allowing other servers to use Clarens as a SAZ proxy if it is more convenient to use web service tools than custom SAZ tools.

Further integration work with the MonALISA monitoring system was also done. Performance statistics including request rates and fault numbers are now reported to a set of station servers for republication to the MonALISA network.

Service, method, and server certificate details are now also published to MonALISA in the same way, to provide the basis of a new discovery service. In tests, the MonALISA-based dissemination of discovery data proved to have lower latency, less of a performance impact on servers to report data, and offers more complete searching and archiving of data by a global repository than the previous web spider-based discovery service.

The web service part of the discovery service is handled by JClarens, which acts as a client to the Java-only JINI network used by MonALISA.

Some parts of the older discovery service are still being implemented, notably the ability to retrieve cryptographically signed service API descriptions (WSDL) from the server. Changes are being made to MonALISA to allow the publication and retrieval of longer text-based monitoring data objects to allow this functionality to be implemented.

Work was started on making Clarens services available over an instant messaging (IM) network instead of using the client/server HTTP protocol. This will allow for efficient communication with transient "servers", including running jobs on compute nodes on private networks or behind firewalls. The bi-directional nature of the IM architecture also allows for asynchronous notifications to be sent to job owners, a key shortcoming of the request-response architecture of the HTTP protocol.

Demonstrations and publications

A presentation on the Clarens architecture and developments was given at the Computing in High Energy Physics (CHEP) meeting held at Interlaken, Switzerland. A paper was also submitted for publication in the conference proceedings, entitled "The Clarens Grid-enabled Web Services Framework: Services and Implementation" as paper 184.

Another presentation was given on the Clarens-based Grid-enabled Analysis Environment entitled "Grid Enabled Analysis for CMS: prototype, status and results", as paper 182.

Effort Report Abhishek Rana (UCSD):

Abhishek split his time equally between documenting the Open Science Grid blueprint activity and the common projects data management focus area. For the data management focus area he deployed "resilient dCache" at UCSD, and worked through a number of issues with the FNAL developers. He is presently deploying dCache on test cluster at FNAL, UCSD, Caltech, and MIT for an sc2004 demo of the Proof Enabled Analysis Center (PEAC).

4.4 D0

SAMGrid activities in this quarter which advanced the PPDG objectives included:

- continuing to add JIM installations at new sites for D0 Monte Carlo production
- continuing operational support for ongoing D0 Monte Carlo production
- developments in JIM to further support CDF Monte Carlo production and D0 reprocessing: updating and improving the JIM interface to workflow management for the two experiments
- beginning deployment of JIM at the Fermilab D0 reconstruction farm, which will pave the way for a fully unified grid job submission to be used for the full D0 reprocessing project, scheduled to begin in January 2005
- working on the connection of the SAM Monitoring and Information Server with MonAlisa

4.5 JLab

During the the July-September 2004 time frame JLab developed two SRM v2.1.1 clients for use by remote collaborators to transfer files to and from JLab. These clients use the SRM Web Service to obtain a transfer URL and Gridftp to transfer files. With these tools remote collaborators interact with the JLab mass storage system. The SRM client and web service have been installed at JLab. The SRM client has been installed at CMU. JLab is currently working with CMU to update their grid environment (Globus installation) and has begun testing the SRM client for data transfers. It is believed that the JLab SRM is the first one to be built on version 2.1.1 of the SRM specification.

JLab also worked on upgrading and deploying the laboratory's local job submission tool (Auger). This upgrade included additions and enhancements to facilitate future integrations with a Grid environment. The two most significant changes included the removal of dependencies on LSF for job monitoring and a refactoring of the authentication module with the intent of adding X509 certificate support (GSI).

4.6 BNL RCF/ACF

Effort report from Gabriele Carcassi of BNL-RCF/ACF on Security

Description:

Gabriele Carcassi is continuing his effort within the BNL site AAA project. He continues to develop GUMS (versions 0.6/0.6.1 released) and maintain its installation within BNL. The GT3 web service prototype of GUMS is ready, but more work needs to be done to understand how to achieve better performance with security.

He deployed new versions of VOMRS for STAR and PHENIX, and is finalizing the transition to the ATLAS VO in Grid3. He has been in contact with BNL cybersecurity to start the investigation and use of SAZ within the facility.

Effort report from Dantong Yu of BNL-RCF/ACF on Instrumentation & Monitoring

Description:

I developed BNL Core Grid service monitoring system for DC2. It collects the crucial status of core grid services: GridFtp, Grid Monitoring Services and Grid gatekeepers, Globus Replica Location Service, DC2 data registration tool (Don Quixote server), Message Passing Server by sending probing jobs periodically. It is based on the existing grid3 software with my local customization. The status page can be found at: <http://www.atlasgrid.bnl.gov/gridcat/>

BNL operators and iVDGL operation center (iGOC) uses this page to monitor the DC2 core servers 7/24 since DC2 production became stable. The emergency alarm will be forwarded to on-call person within an hour.

USATLAS DC2 support

I deployed four GridFtp servers and shared system administration roles on several other core grid servers. I organized group members to configure Tier 1 Grid-enabled facility to support data challenge at BNL and avoid single point of failure. I defined response procedures to different levels of emergency alarms.

High Performance Data transfer.

The project was to fill up the BNL OC12 connection with real data read from disk storage between BNL and CERN CASTOR. We used parallel approaches/peers to achieve aggregated high performance. I created two Gridftp servers for remote disk data transfer and created RRD name to do simple load balancing among these two servers. We achieved 45~70MB/seconds data transfer from CERN castor MSS to BNL local disks with data files produced in USATLAS DC1. I continued investigating the solution to sustain the data transfer for days or weeks. I participate into the Robust File Transfer project with joint effort from CERN and LHC Tier 1 centers and worked with CERN network expert to optimize CERN and BNL link.

4.7 STAR**4.7.1 Monitoring**

Efstratios, from the BNL Information and Technology division (ITD), started a cross-team project (involving the MonALISA developers as well) aimed to study and develop a set of tools providing monitoring data related to local and remote Resource Management System (RMS or batch system) to Meta-Schedulers and this, in a consistent fashion. The underlying idea being that such data could be the base of input for Schedulers, helping in making complex decision. We aimed at both local and remote but our initial testing and development aimed at local non-overlapping clusters.

Custom MonALISA monitoring modules were developed to collect monitoring data from LRMS. We started by using LSF information such as the number of jobs in a queue, the number of jobs in different state (pending, running) for a given time interval. Associated with the already provided information, the goal is to determine, when two clusters are involved in decision making, how “busy” a given cluster really is.

Additionally, two approaches were developed and tested to retrieve the monitoring data. A MonALISA pseudo-client and Web-service based client. Codes were developed to integrate both solutions within SUMS and tests were made by using two separate local clusters. This work was presented at the CHEP04 conference in the Distributed Computing Service session: *Development and use of MonALISA high level monitoring services for Meta-Schedulers* ([track ID 393](#)). Our initial studies showed that while the approach is sound and successful in a rather stable cluster load environment, we were susceptible to wrong decision making, with large adjustment latency if the state of a cluster would rapidly vary. We concluded that not only current values but differentials needed to be observed. This project will require more work to achieve the stability we require that is, a self-adapting system reacting to conditions and changes in load and queue occupancies.

We would like once again to acknowledge and thank Iosif Legrand for his support, advices and consistent help in the course of this project. His invaluable inputs and suggestions helped us in taking the proper steps toward developing the idea within the MonALISA framework.

4.7.2 Job Management**4.7.2.1 The STAR Unified Meta-Scheduler**

Much effort was made in the area. Levente Hajdu worked on reshaping the design and layout of the STAR Unified Meta-Scheduler (SUMS) so the package becomes both easier to later modify (plug-and-play features) and extend (base classes arranged in a more logical fashion). Several improvements were made and especially

- A new queue object implementation brought consistency to the definition of queues or Pools allowing for multiple queues to be taken into account within a priority. A static and dynamic priority was implemented to allow for strict ordering and ordering based on external criterion (such as information coming from information services or monitoring information).
- The above changes led to an easier way to implement new underlying batch systems; to date, we have support for 4 LRMS (LSF, PBS, SGE, Condor), one DRMS (Condor-G) and a hybrid dispatcher (BOSS). The most recent, Sun Grid-Engine scheduler was added and supported without much effort. Our LBNL/PDSF facility, with help from users, is testing this solution as a possible replacement for the currently used LRMS.
- We extended the U-JDL schema to allow for considering Storage Elements and Computing Element resources (memory and disk space mainly).
- A new JobID scheme. While we could not reproduce it, our initial time stamp based JobID had been reported as inappropriate by users (two sessions started within delta time would lead to the same time stamp which was set to the millisecond close). We changed our approach and implemented a universal id inspired mechanism: using MD5 hashes, we tested that this approach was reliable.
- The report log generated by SUMS at submission time is normalized, providing clear information to the users on what SUMS's decision led to as per the job splitting and dispatching.
- A new utility and feature was developed allowing users to re-submit or re-queue, kill and show the status of jobs previously scheduled without the need to know the RMS specific commands. The information is provided in a consistent fashion so the user would not need to interpret the result (status query) depending on what RMS was used.

A large part of the effort was to also work, test and make use of the queue monitoring information mentioned in the previous section. Several algorithms were developed for this purpose and while a fair amount of CPU cycles were needed to test the reliability of the approach, the exercise was very informative as already mentioned. We hope, as a first benefit, to be able to tune our algorithm and recover un-used CPU cycles at local level and will then expand to include our remote sites.

SUMS architecture and design was presented at the CHEP04 conference in the Distributed Computing Services as "*The STAR Unified Meta-Scheduler project, a front end around evolving technologies for user analysis and data production.*" ([track ID 318](#)). In fact, we took the opportunity to mention in this presentation the outcome of this project and impact on our scientific capabilities: to date, we realized that 80% of our active users have switched or are regular users of SUMS, in the past three months, 7 papers were submitted to refereed journals all analysis were performed using SUMS.

4.7.2.2 The Grid Collector

We continued work with the Grid Collector project, representing the next generation of SRM based aware user application. The core of the Grid Collector is an "Event Catalog". This catalog can be efficiently searched with compressed bitmap indices, developed by the Scientific Data Management Group (SDM) at LBNL. Tests show that it can index and search STAR event data much faster than any database systems. It is fully integrated with the current STAR analysis framework so that a minimal effort is required to use the Grid Collector in an analysis program. Furthermore, by taking advantage of the existing STAR replica, meta-data and file catalogs, Storage Resource Managers (SRMs) and GridFTP, the Grid Collector automatically downloads the needed files anywhere on the Grid without user intervention. Wei-Ming Zhang from Kent State University, John Wu from LBNL, and the STAR Software and Computing team located at BNL collaborated to fully integrate the Grid Collector into the STAR framework, and provide support for event based and micro-data summary tape (MuDST). Help pages and tutorials were developed as well as example macros. Several users moved from beta-testers to truly beneficial analysis mode of operation. As the latest example, a Berkeley scientist performed an analysis selecting events containing unusual particle trajectories (or atypical track topology), the events were believed to span two production series within a large minimum biased event sample. The entire event selection process took ten minutes for

a task which used to take almost a week if approached with a traditional “read all file and select events of interest as reading”. We believe that such success stories will spread amongst our collaborators and push this development to its next phase.

The Grid Collector project was presented at the CHEP04 conference as an oral presentation in the Distributed Systems and Experiences session as “*The Grid Collector: Using an Event Catalog to Speedup User Analysis in Distributed Environment*” ([track ID 319](#)). The abstract was submitted by Kensheng Wu and presented by Jerome Lauret.

4.7.2.3 U-JDL developments

During this quarter, we met a few times with David Alexander from Tech-X Corporation with whom we have started a Phase I SBIR envisioned, in part, to study the diverse available high level JDL (or User JDL) and attempt to homogenize the approach by consolidating it into a unique U-JDL. This work came as a natural extension of the URDL published at [PPDG 39](#). Five high level JDL were studied namely

- The JDSL from GGF
- The AJDL fro Atlas
- The URDL (PPDG 39) from STAR and J-Lab
- The DataGrid JDL attributes supported by the EDG WMS release-2 software (ClassAd XML)
- The J-Lab Auger JDL

We identified commonalities and concepts and tried to integrate them into a new high level U-JDL we will be developing within a close philosophy to the modular approach presented in the URDL version 0.4 of the PPDG 39 document. We intend to soon release a new version of the URDL for PPDG to review. In the process, we interacted with the J-Lab team and gather feedback as per the usability of the extended scheme in regards of the Auger system. All functionalities were acknowledged to be present although time constraints did not allow the J-Lab team to pursue a prototyping track at this stage.

We also designed the first version of the WSDL for an Abstract Job handling Grid service although time did not allow starting prototyping. We hope to accomplish this within the coming weeks, in parallel of a final SUMS release which would incorporate the new URDL as a proof of principles.

4.7.3 Data Management

Eric Hjort worked testing the integration of the DRM into VDT and provided feedback to the developers, mainly Alex Sim, as per readiness and usability of the deployment. In order to achieve this goal, and working as a non-root user, he got pacman to install VDT 1.2.1 and then DRM 1.2. The documentation up to this point was reviewed (pacman / VDT / DRM) and seemed very good.

Eric continued support for the STAR data transfers using HRM w/RRS. Recent developments in this area include reworking some of the software used to generate formatted file lists for HRM transfers and increase automation. This is now completely catalog based (older implementation actually looked in HPSS which is slow and not scalable) and works on a much larger scale: all production files with given software library could be requested at once instead of a combination of library, trigger, field, etc... as done before (the directory structure follows an encoding based on those values and the approach was an “abuse” of implicitly hard-coded Meta-Data). In other words, the transfer can be done purely based on a Meta-Data query. Another improvement was that the destination directories in HPSS are now also created automatically. The overall result is increased automation.

Eric also submitted an abstract to CHEP04 regarding use of HRM in STAR. It was accepted for an oral presentation in the Distributed Computing Services track as “*Production mode Data-Replication framework in STAR using the HRM Grid*” ([track ID 334](#)). As he was unable to attend in person, this was presented by Doug Olson.

4.7.4 Infrastructure and other activities

J. Lauret facilitated the communication between the diverse parties interested in having a RACF wide deployed version of the HRM/SRM and initially steered the effort toward the beginning of a project. Zhenping Liu (RACF), Alex Sim & Junmin Gu (SDM) were mainly involved in the development and integration. We are pleased that our efforts have motivated the RACF personnel and are satisfied the project is benefiting the community as a whole after a long usage standing in STAR for the past few years.

Gabriele Carcassi from the RACF has deployed and supported the RHIC and Atlas VO management systems based on VOMRS. This has been of a tremendous help to STAR, our VO mechanism being close to inexistence prior to this deployment.

We acquired a new Linux box which was aimed to serve as a new gatekeeper for the development of SRM related projects as well as resolving our past reported firewall issues (by using two network interfaces, one looking directly toward the outside perimeter). This node came in mid-summer and took longer than expected for setup & configuration by the RACF. While not causing a serious delay it precluded large-scale testing of the Grid Collector before the CHEP conference. A node set up at PDSF allowed the R&D to proceed. We strongly believe at this stage the use of remote facilities to be an essential part of our success and will prospect avenues to consolidate our goals.

We continued the regular schedule of meetings between STAR and Phenix. We discussed about job monitoring in general (BOSS is in use in Phenix as the current solution), shared our views and experiences, exchange ideas on SE approaches in schedulers, reproducibility of results in a distributed computing environment as examples. The meeting also allowed to better understand Phenix's needs as per the desired feature in SUMS. We also allowed Mike Reuter to share our infrastructure for a production on grid demonstrator: as reported in the last quarterly, our own attempts went through initial difficulties and since our observations were that stability could be achieved from the most powerful and best network logistically placed machine, we hope this will facilitate their own testing.

4.8 PHENIX

Mike Reuter has been working on doing PHENIX simulations on the grid. This includes job submission, software deployment, job monitoring and data handling. Jobs are created using a GUI which writes a job configuration file, embedding PACMAN commands for software deployment. The job configuration is an input to SUMS.

The GUI allows user to choose event generator, input parameters to event generators, control for reconstruction. PACMAN calls to retrieve the software are included in the jobs, but Postgres database must be pre-installed for the reconstruction step. Still todo is finish implementation of PYTHIA, vertex location selections for PISA, Tracking and passing of random number seeds.

It has been tested so far on RedHat 8 OS only. Install time is ~10minutes, so the overhead is not large for PISA +reconstruction in each job. Tests so far have been rather small, so scale will be determined by further tests. 100 minbias events in HIJING takes >2 hours; 100 single particle events fully processed in 15 minutes.

Successfully tested SUMS and BOSS+SUMS. Use of BOSS was very easy, thanks to Andrey's work. Boss grabs STDIN, STDOUT, STDERR from SUMS from SUMS, so they get left on the executing machine. This needs to be fixed; a script that wraps BOSS submissions with Globus has been tried. Scheduling + monitoring needs to be combined into the same interface. Main problem that is left is to automate return of output & stdout, stderr files. Mike has identified some needed SUMS modifications to allow successful running at StonyBrook Chemistry cluster and UNM. Output file handling: can do by a simple grid copy command into the job stream (so need Globus deployed). SRM is needed for pre-staging input files. Site selection is currently done by hand; the target directory for output must be pre-existing.

GUI and software should be ready soon and Mike plans to run a demo production in early November.

4.9 Condor

The Condor team continues to lead the VDT build and support effort as well as improvements to Condor. The Condor-C component (somewhat analogous to Condor-G) has been recently developed to manage a VO-specific queue of work at a site. A brief description is included in a presentation on VDT at the Open Science Grid meeting in September, see <http://www.opensciencegrid.org/events/meetings/boston0904/docs/vdt-roy.ppt>.

4.10 Globus – ANL

4.10.1 Coordination and Support

Primary coordination vehicle is via the Open Science Grid blueprint activity. Support through the discuss lists and Bugzilla are available to all experiments.

4.10.2 Globus Toolkit 2.x Updates, Bug Fixes and Open Issues

There have been no PPDG bugs submitted or resolved during the quarter. Bug 1725, which has a Trivial severity, is still open. There were 2 bugs submitted just after quarter end, one of BLOCKER severity and one of CRITICAL severity. We are working with Dantong to better understand these issues. Additional information about Bugzilla bugs can be found at <http://bugzilla.globus.org>.

4.10.3 Globus Toolkit 3.2

GT 3.2 continues to be the current stable release for the Globus Toolkit. It contains both the OGSi based web service components, and the pre-WS components (the components similar to the 2.x releases). The WS components offer some substantial performance improvements over the 3.0.x versions. The pre-WS contain mostly bug fixes. However, GridFTP did include some additional, very useful functionality. It added support for structured directory listings in the protocol, which enabled a multi-file globus-url-copy as well as the ability to specify a directory and have it recursively moved in both globus-url-copy and the Reliable File Transfer service.

4.10.4 GridFTP / XIO

We continue to work on the new GridFTP server. A new development release is scheduled for the end of October. This release will be feature complete, have stable interfaces, and should be very stable for a development release. It would be very suitable for initial compatibility testing, something that will be critical to the ease of transition once the final release is out in February. Over the next quarter no new features will be added and the entire focus will be on bug fixes, robustness, ease of use, and performance.

XIO work over the last quarter focused on providing additional drivers. We completed several new drivers to round out the functionality and usability of XIO. We added a MODE E driver, which enables Open/Close/Read/Write (OCRW) access to the GridFTP data channel (multiple TCP streams). Since MODE E can have out of order arrival of data, we implemented an ordering driver which buffers the blocks and presents them in order to the application. We also implemented a GridFTP driver which allows OCRW access to a file via a GridFTP server. Finally, we implemented a queuing driver to allow multiple outstanding writes. We are also working on a UDT driver (reliable UDP based protocol from Bob Grossman's group at UIC). Currently it is functional, but we are not getting the performance out of UDT that we should. We are currently working with the UIC crew to resolve these issues.

4.10.5 Monitoring and MDS work

The focal point of information services work over the last quarter has been two fold. First, we completed the port from the Open Grid Services Infrastructure (OGSI) to the Web Services Resource Framework (WSRF). Second, we added three new features / components, a trigger service, an archiver, and web based user interface. The trigger service performs some action, such as sending email, when a specified condition

becomes true (eg. disk full). The archiver keeps historical monitoring information and makes it available for later use. Webmds allows user to see monitoring information with just a web browser.

4.11 SRM

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

The main activities in this quarter included deployment of Berkeley_SRM in multiple sites, coordinating and running two collaboration meetings for RRS and SRM, and the writing of draft functional specification for the RRS 1.0 and the SRM v3.0. These activities are described below.

Replica Registration Service meeting and draft spec

The goal of the Replica Registration Service (RRS) is to provide a uniform interface to various file catalogs, replica catalogs, and metadata catalogs. It can be thought of as an abstraction of the concepts used in such systems to register files and their replicas. Some experiments may prefer to support their own file catalogs, and some systems use metadata catalogs or other catalogs to manage the file name spaces. Our goal is to provide a single interface that supports the registration of files into such name spaces as well as retrieving this information.

Early in September we hosted a meeting at Berkeley Lab with participation from CERN and USC/ISI to coordinate the RRS activity. We have completed a draft version 1.0 of the RRS specification and submitted it for comments to the RRS collaboration.

In addition to the RRS, we have identified three other services, namely the Replica Copy Service (RCS), Replica Selection Service (RSS) and the Replica Management Service (RMS). The RMS inherits all the functionality of the RRS, RCS and RSS and provides additional, coordinated capabilities of copyAndRegister and unregisterAndDelete. In August we had a coordination meeting in Chicago for the RMS activity with participation from Condor, Globus (both ISI and ANL) and the University of Wisconsin. All participants showed great interest for a joint interface specification of RMS. We have submitted a draft version for comments.

Storage Resource Management meeting and draft spec

The meeting, held in early September, discussed two major issues: 1) the state of the current SRM specification for versions v1.1 and 2.1.1, and 2) how to proceed with future versions and how they relate to the SRM-Basic and SRM-Advanced versions proposed in the Grid Storage Management (GSM) group at GGF. The main conclusions relative to future versions were:

- We should strive to have a uniform versioning method, where each successive version is backward-compatible with a previous version.
- Future versions should consist of a “core” set of functions and “advanced” functions. The core version, which will be called SRM-Basic should be defined independently of the advanced version, called SRM-Advanced. The advanced functions will be grouped into “feature-sets”.
- There will be multiple SRM-advanced versions, starting with the SRM v3.0. SRM v3.0 will include all the functions of v2.1, the correction of ill-defined functions, as well as the separation of the core and the advanced features.
- It should be possible to map the SRM-Basic version to SRM v1.1. When accessing an SRM-Advanced server, it should be possible to find out what features are supported by that SRM. To support this, there will be a function available with enumeration of the feature-sets.

A draft version of SRM 3.0 was written and will be circulated for comment to the collaboration in the next quarter.

Deployment activities

- **Deployment of Berkeley-SRMs in production at BNL**

The LBNL SRM-team (Alex, Junmin, Viji) provided support to Jane Liu at BNL for deploying the DRMs and HRMs. The support included guidance in setup, resolving problem

that arose because of security, and setting up testing. The Berkeley-SRMs are now deployed in the production system in STAR and ATLAS nodes. There are plans to deploy the Berkeley-SRMs in other US ATLAS sites. A presentation on this experience was given in CHEP by staff members at BNL (see:

<http://indico.cern.ch/getFile.py/access?contribId=345&sessionId=7&resId=0&materialId=slides&confId=0>)

- **Support Ed May for using DRM**

The purpose of this activity is to use DRMs for analysis use cases in ATLAS. The SRM-team help guide Ed in installing DRM from the VDT binaries, and set it up for local configuration. So far, only basic tests were made, and access to NERSC-HPSS by connecting to an HRM in LBNL is underway

- **Support the use of DRM on top of SRB in the UK**

Dimitris Tsirigkas at EPCC, UK has a project of using a DRM (and the DataMover) on top of SRB. The idea is to replace the code in DRM that invokes GridFTP with calls to SRB. Help was provided in understanding how to use the DRM and the DataMover. Both were installed by Dimitris from VDT.

- **Support the deployment of DRM at Caltech for CMS**

Three people from Caltech have experimented in installing and using the Berkeley-DRM (Frank van Lingen, Conrad Steenberg, Michael Thomas). Support was provided for completing this task successfully.

Enhancement activities

- **Enhanced HRM to access GSI-enable HSI**

The Berkeley-HRM which accesses HPSS uses HSI to find out tape-IDs for requested file. It uses this information to order request for files in tape order in order to minimize tape mounting. To keep this importance feature, the HRM had to be enhanced to properly interface to the newly installed GSI-enable HSI.

- **Continued cooperation with the NeST team**

Junmin continued to test the SRM on top of NeST. This helped discover a bug when the gsiftp file is brought into a NeST lot is larger than the lot. This led to a design for an enhanced version of NeST, an activity currently underway.

- **Performed various enhancement to SRMs**

The main features added are: authorization testing from HPSS for sharing file already in the SRM cache, the packaging of the web-gateway to SRMs (to be deployed in VDT), and enhanced logging of HPSS failures to track tape failures.

4.12 SRB

Effort report from Reagan W. Moore of SRB-SDSC on Data Management

Description:

I participated in multiple working groups of the Global Grid Forum, including the Grid File System WG, a Grid Robustness and Reliability group, the Simple API for Grid Applications WG (SAGA-WG), the Preservation Environments WG, and the Grid Storage Management WG. Several standards efforts are going forward:

- integration of the SRM version 3 interface with the Storage Resource Broker (Peter Kunszt, CERN)
- standard APIs for C library, shell commands, and Java classes. The SRB APIs for each access environment were provided to the SAGA-WG to simplify their task in developing a minimal set of access mechanisms

- SRB file system commands were provided to the Grid File System group to demonstrate the set of functions needed in remote file system access.

Effort report from Wayne Schroeder of SRB-SDSC on Data Management

Description:

Support, coordination, and planning continued with the BaBar SLAC team. SLAC reports stable and reliable operation of the SRB. They plan to upgrade to SRB 3.2 shortly, and use the technology to federate two independent data grids between SLAC and IN2P3 in Lyon.

During this quarter the SRB team released SRB version 3.2 and 3.2.1 that contain a number of improvements and bug fixes that will be of use to the BaBar experiment. See <http://www.npaci.edu/DICE/SRB/CurrentSRB/ReleaseNotes3.2.html> and <https://lists.sdsc.edu/pipermail/srb-chat/2004-August/001213.html> for more information.

In collaboration with John Bresnahan of the GridFTP team, we have made significant progress in the integration of a GridFTP driver into the SRB and an SRB interface into GridFTP. In testing, we were able to successfully operate GridFTP as an SRB resource.

SRB now supports the Grid Toolkit 3 version of Grid Security Infrastructure. This was needed for the GridFTP integration work and will also be useful in other projects. A description is available at <https://lists.sdsc.edu/pipermail/srb-chat/2004-September/001299.html>

An obfuscation system has been developed such that SRB user password credentials can be saved in a non-plain-text form. Although based on source code secrecy, this does provide more security by presenting another obstacle to the interception of passwords and credentials. In some ways, this (with SRB ENCRYPT1) is more secure than GSI (particularly for server to server authentication) since GSI credentials are usable as recorded on disk.

Also of interest to the PPDG community, Simon Metson of Bristol University has released his GMCat server that provides a mapping between SRB file spaces and other Grid tools by echoing the functionality of the EDG LRC component of the RLS. See <http://www.npaci.edu/dice/srb/contributed.html> and <http://tuber1.phy.bris.ac.uk:8080/GMCatWS3/>.

4.13 IEPM, Network Performance Monitoring

4.13.1 Bandwidth/Throughput measurement (IEPM-BW)

We presented a talk on the Internet End-to-end Performance Monitoring at the Energy Science Network Site Coordinators Committee (ESCC) meeting, see <http://www.slac.stanford.edu/grp/scs/net/talk03/escs-jul04.ppt>.

Following concerns about the impact of iperf testing on network traffic, we examined the effects and documented them at <http://www.slac.stanford.edu/grp/scs/net/case/iepm-jul04/>.

We added U Victoria to the sites monitored by IEPM-BW.

We installed and studied BWCTL from Internet 2 which we will use for scheduling on-demand bandwidth measurements.

4.13.2 Lightweight Bandwidth Estimation

We assisted the people from SDSC/CAIDA to evaluate pathload on the IEPM-BW testbed.

4.13.3 Bandwidth performance anomalous events

We implemented and extended the [NLANR "plateau" algorithm](#). We have tuned the anomalous event detection using the "Plateau Algorithm" to minimize diurnal effects.

4.13.4 Traceroute Analysis and Visualization

We gave a talk on [Traceanal: a tool for analyzing and representing traceroutes](#)

We improved the traceroute topology visualization tool and applied it to the AMP data and incorporated it into the IEPM-BW traceroute visualization.

4.13.5 PingER

We worked with Florida International University to get agreement to install a PingER monitoring site there. This will be particularly useful to understand South American connectivity. We also successfully set up monitoring to an Indian commercial site (most PingER sites are Academic & Research, and we want to find out if the poor performance to India extends to commercial sites.

We worked with NIIT/Pakistan to develop a mouse sensitive map of PingER deployment.

4.13.6 SC2004 Bandwidth Challenge

We set up a collaboration with Caltech, FNAL, University of Manchester, England, several companies (e.g. Chelsio, S2io, Sun), ESnet, National Lambda Rail and other to participate in this year's SC2004 Bandwidth Challenge. For more details see the web site at <http://www-iepm.slac.stanford.edu/monitoring/bulk/sc2004/hiperf.html>.

As part of this we secured loans of two 10GE wavelengths (from NLR and ESnet/QWest) from Sunnyvale to Pittsburgh (this year's site for SC2004), the loan of eleven Sun Opteron compute servers, 11 10GE interfaces, three Sun file servers, Cisco equipment (XENPAKs and routing blades), space at the Sunnyvale CENIC point of presence.

4.13.7 Proposals and Representation

We submitted and had accepted a proposal to the US Department of State and the Pakistan Ministry of Science and Technology for *Measurement and Analysis for the Global Grid and Internet End-to-end performance (MAGGIE)*

We amended the DoE/SciDAC proposal on *TeraPaths: A QoS Enabled Collaborative Data Sharing Infrastructure for Peta-scale Computing Research*, had it accepted and received funding. A two page summary of the new proposal can be found at <http://www.slac.stanford.edu/grp/scs/net/proposals/iepm-bw/dgnmi-2p.doc>.

SLAC is a partner in the UltraLight optical testbed proposal (led by Caltech) which was funded by the NSF. We collaborated with Texas A&M, NASA and others to submit a proposal to NASA on developing and monitoring IP based protocols for NASA satellites etc.

We attended the NASA/LSN workshop on Optical Network Technologies at NASA Ames. We prepared and gave a presentation on WAN Monitoring Issues (see <http://www.slac.stanford.edu/grp/scs/net/talk03/jet-aug-4.html>) and also served on a panel.

Submitted paper with FNAL to CHEP04 on *Wide Area Networking System for HEP Experiments at FNAL*.

We attended the DoE PI network research meeting at FNAL and made two presentations:

- *TeraPaths: DataGrid WAN Monitoring Infrastructure* see <http://www.slac.stanford.edu/grp/scs/net/talk03/scidac-dwmi-sep04.ppt>.
- *PingER Project*, see <http://www.slac.stanford.edu/grp/scs/net/talk03/scidac-pinger-sep04.ppt>.

5 Appendix

5.1 List of participants

TEAM	Name	F	Current Role CS	Systems and Production Grids	Job Mgmt	Data Mgmt	AAA	Grid Analysis and Catalogs	Other: Web Services, Evaluations Interoperation, etc.
Globus/ANL	Ian Foster	Y	Globus Team Lead, GriPhyN PI, iVDGL, GriPhyN			x			
	Mike Wilde	N	GriPhyN coordinator, ATLAS- CS liasion	x		x			
	William Allcock	Y	GridFTP	x		x			x
	Von Welch	N	CAS				x		
ATLAS	John Huth	N	ATLAS Team lead, GriPhyN Collaborator	x					
	Torre Wenaus	N	LCG Applications liason		x	x		x	
	L. Price	N							
	D. Malon	N	Database/POOL Liason					x	
	A. Vaniachine	N						x	
	E. May	N	Testbed applications	x		x			
	Kaushik De	N	Testbed applications	x					
	David Adams	Y	Distributed analysis					x	
	Wensheng Deng	Y	Metadata catalogs			x		x	
	R. Gardner	N	IVDGL coordinator, Atlas Grid Tools		x	x			x
	G. Gieraltowski	Y	Interoperability	x				x	x
	Dantong Yu	Y	Monitoring and VO	x				x	
	Gabrielle Carcassi	Y		x	x				
BaBar	Richard Mount	N	PPDG PI, BaBar Team co- Lead						
	Tim Adye	N	BaBar Team Co-Lead						
	Robert Cowles	N						x	
	Andrew Hanushevsky	Y				x			
	Adil Hassan	Y				x			
	Les Cottrell	N	IEPM Liaison	x					
	Wilko Kroeger	Y				x			
CMS	Lothar Bauerdick	N	CMS Team Lead. GriPhyN collaborator						
	Harvey Newman	N	PPDG PI. GriPhyN collaborator, Co-PI iVDGL						
	Julian Bunn	N	CMS Tier 2 manager, GriPhyN & iVDGL collaborator	x				x	
	Conrad Steenberg	Y	CS-8:Analysis Tools, GriPhyN collaborator					x	x
	Frank Lingren	Y	CS-8:Analysis Tools, GriPhyN collaborator					x	x
	Iosif Legrand	N	CS-8:Monitoring Tools						x
	Vladimir Litvin	N	GriPhyN collaborator		x				

	Michael Thomas	Y	CS11					X	X
	James Branson	N	CMS Tier 2 manager	x					
	Ian Fisk	N	CMS Level 2 CAS manager, iVDGL liaison	x					
	James Letts	Y	Working on VDT testing scripts	x					
	Saima Iqbal	N	web technology evaluation					x	
	Suresh Man Singh	N	grid deployment	x					
	Anzar Afaq	Y		x	x			x	
	Greg Graham	N		x	x			x	
Coordination	Ruth Pordes	Y	PPDG coordinator			x			
	Doug Olson	Y	PPDG coordinator			x	x	x	
	Miron Livny	Y	PPDG coordinator		x	x	x	x	
	Joseph Perl	Y	CS-11 co-coordinator					x	
D0	Wyatt Merritt	N	D0 team Lead	x	x				
	Igor Terekhov	Y	JIM Team Lead	x	x				
	Andrew Baranovski	Y		x					
	Gabriele Garzoglio	Y		x	x	x			
	Sankalp Jain	Y	Through contract with UTA CSE Department	x	x				
	Aditya Nishandar	Y	Through contract with UTA CSE Department	x	x				
SRM/LBNL	Arie Shoshani	y	SRM Team Lead. GriPhyN collaborator			x			
	Alex Sim	Y				x			
	JunminGu	Y				x			
	Viji Natarajan	Y				x			
SRB/UCSD	Reagan Moore	Y	SRB Team Lead. GriPhyN collaborator			x			x
	Wayne Schroeder	Y	CS-8: Web Services			x			x
JLAB	William Watson	Y	JLAB Team Lead			x			x
	Sandy Philpott	N	facilities	x			x		
	Andy Kowalski	N				x			
	Bryan Hess	Y	Web Services			x			x
	Ying Chen	Y	Web Services	x		x			x
	Walt Akers	N	Web Services	x		x			
STAR	Jerome Lauret	N	STAR Team Lead	x	x				x
	Dave Stampf	N		x					
	Richard Casela	N		x					
	Efratios Efstathiadis	N		x					
	Eric Hjort	Y		x		x			
	Doug Olson	N		x		x			x
	Levent Hajdu	Y	Star scheduler and JDL	x	x	x			
Condor/U.Wis consin	Miron Livny	Y	PPDG PI, PPDG Coordinator. GriPhyN collaborator	x	x	x			x
	Peter Couvares	Y				X		x	
	Alan DeSmet	Y				x		x	
	Alain Roy	N				x			

	Todd Tannenbaum	Y			x				
	Jeff Weber	Y	CDF, D0 support	X					
Globus/ISI	Carl Kesselman	N	Globus/ISI lead						
	Ann Chervenak	N				x			
PHENIX	David Morrison	N	Team lead						
CDF	Frank Wuerthwein	N							
	Rick Snider	N							
ALICE	Larry Pinsky	N							

5.2 Meetings

Wednesday, July 7, 2004

PPDG weekly meeting

URL: <http://www.ppdg.net/mtgs/phone/040707/default.htm>

Thursday, July 8, 2004

PPDG CS-11 - grid data analysis phone meeting

Monday, July 12, 2004

OSG Blueprint Meeting, Madison

URL: <http://www.ppdg.net/pa/ppdg-pa/blueprint/index.html>

Tuesday, July 13, 2004

OSG Blueprint Meeting, Madison

URL: <http://www.ppdg.net/pa/ppdg-pa/blueprint/index.html>

Wednesday, July 14, 2004

OSG Blueprint Meeting, Madison

URL: <http://www.ppdg.net/pa/ppdg-pa/blueprint/index.html>

PPDG weekly meeting

URL: <http://www.ppdg.net/mtgs/phone/040714/default.htm>

Thursday, July 15, 2004

OSG Blueprint Meeting, Madison

URL: <http://www.ppdg.net/pa/ppdg-pa/blueprint/index.html>

Wednesday, July 28, 2004

PPDG Steering meeting

Thursday, August 5, 2004

PPDG CS-11 - grid data analysis phone meeting

Wednesday, August 18, 2004

PPDG weekly meeting

URL: <http://www.ppdg.net/mtgs/phone/040818/default.htm>

Thursday, August 19, 2004

PPDG CS-11 - grid data analysis phone meeting

Monday, August 30, 2004

Replica Registration Service meeting, LBNL

Tuesday, August 31, 2004

Replica Registration Service meeting, LBNL

Tuesday, September 7, 2004

OSG Blueprint Meeting, MIT

URL: <http://www.opensciencegrid.org/events/meetings/blueprint0904/>

Wednesday, September 8, 2004

OSG Blueprint Meeting, MIT

URL: <http://www.opensciencegrid.org/events/meetings/blueprint0904/>

PPDG weekly meeting - cancelled

Thursday, September 9, 2004

OpenScienceGrid Workshop, Boston

URL: <http://www.opensciencegrid.org/events/meetings/boston0904/>

Friday, September 10, 2004

OpenScienceGrid Workshop, Boston

URL: <http://www.opensciencegrid.org/events/meetings/boston0904/>

Wednesday, September 15, 2004

PPDG weekly meeting

URL: <http://www.ppdg.net/mtgs/phone/040915/default.htm>

5.3 Related Publications