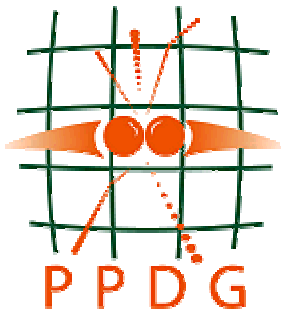
Particle Physics Data Grid Collaboratory Pilot

Quarterly Status Report of the Steering Committee, April - June 2002

13 Sept 2002



1 Project Overview

1.1 Highlights

This quarter marks the end of the first year of the SciDAC funded PPDG Collaboratory Pilot. Several activities in this quarter marked that transition from the first year – where the focus is on peer experiment-computer science group projects and single collaborator activities – to the second year – where we need to a) focus on sharing of technology , interfaces and experiences towards more common grid services in the experiment applications b) move up the service stack to address higher functional levels of end to end experiment grid systems.

Several small but significant steps were made towards the PPDG goals of Grid services integrated into Experiment Data processing production systems. A news brief was sent to DOE to mark the simulation of 50,000 CMS events over a 5 site Grid. The events were accepted as part of the official CMS simulation production - albeit only constituting less than 10% of the simulations.
http://www.ppdg.net/docs/news/news-update-cmstestgrid-17may02.doc

As end-to-end applications that rely on Grid Services become more of a reality the need for and importance of troubleshooting, errors has taken on more significance. Together with collaborators on other HENP projects we submitted a pre-proposal to DOE for discussion and feedback. DOE is working with NSF to plan a workshop to address this broad issue in order to start planning how to move forward to define projects and activities to best benefit the community. In the short term PPDG will undertake pragmatic, small tasks as needed for our experiments' program.

As planned, we held a series of internal reviews of the project activities – joint experiment-computer science group projects. These were regarded as valuable in providing a technically focussed forum for interchange between those working on the project, other members of the collaboration acting as reviewers, and the executive team. Following the reviews the Project Activity and Cross-Cut Collaboration Activities were realigned to provide more focus on the CS – common services – approach to Application Grid Service needs and components.

This also supports our increased collaboration with peer HighEnergy and Nuclear Physics Grid projects whose goals include delivery of middleware to an overlapping set of experiments: the US Physics Grid Projects – Trillium – iVDGL, GriPhyN and PPDG – and the EU Physics Grid Projects – in particular the European Data Grid and DataTAG.

Shazhad Muzaffar who had provided significant GDMP development and support both for PPDG and EDG left to take up a position at Cern, and transitioning his developments to support for the CMS analysis tools (CAS-IGUANA).

1.2 Project Management and Organization

The Project Plan for Year 2 was written and endorsed by the Steering Committee. It was presented to DOE for discussion at the visit on June 7th. While not a detailed task and project list, the plan for Year 2 outlines the focus of the project and the path we will follow.

Joseph Perl was identified to help coordinate across the project, to work with Doug Olson on the CS-11 Analysis Tools activity.

The collaboration took the very difficult decision to redistribute some of the Year 2 funds. Each of the experiment collaboration teams have contributed funds to a cross-project fund for Education and Outreach (based at Fermilab) and Web Site management and design, and support services (based at LBNL). These funds have not been used to date. There will be specific proposals and reporting to the collaboration as they are used.

Additionally PPDG have provided funds to the Globus/ISI group to enable contributions of ISI to PPDG specific extended requirements and integration tasks:

”PPDG CS-5 Reliable Replica Management Services. PPDG is collaborating with the European Data Grid on the GDMP publish/subscribe application for file replication. The next phase of EDG data access (WP2)development will rely on the Giggle replica location service from Globus and enhanced replica management service. PPDG specific requirements for this service include support for a Master Replica. This will be included in Giggle over the next few months, interfaced and integrated with the rest of the EDG WP2 and available as part of the overall file replication services for the PPDG experiments. ISI will continue to collaborate with the PPDG project activities and cross cuts for Reliable (File) Replica Management, and be available for presentations, consulting and to discuss further collaborative work on request.

PPDG CS-3 Monitoring and Status Reporting: PPDG participates as part of the GriPhyN/iVDGL/PPDG joint monitoring working group and HICB/JTB GLUE efforts. In particular the GLUE effort is requesting extensions to and modifications of MDS, the sensors, the default and supported schema, integrating with the experiment monitoring frameworks and monitoring applications. ISI will increase its collaboration with PPDG experiments and their integrated test beds in the support for MDS core infrastructure.”

1.3 Plans for the next Quarter

During the next quarter we will attend the High Energy Intergrid Coordination Board meeting, GGF and HPDC – which is collocated with the HICB. We will report on the progress with Interoperability – especially authentication and certificates, glue schema and issues with Authorization and Job Execution.

We will complete the planning for the joint PPDG, GriPhyN, iVDGL collaboration meeting in December at SDSC.

We hope to have an opportunity to follow up on issues of catalogs and troubleshooting.

1.4 Trillium Collaboration: PPDG, GriPhyN and iVDGL

The increased coordination and synergy between the three US Physics Grid Projects – GriPhyN, iVDGL and PPDG – has led to the concept of the overall “Trillium” collaboration and presentation to our European and other partners. This strategy has strengthened the discussion of common and convergent goals; overall definition of scope and deliverables; and discussion of prioritization and milestones resulting in the movement of resources between the three projects as needed.

The fact that LHC is a significant focus of the US HEP program over the next few years, coupled with the fact that ATLAS and CMS are collaborators on all US Physics Grid projects, together with the shortfall in funding for the LHC S&C projects, together with the progress and successful collaborations in the US, has resulted in Trillium taking a major focus on the LCG project deliverables and milestones. This situation is leading to responsibilities and expectations for production deployment and middleware software support and maintenance that stress the expectations and plans of the Trillium projects to date. It is a measure of success of the projects that we are being approached and relied on for development, integration and support in these areas. It is however clear that the projects are stretching rather limited resources to carry out these duties. The iVDGL management must address funding and resource issues in the remainder of 2002.

1.5 Virtual Data Toolkit (VDT)

PPDG is contributing to the Virtual Data Toolkit (http://www.ivdgl.org) which is a deliverable from the GriPhyN and iVDGL projects. US CMS UCSD PPDG manpower, James Letts, is developing a test suite and interface to help with the testing of the toolkit and validation of installations, etc.

1.6 Collaboration with EDG and EDT (European DataTAG)

Collaboration with the EDG occurs both through working with the middleware workpackages and the computer science groups that are part of PPDG. Specific work was accomplished with WP1, WP2, WP5 and WP7 (Networking). Collaboration with DataTAG is happening through the GLUE ivdgl/datatag effort – to which PPDG is providing significant resources – glue schema effort from DMF and Globus (see below); authentication and authorization interoperability – CMS Caltech Conman and SiteAAA. Doug Olson is providing coordination between all these activities. The ATLAS and CMS test grids are collaborating through DataTAG WP4.3, 4.4 – led by Flavia Donno – in preparing to extend their sites to include European collaborating institutions.

Doug Olson, Ruth Pordes, Peter Couvares are members of the High Energy Physics Intergrid Joint Technical Board, which has sponsored the GLUE schema document (http://www.hicb.org/glue/glue-schema/schema.htm). Several other PPDG members- Jenny Schopf, Brian Tierney, Conrad Steenberg, Ewa Deelman and John McGee – have contributed to the Glue and JTB work.

2 Common Service Areas

2.1 Introduction

Text contributed by the individual teams has been transferred verbatim to the corresponding CS area. While this may lead initially to some awkwardness in construction, we hope that in future teams will report following CS guidelines as a help towards encouraging common work across the groups.

The PPDG architectural components as presented briefly in the last quarterly report have been further defined below. The next steps will be to start identifying existing protocols, required functional interfaces between the services and use cases of end to end applications using the services, based on the PPDG team applications and grid systems.

2.2 CS-1, CS-2 Job Description Languages, Management and Scheduling

2.2.1 Job Scheduling and Management Focus Meeting

All experiments and computer science groups attended this meeting in May and discussed common components and services towards grid job scheduling. All experiments plan to deploy some monte carlo or data processing applications over the grid in the next twelve months. A summary document refining our understanding of the steps in a Grid Job, the planning and scheduling management and interfaces was written.

2.2.2 Hardening of Condor and Globus technologies

The focused effort to run simulation production on a CMS TestGrid of 5 sites required significant effort from the US CMS PPDG, GriPhyN, iVDGL and US CMS S&C project, as well as the supporting computer science groups. This required modifications and bug fixes to condor-g, condor, and globus components, as well as the application MOP and IMPALA frameworks.

As a result of the experience on the US CMS TestGrid, and virtually simultaneous experience of the EDG Testbed 1.1, the Globus GASS-cache server is being extended and redesigned by the Globus and Condor developers. This work will benefit HENP grid projects in Europe as well as in the US.

2.2.3 Collaboration with EDG WP1

A 2 day requirements and design meeting was held in Italy in May with the EDG WP1 and WP2 leadership to better understand and design integration of the data and job management services for Release 2. This design is in line with the PPDG Job Management strategies developed at the May focus workshop. EDG is

providing enhancements to the DAGMAN service which which will additionally satify requirements for PPDG. Better definition of the interfaces to the WP1 Resource Broker should allow future versions to be integrated into other data handling systems.

2.2.4 Extensions to Condor and ClassAds for D0

The D0 job and resource management sam-grid project will use Condor Matchmaking for the Resource Broker functionality. This required migration of the matchmaking service to Condor-G, extension of the inherent negotiator to include determination of the Gatekeeper to which the job will be submitted, to be determined as the job is submitted; and extension of the ClassAds to call user functions.

2.2.5 ATLAS Distributed job management

The primary activity in the area of distributed job management over the last quarter was production of of simulation data for ATLAS DC1 (data challenge 1). Pavel Nevski at BNL is coordinating this production and integrating it with various GRID tools, in particular Magda and a VDC (virtual data catalog). During the last quarter, he and Alexandre Vaniachine at ANL developed a toolkit was developed to manage production using a VDC. BNL DC1 production is being carried out in that environment. Description of the VDC and the catalog itself (at least for ATLAS members) may be found at

http://atlassw1.phy.bnl.gov/NOVA/VDC.htm

The virtual data catalog is split into a set of logically independent orthogonal components including site description, software signature description, physics application parameters, etc. Production parameters that are subject to changes during the DC1 cycle are specified in the catalog. The input to one processing step (transformation) may be the output of previous steps described in the same catalog. This allows the possibility of dynamically regenerating data from the initial specification or any intermediate stage.

The DCI simulation production code was extracted and bundled so that it can be distributed to production sites using Pacman. External dependences (MySQL, ROOT I/O, Compiler-specific libraries, etc.) are encapsulated in a "sandbox" which can be exported to any Linux operating system and executed in a consistent manner. The distribution is maintained in the BNL_PAS Pacman cache. Other sites, especially those in the U.S., can use this Pacman installation and the VDC to contribute to DC1 production. This distribution is based on ATLAS production release 3.2.1.

2.3 CS-3 Information Services

2.3.1 Joint PPDG, Griphyn iVDGL Monitoring Project

2.3.1.1 Talks

Jennifer Schopf was asked to present the work by this group on monitoring requirements at the CMS Grid Production Workshop at CERN in June. This talk served to focus a half-day discussion on monitoring, and was well received.

2.3.1.2 Requirements document

Work on the requirements document did not move forward much this quarter. Yujun Wu provided some edits, but other editors that were working on this document have not followed through. We have identified this as a focus for the next quarter and hope to have a new draft of the document out (in part based on feedback from the CERN presentation) in August. The current draft document is available at http://www-unix.mcs.anl.gov/~schopf/pg-monitoring/ReqDoc/mon_req.v3.pdf.

2.3.1.3 Meetings and scope-issues

This group met formally as part of the GriPhyN meeting in April where the requirements document was discussed, and new editors were assigned. This meeting also resulted in the formation of a second mailing list to deal with non-Grid level monitoring issues (site monitoring or application monitoring) (ppdg-mug@ppdg.net). This focus of the ppdg-mug list has not yet been fully clarified, but there will be coordination between the two lists to make sure that no toes are stepped on.

2.3.1.4 Glue Schema work

The main effort by this group has been in coordination with the GLUE-schema effort (http://www.hicb.org/glue/glue-schema/schema.htm). In early March the JTB started a group to define, publish, and enable the use of common schemas for interoperability between the EU physics grid projects (focusing on EDG and DataTag) and the US physics Grid projects (focusing in on PPDG, GriPhyN and iVDGL). We are not proposing the adoption of any previously specified set of schemas, rather a new, unified schema is being developed; with the goal of having schemas defined for use in LDAP, SQL and XML.

This group had weekly calls which resulted in over 700 pieces of email, 14 versions of a comparison spreadsheet, and 9 versions of a structure document all focused on the definition of the needed schemas for compute elements. This work is within epsilon of being completed, and implementations by both Globus/MDS and EDG/DataTag are expected to be completed by the end of July.

Discussions of the storage element schemas have begun, and will in part incorporate information from the CIM schema work. A full timeline is available from the web site.

In addition, a small testbed has been set up for the testing of these schemas. Ewa Deelman has been coordinating this effort on the US side. There will be 4 sites overall – individual nodes at ISI and ANL, an ATLAS CE running LSF and part of the CMS condor pool. A GIIS has been configured at ISI and is up and running. The test for the CE schemas is planned for the second week of August.

2.3.1.5 References

Please note monitoring work for individual groups are discussed in other sections of this report. The web page for the joint PPDG/GriPhyN/iVDGL project is located at http://www.mcs.anl.gov/~jms/pg-monitoring/.

2.3.2 Globus MDS work

Globus members have taken a leadership role in the GLUE-schema work that is defining a joint-schema with DataTag and EDG for interoperability. More details on this work is listed in the section for the joint PPDG/iVDGL/GriPhyN Monitoring project.

There will be a beta release of MDS 2.2 by July 31, 2002 that will include much added stability and tracing/debugging additions and support for the GLUE-schema for the current information providers. In addition, this release will have new providers for GridFTP log data, TRU64, AIX and MDS components. There will be bug fixes for the mutual authentication of a GIIS-GIIS-GRIS setup, and a correct build process on AIX, as well as more functional and performance testing.

2.3.3 Collaboration with IEPM, Network Performance Monitoring

Contact: Les Cottrell

2.3.3.1 Measurement toolkit extensions

We have added Web100 to the measurement toolkit, and have selected an initial set of Web100 parameters to record. We are working on comparing results from Web100 throughput reports with those from the applications and from passive (NetFlow) measurements. We have also looked at the impact of high throughput on smoothed round-trip reported by Web100 and reported on this in a couple of presentations.

We have extended the prediction tools to incorporate exponentially weighted moving averages. It appears to make little, if any improvement compared to an un-weighted moving average.

We have been measuring and analyzing disk system performance to try and understand the effects of file system, caching, and committing writes, see Disk Throughputs.

2.3.3.2 Measurement tools (sensors)

Working with Richard Hughes-Jones, we have selected reasonable parameters for, and incorporated UDPmon into a test version of the toolkit. We are also separately making measurements with GridFTP,

though incorporating into the toolkit requires more work on certificates. We are working on comparing the results from these two sensors with iperf, bbcp, and bbftp

We are working on extending the InCITE topology/tomography tools to make them more portable (remove MatLab). To assist in this we have a PhD student from Rice visiting SLAC for the summer. Jiri is working to provide improved topology/tomography visualization tools and incorporating them with the traceping measurements (see Internet Tomography). We hope this will provide improved ways to visualize the network links on which the Grid is built. Warren also modified traceping to increase its robustness.

We have incorporated modifications to PingER from NASA and EDG into the standard distribution.

2.3.3.3 Setting up new Sites

We successfully ported the measurement toolkit from Solaris to Linux. The main challenges were related to the different ways the two operating systems treat multiple tasks in Perl. Following this we worked with Richard Hughes-Jones, of the European Data Grid (EDG) who visited SLAC for a week, to port the toolkit and start the measurements on a host at Manchester University. As part of this we parameterized the features that were measurement-host dependent and created a configuration file for each monitoring host to store the parameters. We also carefully documented the steps needed and built tools to enable automating the port procedures. Following this we successfully used the nascent tools to port and bring up the toolkit to FNAL, working with Frank Nagy of FNAL. An extra challenge at FNAL was successfully integrating the toolkit to work with FNAL's strong authentication. FNAL is now making measurements to a single host, extracting the data, analyzing and reporting the results etc. via the web. The next steps are to further improve the porting tools and the toolkit, automatically upgrade the toolkits at FNAL & Manchester, and work with Manchester and FNAL to set up sets of remotely monitored hosts for each site.

We have added Manchester University, NASA to the remote sites monitored by IEPM-BW from SLAC. For Manchester we measured over 460 Mbits/s, see Bulk Throughput - Manchester University.

2.3.3.4 Data Availability

We have installed and brought up Globus MDS and are working to install the PingER data. We are working on a web interface to provide access to the XIWT/CNRI/Telcordia/IPEX passive monitoring data.

2.3.3.5 Papers, Proposals, Meetings, Presentations

Andy Hanushevsky and Les Cottrell submitted a paper on data compression, and Les and Connie submitted a paper on the new IEPM-BW measurement infrastructure.

Les visited Romania at the invitation of the Romanian Ministry of Telecommunications and Information Technology and the Ministry og Education of Research. He was part of a team from the west (including Harvey Newman, Ian Foster and David Williams of CERN) to provide assistance for "The Romanian Potential regarding Grid Activities and Distributed Computing". Presented talk on Worldwide Network Performance and Monitoring. On the way to & from Romania, Les met with Peter Clarke of UCL and Richard Hughes-Jones of Manchester to go over EDG activities and see how best to coordinate our efforts (see trip report).

We are putting together a draft of the Maggie (Measurement and Analysis for the Global Grid and Internet End-to-end Performance) proposal. The initial deadline has been postponed and the proposal will be submitted in August for consideration for FY03.

The following talks were presented:

- Overview of IEPM-BW - Bandwidth Testing of Bulk Data Transfer Tools, presented by Connie Logg at the May 2002 Internet2 Conference in Arlington, VA.
- Comparing Throughputs as Measured by Active and Passive Methods, presented by Connie Logg at the May 2002 Internet2 Conference in Arlington, VA.
- International Connectivity and Performance for the HENP Community, presented by Connie Logg at the May 2002 Internet2 Conference in Arlington, VA.
- INCITE Review, presented by Les Cottrell at INCITE Review at Santa Fe, May 2002.

- Experiences and results from implementing the QBone Scavenger, presented by Les Cottrell at the CENIC meeting May '02 in San Diego.

2.3.4 Monitoring Frameworks

2.3.4.1 SRM File transfer Monitoring Tool (FMT)

1. We completed the development of the FMT as a web-based tool. This tool presents in a graphical form the status of each file requested in the multi-file request. The possible status is: transfer completed, transfer completed and the file released, transfer in progress, and trnasfer did not start yet. The graph gets updated every few seconds, so one can visually track progress of files in transfer. The tool also provides statistics on the average transfer rate per file.

2. FMT was installed at BNL and PDSF. Plans for it use in the next quarter were discussed.

2.4 CS-4 Storage Management

2.4.1 STAR-DDM (STAR-LBNL/SDM)

In this quarter a number of problems were solved that enabled HRM technology to finally go into service as a production tool for the STAR experiment. The main problem of GridFTP transfers "hanging" was solved with the release of Globus 2.0 in April. The other problems were solved by reconfiguration and upgrades to the grid nodes used for STAR data transfer. The result is that data transfers using HRM are very reliable and an overall transfer rate of 5 MB/s is typical. Future work will involve improving the transfer rate and development of more sophisticated tools to interface to the STAR file catalog.

To monitor file transfer a new tool was released by the SDM group called the File transfer Monitoring Tool (FMT). This web-based tool presents in a graphical form the status of each file requested in a multi-file request. The graph gets upgraded every few seconds so one can visually track file transfers and also provides statistics including the fraction of the request completed and the average transfer rate for each file.

Other efforts during this quarter included meetings and discussions regarding distributed computing in STAR. At this point we are working on the initial installation of VDT with the idea of evaluating Condor and GDMP for use in STAR.

2.4.2 GridFTP-HRM

The software that interfaces gridFTP to HRM for the purpose of enabling gridFTP access to HPSS was completed, and tested thoroughly. This componenet intercepts a gridFTP request for both get and put by a small modification to the gridFTP deamon. It then invokes HRM to get the requested file from HPSS, and then lets the gridFTP complete the transaction. Similary, for "put", it lets the gridFTP deamon deposit the file in HRM's disk, and then HRM schedules a transfer into HPSS.

2.4.3 HRM use of bbftp

For comparison purposes as well as accommodating clients that prefer to use various transfer protocols, we developed a version of DRM/HRM that invokes bbftp as the transfer protocol. It takes advantage of large windows capability, as well as parallel streams. This demonstrated the flexibility of DRM/HRM to use multiple protocols. We did not implement this as a dynamic feature. It is a parameter set in the config file.

2.5 CS-5 Reliable File Transfer

2.5.1 Globus RFT

The Reliable File Transfer Service (RFT) will be Globus's first OGSA service, completely up to the current specification. There will be a version demonstrated at GGF5/HPDC (July 2002), but it is a work in progress. It currently is built on top of the GSI security, and allows single-file transfer only (URL to URL), although there are plans to develop a separate queue service. It does not have the knobs to allow for parallelism or changing the buffer size, but these will be added in before an alpha-release is made.

The Globus project has been interacting strongly with the SRM project, and will include JLAB and Fermi/D0/SAM in future discussions.

An alpha-release, and solicitation of alpha testers, is planned for Fall. Additional information is available at http://www-unix.mcs.anl.gov/~madduri/RFT.html

2.6 CS-6 Robust Replication

2.6.1 GDMP

In this quarter, main emphasis was put on the first internal GDMP review that took place on 23 April 2002 at Fermilab. The review was very successful and the results can be found at:

http://cmsdoc.cern.ch/cms/grid/docs/ppdg_internal_review-gdmp-final.pdf

Shahzad Muzaffar has left the GDMP team which is a major cut down in man power for development work. However, the GDMP team does not foresee much development anymore except a possible integration with the Replica Location Service (RLS). Support is now a major task and the Condor team now officially supports GDMP related issues on the U.S. side of the project.

In the EDG testbed several tests have been done with GDMP 3.0-x and a few bugs have been identified and fixed by the GDMP team. The latest release is GDMP 3.0-8 and all GDMP 3.0 versions are compatible.

2.6.2 BaBar Database Replication (BaBar-SRB)

The SRB collection management system now supports authentication using the GSI security infrastructure for the BaBar experiment. This required interoperation with afs tokens to use local disk to store some of the SRB files. After further testing, the SRB instance for the BaBar experiment will switch to using GSI authentication in production. A PPDG certificate authority was used to issue the GSI certificates. Documentation has been written for the procedure that was followed to set up the SRB environment with GSI.

A SRB instance has been installed at in2p3 on a Solaris 7 machine. A test has been done for reading/copying data between SLAC and in2p3 using the SRB Sget command from disk and hpss. In support of this effort, modifications were made to the SRB environment to providea new resource management command to remove an SRB resource. The SRB mechanisms for implementing a proxy operation to access HPSS at SLAC were explained in detail.

Coordination meetings between SLAC and SDSC are now being done on a bi-weekly basis, after the GSI enabled SRB server was successfully installed at SLAC.

Work is ongoing to complete the implementation of a BaBar specific set of RDBMS tables holding metadata necessary to load objectivity databases into a federation. BaBar is working on using these tables within the SRB to allow users to replicate/move BaBar specific collections. Some tests have been done with a partially complete set of tables.

2.6.3 Globus ISI RLS work

Development continues on the prototype Replica Location Service (RLS), a distributed service for maintaining mappings between logical file names and physical file locations. Support for a master copy attribute (requested by PPDG) was added. Functional testing by both the Globus group and EDG WP2 took place. The API has been adapted slightly, and performance measurements are continuing. These results will be presented at SuperComputing'02.

An alpha release for general testing was made available in June 2002. The code will be available as a standard part of the toolkit in the next quarter, and we will more aggressively be looking for beta-testers at that time. Previously, we approached Wensheng Deng, LBL (Atlas, Magda) about alpha testing the current code but this has not moved forward.

Our plans for the next quarter include encouraging the testing of this code with other application groups (at least CMS and ATLAS, and possibly a third), with the thought that after testing we can determine whether we need to modify the update algorithms between the local and remote index nodes. In addition we will evaluate the performance of the current scheme - we may need to add some compression. After several months of testing, code improvements, etc. we will encourage people to switch over from the old code Replica Catalog code to the new, but the time-frame for this is about 4 to 6 months in the future.

Additional information is available at http://www.isi.edu/~annc/RLS.html

2.6.4 JLAB-Replication (JLAB-SRB)

JLAB and San Diego have reached agreement on the initial set of web services that will be created for demonstrating access to data resources at both sites. The basic approach is to implement a data grid interface that integrates a SOAP interface on top of "file" transport protocols. The integration effort had to recognize grid infrastructure support for authentication and authorization, correctly use the underlying logical name space for digital entity names, specify the web services that would be provided at each site, and agree on use of client libraries. Each of these areas required agreement on the approach, which is outlined below: Note that JLAB and SDSC represent two separate virtual organizations (VOs).

1. Authentication: user & virtual organization (VO)
Users can belong to multiple virtual organizations. Some web services will be able to handle multiple VO's concurrently. Thus the SRB collection-ID represents a VO.
a) By default, users will authenticate using https with X.509 certificate. This will say who he is. Initially JLAB can do this with unmodified web servers, and slightly extended java classes to help gain access to the certificate by which the user connected. The SRB already supports this capability.
b) Virtual Organization will be an optional argument (string) of each service invocation which needs it, and will default to either the default VO of the service (e.g. a single VO service), or a default for that user (maintained elsewhere, not specified here) or remembered from the last call the user made (sticky default).
c) Proxy credentials will be passed by either:
option 1: passed as an optional argument (safe, since encrypted by https) -- this is currently how JLab is doing proxies
option 2: passed by GSI using a GSI enabled/modified web server or other modified SOAP server

2. Authorization:
This is not specified by the OGSA web service definitions (yet). The service will have some mechanism for deciding what privileges a user has. Example: user is mapped to a local unix account, and access control is done via unix user, group conventions. Example 2: service maintains a database of access control lists per digital entity, as currently supported by the SRB.

3. Data Model:
a) Network Objects are defined by a Replica Catalog (service knowing where files can be found), a Site (node on the data grid having web services), and a File Server (source or sink of data bytes, typically one protocol per instance). The SRB will function as both replica catalog and site.
b) File Objects will be accessed by one of three possible naming conventions:

i) GFN = global file name, a globally unique string (within a single VO)
ii) SURL = Site URL, a concatenation of a network reference to a site (path to the SOAP service) and the string by which a site knows a file. JLab has adopted a format for the SURL in which these two parts are separated by a double colon (::)
iii) TURL = Transfer URL, a concatenation of a protocol specification, a host and port, and a file path as known by that file daemon; this is a standard web pointer like ftp://ftp.jlab.org:5555/path/filename

4. Web Services
We will first target file copying / replicating services. The implementation steps are:
Step 1: identical WSDL servers for a replicate command (internal to one system, make a copy of a file server-to-server)
Step 2: 3rd party copy between dissimilar servers using the same WSDL as for step 1.
2A: command is given to SRB, remote server is JASMine, SRB treats this as a copy to/from operation. If the SRB can register the remote file into a SRB collection, then at the end SRB knows about both copies of the file.
2B: command is given to JASMine's replication service, and the remote server is SRB. JASMine negotiates protocol with SRB (probably using gridFTP), transfers the file, and at the end JASMine knows about both copies of the file (since it's replica catalog only keeps references to files as SURLs, it can keep the reference to the SRB file even if it is not managing the file).

5. Client Libraries
We agree that there should be client libraries which in some cases wrap multiple network operations into a simple form for higher level code. No specification has been done yet, except that there should be a way of opening the byte stream for a data set specifying only the GFN (library finds an instance, gets access to it, opens stream, deals with protocol issues, etc.). This is equivalent to defining a storage repository abstraction for the set of operations that will be supported at remote sites. The SRB provides one abstraction. The XIO remote file system under implementation within Globus will support a subset of the SRB remote operations.

2.6.5 SRB based Replication Services

This quarter we have focused on providing broader access to the SRB technology within the high energy physics community. With the installation of the SRB data handling technology at in2p3, we have started interactions with additional sites in Europe. In particular, through the NSF GriPhyN project, we supported Frank van Lingen at SDSC for 2 months to learn the SRB technology. Frank has returned to CERN and given talks on the SRB capabilities to the high energy physics community. One immediate reaction has been the desire to test replication performance between the GDMP system and the SRB system. Of great interest is an identification of the technology differences that cause performance differences between the two systems. The hope is that such a comparison can lead to a better solution for replication execution, and determine the impact of wide area latency management mechanisms. The contacts are Asif Jan at CERN and David Jones, a PhD. student from the University of Liverpool. Both are participating in the CMS experiment.

The use test case of interest is to move 26,000 files that have a size of 20 MB each from one storage site to another. Currently these files are stored in HPSS at CERN. Replication of such a large amount of files created some difficulties with GDMP. The successor of GDMP (called reptor) will not be available until the end of this year.

2.7 CS-7 Documentation

Document below are posted at http://www.ppdg.net/docs/documents_and_information.htm.

Reports, Documents and Papers		Date/Version
PPDG-18	PPDG Year 2 Plan	6/07
PPDG-17	Job Scheduling and Management Focus Meeting Report	5/02
PPDG-16	iVDGL WorkPlans	
PPDG-15	Reports of the Internal Reviews and Action Items	4/02

Talks and presentations:

Presentations & Publications	
July 2002	Talk to BaBarGrid
June 2002	Talks at DOE: Richard, Miron, Harvey, Ruth; Talks at INFN Grid Workshop: Miron, Ruth (Trillium, Glue); Talks at CMS Grid Production Workshop. Dougs talk at ACAT
May 2002	Talk at Atlas S/W workshop
April 2002	IEEE Mass Storage Meeting - paper from Arie
March 2002	SDM-ISIC meeting and Talk by Doug; Talk at EDG Workshop

2.8 CS-8 Evaluations and Research

The work done by the PPDG teams over the part year has coalesced the understanding of the end to end and experiment grid needs. The focus has moved from evaluation and research to work in the common services areas and integration and deployment of "working" grids.

2.9 CS-9 Security, Authentication, Authorization, Accounting

2.9.1 Certificate/Registration Authority

The DOE Science Grid Certificate Authority (CA) underwent some small changes and improvements that will favor long-term stability. The most noticeable change is use of the domain name www.doegrids.org as the public location of the CA. This change is to reflect the fact that this CA serves a broader community than just the DOE Science Grid Collaboratory Pilot, which uses www.doesciencegrid.org for it's project URL. The role and definition of the Policy Management Authority (PMA) is becoming clarified and a document that serves as it's charter is in draft and undergoing iteration.

There is current work to establish an ldap directory which holds all of the certificates issued by the CA and serves as the definitive source for public access to these certificates. Details of how this directory is

organized are relevant to other work (under the Site-AAA area) and there are some discussions related to this. A test version of this publishing directory is available now and the production version should be in service in a few weeks.

Additional Registration Authorities joined this CA in this quarter where the most significant (from a PPDG point of view) is an RA for iVDGL (see www.ivdgl.org). This is important due to the role of iVDGL serving a largely overlapping applications community with PPDG and it's mission to establish a true international virtual data grid laboratory. With help and encouragement from PPDG, and endorsement from DOE, the iVDGL was able to establish an RA with the DOEGrids.org CA and it is now beginning operation.

There are some minor issues that develop and get resolved and some issues which are more significant. Probably the most significant issue is on how Certificate Revocation Lists (CRL's) are handled. The current Globus GSI infrastructure has only a token acknowledgement that one should be able to check the validity of a certificate but the infrastructure for managing CRL's is inadequate. It may well be that CRL's are not the best solution to the problem. At present, with some manual intervention on the part of the CA managers, up to date CRL's are published and available to the sites which need them.

2.9.2 Site-AAA

The Site-AAA proposal received funding this quarter and has begun activities at all 5 of the proposing sites: BNL, FNAL, LBNL/NERSC, SLAC, TJNAF. Details of the activities in this cross-cut project (including detailed site reports for this quarter) are described at http://www.ppdg.net/pa/ppdg-pa/siteaa/ and a mailing list[1] with web archive is established to support communications for this effort. In addition to the laboratories, there are participants in this effort from the Condor and Globus projects to ensure effective coupling between the middleware developers and the site security administrators. Conference call meetings are held every two weeks to report progress and current activities, explore common interests in solutions/issues and discuss proposed solutions.

Contacts with the corresponding efforts in Europe in the DataTAG and EU-DataGrid projects has been established. Discussions at a BOF at the GGF5 in Edinburgh July 23 will explore the possible creation of a Grid Forum Research Group to develop a site infrastructure requirements document jointly. A working list of issues identified as problems for operations of large sites is posted on the website and reviewed monthly for change in status and understanding. All sites have begun dialogues with their management and local computer security experts to begin identifying policy concerns and issues.

All sites have identified local integration projects for gaining expertise with the Globus toolkit and to inform comments. Three Sites (FNAL, JLAB, and SLAC) have begun integration of their mass storage systems with the Globus file transfer mechanism (GRIDftp) with FNAL and JLAB reporting functional prototypes. Two sites (LBL and BNL) have identified management of user groups and accounts (the VO membership structures) as projects and have begun their efforts. LBL is deploying the EDG tools and working on integration with their internal system and BNL has begun definition of a tool for coordinated account creation and tracking. FNAL has deployed the KCA system from the NMI project and is working integration issues between its internal Kerberos infrastructure and a PKI based GRID.

A major issue identified by this project was the user based, sitewide authorization decisions required by sites were not possible with initial Globus designs. The Globus Community Authorization Service is being redesigned to address this issue. The GRAM protocol is being revised as GRAM2 for GT3 release and expected to address some of the issues as well. Consensus is this is the key authorization issue from the site perspective.

2.9.3 VO Management

In order to manage authorization in the Grid computing model, the concept of Virtual Organizations was developed (by others). This involves placing users in different groups or categories and managing access to certain resources on the Grid based on these groups, instead of on an indivudual basis, thereby reducing

[1] http://www.ppdg.net/mailman/listinfo/ppdg-siteaa

the load on site administrators. To aid this process a tool to manage VOs was developed that provides a graphical user interface to the user, group and certificate information, as opposed to the cumbersome collection of scripts that are currently used. This tool has a web page at http://heppc22.hep.caltech.edu/groupman.

2.9.4 Globus Site-AAA work

ANL is actively participating in the PPDG Site-AAA group, including bi-weekly calls and hosting of the group's first meeting at ANL on April 26th. Based off of the feedback we have solicited from this group, we have redesigned our Community Authorization System (CAS) to better meet their needs (as described in the following section on CAS).

Von Welch has taken a leadership role in defining the PPDG security plans via the PPDG-AAA group and trying to help guide them also lines that are compatible with other Grid deployments like DOE SG. In addition, Von Welch and Steve Tuecke has helped plan and coordinate the group's efforts to present their material at the upcoming Global Grid Forum in Edinburgh (GGF-5), one of the milestones for this group.

2.9.5 Globus CAS

Based on feedback from the PPDG Site-AAA group, we have changed the design in our CAS prototype code to allow a relying site to be able to authenticate the connecting user itself. In the initial prototype, alphaR1, the CAS server authenticated the user and passed the identity to the relying site. This redesign will appear in the next release of the CAS prototype, alphaR2, and is expected to be available by the middle of July. At that time we will actively seek alpha testing.

We also produced a modified version of CAS to work on the CMS testbed at the end of March. This required modifying CAS to work with Python version 1.5 (which is what comes on RedHat 6.2 as run on the testbed). This was given to CMS for deployment, though it's unclear that any work has progressed since then. Future releases of CAS will support Python v1.5 so that CMS can use them without modification.

Additional information on CAS can be found at http://www.globus.org/Security/CAS/

2.10 CS-10 Experiment Grids and Applications

2.10.1 ATLAS

2.10.1.1 ATLAS distributed data manager, MAGDA (ATLAS-Globus)

The principal goals of the Magda project for this period were participation in the ATLAS Data Challenge 1 (DC1) and significantly improving the speed of data transfer between BNL and CERN. At the same time, Magda developers actively participated in preparation for the US ATLAS July testbed production and for the Super Computing 2002 (SC2002) demos.

The web interface of Magda was enriched during this period.

1. A web interface for filling collections was developed. Users choose a collection from a list to fill. The collection can be filled with the content of a selected location, or all logical files matching specified keys.
2. Work on a web interface for deletion has been started. At the moment, users are allowed to delete all file records at a specified site location.
3. Web file listings can now be sorted by subdirectory.
4. Users can activate and deactivate replication tasks from a password-protected web page.
5. Users creating a file location with the web interface can now specify whether or not spider should scan that location.

The data replication part of Magda continues growing. It is now forbidden for two active replication tasks to share the same input collection and/or caches. Data replication has been extended to support disk-rftp and disk-castor transfers. Replication of dynamic collections of files now supports the BNL HPSS source. A replication task can be started even before the production jobs are submitted. This greatly increases the efficiency of using network bandwidth. A new conduit was open in the BNL firewall in order to run bbftp in ssh mode. The bbftp protocol has been integrated into Magda and is heavily used for ATLAS DC1. The bbftp-armed Magda works quite stably when using local disks as caches one both ends from BNL to CERN. The US ATLAS ftp gateway aftpexp has been upgraded. With his DOE certificate, Wensheng Deng successfully ran globus-url-copy and to transfer data from BNL to CERN.

Much effort had been put into the Pacman distribution of Magda with the goal of simplifying the installation of Magda-related software. The distribution includes MySQL client, perl, perl add-on modules and Magda itself. Two versions of distribution were made and put in the Pacman BNL PAS cache. One requires root privilege. With a single Pacman command, users can smoothly install all components of the distribution and then easily run Magda after executing a Pacman-generated setup script. Pacman was enhanced in response to requirements that grew out of this effort.

Pacman is being used to deploy Magda at the eight US ATLAS testbed sites and will be cataloging and moving files in the upcoming testbed production. Magda will be a part of the US ATLAS testbed demos at SC2002.

Atlas collaborators at Milan are examining Magda and will be integrating it with GDMP. Oslo NorduGrid developers plan to evaluate Magda and run their own Magda server.

Further integration of Globus tools, particularly GridFTP and remote command execution for more flexible Magda usage, are part the short-term plan. The plan also includes investigation of a new logical entity ‘farm', distinct from 'site', to accommodate locations distributed across the local disks of a processor farm. This and the integration of integrate of Magda collections with datasets will facilitate the dataset-based distributed analysis being developed at BNL.

2.10.1.2 US ATLAS Grid Testbed

During the second quarter of 2002, the U.S. ATLAS GRID testbed concentrated on deployment of grid middleware and Monte Carlo production. In April, a workshop was organized at the University of Texas at Arlington. See http://heppc1.uta.edu/atlas/workshop_april_2002/index.html for agenda and transparencies. The command-line GRATS toolkit to run MC production jobs was released in mid April ater this workshop. This toolkit uses Globus version 2.0 to run athena-atlfast MC jobs on any grid-enabled or afs-enabled site. It was successfully tested on all eight ATLAS testbed sites, two CMS, two D0, and two EDG sites under Red Hat Linux versions 6.x-7.x.

The Core/Grid Workshop was at BNL in May (http://www.usatlas.bnl.gov/computing/software/core-grid-200205/) and the Grid workshop at Boston University in June (http://physics.bu.edu/%7Eyoussef/meetings/june_2002) were used to solidify plans for a completely automated Grid based MC production system on the Testbed.

During June, a lot of work was done to prepare the Pacman packages for automatic installation of required software on the testbed. In addition, various MDS based monitoring tools (Gstat, Pippy) were developed for QoS on the Grid. Both GRATS and Grappa continued to be improved in preparation for production. More details about the software and tools developed by the Testbed group can be found at http://heppc1.uta.edu/atlas/grid-testbed/index.htm.

2.10.1.3 EDG Testbed Interface

Jerry Gieraltowski at ANL continued his work with the EDG testbed. Evaluation of the testbed middleware focused on data handling. Several different atlfast jobs were submitted to the testbed and data was stored and replicated at multiple sites using the GDMP functionality. The following aspects of data handling were tested and bug reports generated where appropriate:

a) Data resource identification - the ability to map logical and physical file names to specific site locations for the referenced data. Assessment: identification works well but the BrokerInfo file created still has unresolved bugs

b) Data registration and publication - the ability to register a set of files on an SE node, publish its existence to a master replica catalog and replicate it to nodes on other sites in the test bed. Assessment: GDMP 2.1 was used for this effort and works well. Minor problems were resolved with the administration of the ATLAS/GDMP configuration files at the various Testbed sites.

Current Status of Bugs Reported: 4 -still open, 13-resolved

Work is in progress on a document presenting a "User's Guide for the ATLAS End User". It is intended to provide a concise and simple user's guide covering topics such as: "Getting Started", "Job Submission", "Data Management". See http://www.hep.anl.gov/gfg/anl-atlas/edginfo/userguide.

2.10.1.4 BNL tier 1 activities

BNL is the U.S. tier 1 center for ATLAS and, as such, is providing a large compute farm and extensive data storage capability for ATLAS in general and for U.S. ATLAS institutions in particular. A number of activities are underway to provide a GRID interface for the BNL center. Dantong Yu is carrying these out.

Over the last quarter, the following were carried out in the area of data replication:

1. Set up GridFTP server in BNL and deployed GridFTP client at CERN, LBNL and Oklahoma University.
2. Participation in the ATLAS Data Challenge with integration of the BNL-CERN data transfer into Grid environment.
3. Optimization of data transfer performance, plan on designing efficient data replication architecture. Some results can be found at:

   Bandwidth:

   http://www.acf.bnl.gov/UserInfo/Facilities/Grid/OC-12.pdf

   http://www.acf.bnl.gov/UserInfo/Facilities/Grid/OC-12.jpg

   Timing:

   http://www.acf.bnl.gov/UserInfo/Facilities/Grid/time.pdf

   http://www.acf.bnl.gov/UserInfo/Facilities/Grid/time.jpg

In the area of AAA (authentication, authorization and accounting), an ATLA VO (virtual organization) server was deployed at BNL. The VO server users are divided into groups. Tools were installed to enable the VO manager and group managers to add, remove and group certificates. The relevant URL is ldap://spider.usatlas.bnl.gov:6200/o=atlas,dc=ppdg-datagrid,dc=org Work is underway to adapt and web interface for the server. Discussions were begun on the problem of mapping the VO certificates to local accounts.

2.10.1.5 Data signature

There was no modification of the data signature classes during the quarter. The persistence part of the CERN LCG project started up and many ATLAS PPDG members are contributing and monitoring that project to ensure (among many other requirements) that the ATLAS data signatures will be supported. We have begun to integrate some of the data signature concepts with those of the virtual data catalog. See

http://www.usatlas.bnl.gov/~dladams/ppdg/talks/020506_virtualdata.ppt

2.10.1.6 Distributed analysis

A new project led by David Adams at BNL was initiated at BNL this quarter. The goal is to enable distributed analysis of large volumes of ATLAS event data where large means more than can be conveniently accessed directly from a single machine. We anticipate that data will b distributed over local disks on one or more farms and that the user will have a single entry point (e.g. using ROOT) that will communicate with remote GRID sites. Although there are many issues in common with batch-oriented production running, we are most interested in issues specific to interactive analysis such as prioritizing tasks, dedicated resources and the retrieval of partial results.

This project was first described in a talk at the BNL PPDG meeting last May. The talk may be found at

http://www.usatlas.bnl.gov/~dladams/ppdg/talks/020507_analysis.ppt

Another presentation was made at the ATLAS software workshop at the end of May. It can be found at

http://www.usatlas.bnl.gov/~dladams/ppdg/talks/020530_analysis.ppt

ATLAS is preparing to begin Data Challenge 1 (DC1) and we plan to use data from DC1 for our initial efforts. ATLAS has chosen a ROOT-based format for its Monte Carlo generator output files and we plan to begin with those files. We have learned how to read the files and have run some simple single-file analyses.

Most of our effort this quarter has been focused on defining and implementing means to specify and access data of interest. For this we introduce a composite dataset as described in another talk at the ATLAS software workshop at

http://www.usatlas.bnl.gov/~dladams/dataset/talks/020528_dataset.ppt

and in the dataset document which can be accessed from the dataset web page at

http://www.usatlas.bnl.gov/~dladams/dataset

That page also provides links to two new software packages: dataset which defines the base class Dataset and associated classes and the package root_dataset which provides a dataset implementation for reading and accessing the current ATLAS ROOT file format. We are able to extract Monte Carlo generator data from these datasets and are now working on an XML-based persistency mechanism for these and other types of datsets.

2.10.2 BaBar

Within BaBar we intend to use the SRB to handle the data distribution initially between Tier A centers. The metadata catalog, based on the SRB Mcat, is expected to be of wider use (e.g. finding out which databases belong to which collections, when the data were loaded into the analysis federations, the size of collections etc).

During this quarter we have been able to setup test SRB servers at in2p3, Lyon and at SLAC. We have carried out tests of moving databases from HPSS at SLAC to in2p3 using the native SRB commands. These tests have enabled us to understand some of the tools we need to build in order to achieve a production system (eg although we have found the Sget command to be useful, we would like to be able to incorperate other file transfer mechanisims, such as bbcp within Sget and we would also like to be able to wrap Sget to incorperate our integrity checking tools once the data has been transferred). We are in contact with the SRB developers who are helping us to build the tools that we need. The replication tests were done using the SRB's native security infrastructure.

Since these tests were done we have built the SRB servers with the Globus GSI libraries and have been able to use PKI certificates to access the data stored in the SRB. We have not yet carried out a test of moving data from SLAC to in2p3 through the SRB with this new system yet, however data access and movement tests within SLAC have worked.

All the SRB tests to date have been achieved using the SRB's Mcat holding file- specific information and a set of tables containing BaBar Objy database specific information. The BaBar tables were very rudimentary containing the minimal information needed to copy a collection. Since then, we have been working on designing a schema for the Objy database specific information that meets the needs of the Tier A file replication as well as the needs of BaBar users wishing to use the Objy database. We have captured practically all the use-cases, built several tests and expect to install the production tables complete with load and query applications before the end of the next quarter.

2.10.3 CMS

2.10.3.1 CMS Testbed

Stu Martin (a senior Globus developer) was assigned to help with CMS testing and bug fixes to CMS testbed, working half time from June to September. This has resulted in Globus having a much more active role in the CMS testbed in terms of triaging bugs and turn-around for basic questions resulting from the use of the testbed for the production runs. For example, the issue with scalability of the gass-cache has been resolved, and a plan of development agreed on.

2.10.3.2 Caltech

1) Eric Aslakson

Eric's activities during this period were divided between developing software for physics analysis and designing software for production activites related to scheduling. In the analysis software area, Eric completed work on a Root-Tree to SQL database insert program. He generalized and extended code supplied by Krzysztof Nienartowicz, of the IT group at CERN, to use Microsoft SqlServer as well as Oracle databases. He also extended the modules to map arbitrary Root group names into SQL table names, enabling conversion of (almost) arbitrary Root trees.

Due to design requirements that the code run unattended, Eric also converted batched database inserts to ODBC function call inserts. This allows more robust error recovery for unattended inserts. This ODBC layer also facilitates pluggable SQL database usage. He then started proving ODBC drivers on Linux for Oracle and SQLServer. He also began work on a standard template library based objectification of SQL result sets. This layer will facilitate front end caching and easy tranport of SQL results sets optimized for high latency network connections. This software will plug into Clarens for use in Grid Analysis Environment prototype.

In the production area, Eric continued to participate developments in the PPDG monitoring area by participating in the very few phone conferences occuring. He continued design of the Execution Priority Manager, a module working in concert with the DAG processors and grid nodes to attempt a guarantee of throughput promises for production activities. This work will continue in August when he will be working with the relevant groups at Fermilab to implement this module.

2) Edwin Soedarmadji

In this reporting period, Edwin concentrated on the tasks : (1) reverse-engineering a portion of JetMet analysis code produced by Pal Hidas while at FNAL (2) writing programs that convert JetMet PAW ntuple into raw text (3) building a SQLServer database and (4) import the converted ntuples into the database.

These tasks were recently completed and now serve as the foundation for the next set of tasks. The SQLServer database will be used in conjunction with an Oracle database at CERN (under construction) to serve out JetMet analysis data in a more efficient and granular way. Currently Edwin is working on a set of web services that would allow queries to be submitted and executed against the Jetmet database by a client software.

3) Saima Iqbal

Saima's main task is the evaluation of Oracle9i for the CMS Event store. Oracle9i implements an Object Relational Model. The object type maps closely to the class mechanisms found in C++ and Java, as a result the object based code that is already in use by CMS would not need be changed or rewritten. Only an SQL wrapper is required to store and access data from Oracle database.

Saima's evaluation started by investigating the Object Relational features of Oracle. She measured the overheads on the objects after storing them in the Oracle9i database. For these overhead measurements simple objects, embedded objects and objects with REF were created in the form of Object-Tables. These object tables were loaded with simple data from vectors and hits. In the Object-Table each C++ object corresponds to one row and each attribute maps to the column in the row of the table.

LOADING AND ACCESS OF TAGS DATA FROM ORACLE 9i DATABASE:

As a follow up activity Saima looked at the proposed architecture for a CMS event store using Oracle 9i and its related tools. A small-scale data warehouse prototype for CMS TAG data was developed by her. To evaluate the performance of Oracle for the persistency of Tags data the Java classes TagLoader.class and TagAccess.class developed at Caltech for SQLServer were converted for Oracle. Preliminary results show a data warehouse ingestion speed of 25 MB/s. The tuning of this process will improve over time.

ORACLE 9i WEB APPLICATION:

In the next step (loading and accessing the Tags Data) an Oracle base Web Application was developed by Saima. After the comparative study of web applications developed by using Microsoft FrontPage, Perl, CGI (Common Gateway Interface) and Java, the web application was designed in Java/JDBC (using Java Server Pages). It is platform independent and its byte codes can easily be run on any machine providing a JVM (Java Virtual Machine). It has an open structure to avoid restricting users to specific hardware and software configurations. It does not require the Oracle Client installation. The demo of this web application was presented by Saima during CMS Week June-2002.

ARCHITECTECTURE TO USE ORACLE 9i IN GRID ENVIRONMENT:

A proposed architecture makes use of Oracle 9i data warehousing in a Grid environment, enabling hybrid data access with Oracle9i and Microsoft SQL server. For this purpose Saima began work on loading the Jetmet Ntuples version 2.06 in the Oracle Server at CERN.

4) Suresh Man Singh

Suresh's main activities during the period were focused on upgrading hardware and software for Caltech's AMD cluster in the USCMS grid testbed. He upgraded the RAID software of the Network Attached Storage (NAS) unit of the cluster. The Linux kernels of all the cluster nodes were upgraded to 2.4.18 allowing improved performance. Various security related fixes or upgrades were made to vulnerable packages. These included the Apache web server, openssh 3.4 (secured shell), openssl, python etc. The time protocol NTP was upgraded on all the nodes and the head node was synchronized to CDT - as required by the distributed Monte-carlo Production (MOP).

Suresh installed both VDT 1.1 server and client on the cluster's head node (citcms.cacr.caltech.edu) which includes Globus 2.0. As part of the host, service and user certificate migration process from Globus signing to DOE Science Grid certificates, he obtained the DOESG signed host certificate for the cluster's head node and installed it. This head node was registered with University of Florida's GIIS MDS server and can be queried by Globus grid-info-search command. Suresh is in the process of upgrading the cluster's local batch scheduler package to the most recent release of Condor to 6.4. This will be used for distributed MOP production by providing a hook from the Globus jobmanager. At the moment only the "FORK" jobmanager of Globus is active.

5) Conrad Steenberg

Work on the Clarens web services infrastructure is described in section on CS-11 (2.11).

Work on VO Management (groupman) is described in the section on CS-9 (2.9).

6) Vladimir Litvin

Vladimir organised the production of 50k events for forward jet-tagging mechanism studies, and 25K events for calibration studies. He made a presentation at ACAT2002.

He prepared analysis software for July CMS milestone in the Egamma group. He analysed five Level1 collections for the DAQ TDR, and five full datasets for the pixel isolation studies. An importer for data files from CERN has been installed, and 0.8TBytes of data moved from CERN and analysed.

7) Julian Bunn

Planning for demonstrations at iGrid2002 and SC2002 is advancing, with discussions centering around which components of the Grid Enabled Analysis Environment will be ready, and which need to be "glued" together to make a sensible application. For iGrid2002 we are collaborating with Bill Allcock and others in the Globus team at Argonne. The intention is to show a ROOT front end client on the show floor, which

talks to Clarens servers at Caltech, Argonne, CERN and possibly UCLA and Rio. GridFTP will be used to move large data files generated on the server by Clarens to the client machine.

We want to show a truly heterogeneous GAE at iGrid2002 and SC2002, with Oracle9i storing JETMET data at CERN, SQLServer storing a complementary set of JETMET data at Caltech, and then perhaps more data in PostgreSQL or MySQL at UCLA and Rio. Bandwidth usage will need to be very heavy in order to interest participants at the conference, so we are adopting a somewhat artificial analysis scheme which involves the use of GridFTP to move files across the WAN. Perhaps at SC2002 we will use Clarens instead of GridFTP, if we can reproduce the streaming performance of GridFTP on the special striped servers being provided by the Globus team.

A particular item of work is on developing a database layer to which the Clarens server talks in SQL, and which is responsible for returning a vector of physics objects. Alternatively this layer would generate a file containing the objects, which may then be transferred using GridFTP.

We are committed to using the Globus authentication methods as part of the demonstrations, and as part of the GAE being developed. We will use DoE Science Grid certificates, which are being obtained for the servers and users involved.

As far as hardware for the demonstrations is concerned, we have spent some time analysing the market for the best motherboard to use to maximise the Gbit NIC performance of a server. With help from Ian Fisk (UCSD) we have ordered two powerful servers based on the SuperMicro P4DP6 board, each equipped with dual SysKonnect SK9843 optical NICs. Tests of the network performance in back-to-back LAN configurations are proceeding. We also have acquired Dlink Gbit NICs for testing. A further server has been ordered, based on the SuperMicro P4DP8-G2 motherboard, featuring on-board dual Gbit NICs (copper). This server will also be tested. Finally, once a suitable configuration has been proven to allow sustained high bandwidth transfers in the LAN, we will purchase several for placement at CERN, Caltech and at the STARLIGHT in Chicago.

Work on optimising the WAN performance of the GAE is an important aspect of our activities, and we have started a very fruitful collaboration with Steven Low's (Caltech) group on Adaptive Queue Managament for TCP, which promises much better performance for large data transfers in the WAN. This work is being partially funded by a successful proposal we submitted to NSF.

2.10.4 D0

Work continues to flesh out the software needed to complete the proposed SAM-Grid architecture. Progress has been slow while Gabriele was devoted almost full time to CDF. Work is ongoing with the Condor team to enable the Match Making Service they are providing and integrate it into the system. This work has been delayed due to Condor collaborators' delays with the Condor development release, as well as contributions to other D0 and CDF sam related tasks. Towards the end of June, we received the Condor release implementing the two changes in the classAds and in the Match-Making service. We have started testing the changes, employing a graduate student and a fraction of an undergraduate student, resulting in a successful demonstration. The gatekeeper for a job has been selected by the MMSusing the information from an externally provided script. There are a number of outstanding technical issues which we need resolved by the Condor team.

Looking at the larger picture our proposed SAM-Grid architecture is composed by 3 major modules: 1) The Data Handling System (SAM),2) The Job Management, and 3) The Monitoring and Information Services. While concentrating on the area of the job management, we are also exploring MDS and other needed middleware for the information services. SAM will be used for the Data Handling system with some modifications.

Igor and Gabriele presented at ACAT2002 in Russia. Igor’s presentation was a plenary talk covering data handling at D0, and included many details of D0, SAM, and SAM-Grid. Gabriele’s talk concentrated on the SAM-Grid architecture, plans and progress. We are also preparing for SC2002 display in which we hope to highlight much of the SAM and SAM-Grid work. The display will be Grid oriented, and is being done in cooperation with CMS, BaBar, CDF, D0.Of course, Condor is a major part of our Job Management architecture and they will be represented too.

Much of our time over the last 3 months has been spent helping CDF evaluate SAM and begin to integrate it into their data handling system. Gabriele has been diverted almost entirely away from the Grid efforts to the CDF project and this has postponed many of our anticipated goals. Lee has also spent a large amount of time working with CDF to coordinate the SAM project to include their needs, as well as D0's.

Igor and Gabriele have spent a large amount of time training 4 students who are working with us for the summer. Two of these are graduate students from the University of Texas Arlington who were hired through the D0 collaboration with the HEP department there. These students have been involved with our Job management, and monitoring and information logging projects. We hope to continue them on beyond the summer as they have become very useful to the program.

Les has been working to enable additional monitoring for the network data transfers for the existing D0 system. We are working with the SLAC networking monitoring group (Les Cotrell), and using the software tools they have developed for the IEPM project. Our networking department (DCD) has deployed a monitoring node at FNAL. We have begun to deploy clients at D0 processing and analysis sites to better understand the network capability of our growing system. At the same time, we are developing better tools to monitor the actual data flows throughout the deployed base of D0 SAM stations worldwide. This later work is anticipated to be used to generate displays for one panel in the SC2002 presentation, mentioned above, being prepared jointly between FNAL and SLAC.

D0 has successfully demonstrated D0 MC job submission (the entire chain of the job, starting from generator, geant, noise addition, reconstruction, and reco-analysis, except for logging to the SAM system) and monitoring at the July workshop in Oklahoma university. This was one of the primary milestone goals that UTA was assigned to accomplish as part of the D0 Grid effort.

In that demo, a few D0 MC jobs have been submitted from UTA central farm site to remote farm at OU where the scheduling was controlled by the UTA generic MC farm control software. The two monitoring software were a generic UTA monitoring through globus job submission and an MDS based resource monitoring, called McView. These monitoring software were developed by Tomasz Wlodek whose new affiliation is BNL, as of Sept. 1.

The same job submission script has been implemented on one of the machines at FNAL and has been tested successfully, submitting a D0 MC job to UTA farm from FNAL.

2.10.5 Jlab

In this quarter Jefferson Lab has continued its developments on two fronts. First, it has added a web services interface to JASMine which is consistent with the SRM specification developed with LBL and Fermilab. Second, it has begun prototyping policy based file replication, using a web services replicated data grid which includes additional SRM functionality beyond those currently specified by the SRM document.

For the first task, the Globus Java Cog, along with some code from the recent Globus OGSA-TP-2 release, was used to provide a web services environment that suppports GSI authentication for the JASMine SRM implementation. [JASMine is Jefferson Lab's java based storage management software, supporting an STK silo and disk pools.] Work with Fermilab and LBL has continued toward the goal of interoperable SRMs from distinct code bases through web services.

This system has been tested on a prototype grid between Jefferson Lab and Florida State University, where batch jobs submitted at FSU used the SRM to fetch remote data. Initial results are promising.

For the second task, a first prototype of policy-based file replication was implemented as server processes which are clients of the replicated data grid. These prototype clients serve two purposes: (1) they enable us to test the functionality contained in the ReplicaCatalog and an SRM, and (2) they enable us to deploy useful services to a small production environment.

This second project uses an SRM implementation which is independent of a back-end storage management system. This SRM (JSRM, for Java SRM) supports access to configured portions of a conventional file system (based upon an XML configuration file and upon user access permissions). These conventional areas are referred to as "unmanaged" file systems in that there is no disk cache manager in use. JSRM can

also support a back-end storage management system such as JASMine or any other SRM compliant system. (Testing of a full integration of JSRM and JASMine should be done in the next 6 months).

For policy based replication, we have developed a spider base class (in java) that walks a directory tree looking for files to work on. A derived class enters into the replica catalog any new file (defined as not in the replica catalog), and requests the file to be transferred to another grid site according to a configuration file. In this version, the spider uses file path mapping from source, to replica catalog, to remote site. A user or batch job only needs to drop a file into a directory being watched by the spider, and the transfer and publish will be done automatically. Another version of the spider (currently used on the receiving end), pushes files into the JLab silo, and also enters replica info into the replica catalog.

We have also developed a publish utility (command line) for cluster users to automatically publish and archive data produced by a batch job. This utility works with another spider variant to complete the task.

In each of these cases, the spider does the registration and file transfer requests using its own credentials, removing the need for un-expired user proxy credentials. In order to preserve the correct ownership of the entries in the replica catalog, we have implemented extended security features in the replica catalog, allowing privileged users (known to the replica catalog) to make entries owned by other entities. In this way we have the notion of grid resident processes which are analogous to root processes on a local system, but with privileges circumscribed by the web services being used.

2.10.6 STAR

The STAR work on reliable file transfer and storage management is described in section 2.4.

There is also work beginning on job scheduling with current plans described at http://www.star.bnl.gov/~carcassi/scheduler/dev/plan.html. The goals of this work are:

- Provide a tool usable for year 2002-2003.
- Freeze the user interface for job submission for the years to come.
- Define a software architecture that allows migration to other GRID tools.
- Provide user interface for scheduler policy management.

There is also work started in the area of distributed databases, that are needed now for STAR and are a necessary part of the capability for distributing compute jobs on the grid. This work is described at http://mysqlforgrid.sourceforge.net/.

2.11 CS-11 Grid Interface with Interactive Analysis Tools

This area on Interfacing and Integrating Interactive Analysis Tools with the Grid and Identifying Common Components and Services was launched in this quarter. A workshop was held in Berkeley in June. Agenda and presentations are available at http://www.ppdg.net/mtgs/18jun02-lbl/agenda.htm. A summary presentation of this workshop was presented at the Advance Computing and Analysis Techniques Workshop in Moscow in July (http://www.ppdg.net/pa/ppdg-pa/idat/olson-acat2002-cs11.pdf).

The initial goals of this area in PPDG are to:

- assemble people from the experiments and the grid middleware groups who are working on defining, integrating or developing interactive data analysis tools and the grid services that these tools require
- assemble reading list of relevant papers, web sites, etc.
- produce white paper
    - describe each experiments needs

  - summarize common or aggregate needs
  - evaluate middleware w.r.t. needs
- look at examples of tools that interface to the grid
- develop a roadmap for grid-based interactive data analysis.

Given the early stage of development of production grid services for end-users it will clearly be necessary for a few prototype development projects to explore and demonstrate to physicists the opportunities and possibilities of the grid regarding interactive analysis.

A document describing the first look at requirements is under development and will be available from the PPDG web pages when ready.

2.11.1 Clarens

Work on the Clarens web services infrastructure continues, with improvements in robustness and functionality. Client functionality was expanded, with a ROOT remote file access method and an ftp-like command like file download client recently being the main highlights. The Clarens project has a new development home where a file repository (CVS), web page, and mailing list functionality is available in order to enable wider collaboration. Work started on developing and deploying Virtual Organization (VO) management tools to ease Grid site security management.

1.Presentations

- "Clarens remote analysis enabling environment"CMS Tutorial week, May 1 - 4, UC San Diego

- "LDAP CA Authorization management" PPDG/Griphyn All hands meeting, April 23 - 25, Argonne NL,

2.Project management

The Clarens project home page was moved to http://clarens.sourceforge.net which provides managed web hosting to open source projects. A project mangement interface can be found at http://sourceforge.net/projects/clarens/, with access to bug tracking, a versioned file repository, and mailing lists.

3.Tool development and software changes

- Clarens services added

Added file download and browsing service added. Since most forms of data are still stored in files, this service is a logical basic compoenent of Clarens. The service streams data from a file to the client in raw form, or provides file listings for the directory structure of the file system. User files may be published from a special directory created in the user's login directory, and accessed using the well known tilde notation (~user/).

- Service description publication

Automatically generated descriptions and API documentation is provided using the system.Listmethods, system.methodHelp, and system.methodSignature API calls

- Clarens client side

Two clients were implemented for the file access service, a command-line ftp-like interface, as well as enabling the ROOT analysis package to access remote files tranparently

3 Single Collaborator Reports

3.1 ANL – Globus

Most of the Globus – PPDG related work is described in the sections above on CS-2, CS-3, CS-5, CS-9 and CS-10.

3.1.1 Coordination and Support

One of the main contributions to the PPDG project by the Globus team has been the continued interactions and support of the individual applications. One factor in this has been the upgrade of the Globus bug-tacking system to Bugzilla (http://bugzilla.globus.org/bugzilla/), which allows you to select the PPDG project, allowing the PPDG globus liason (Jennifer Schopf) to be alerted to this bug and to easily track it.

Additional interactions in terms of coordination and support of the PPDG applications included, but are not limited to:

General PPDG support

- Participation in general PPDG meetings, including
    - Scheduling Workshop, Fermi Lab, May
    - Strong participation in internal reviews process
- Meetings with Brian Tierney and D0 monitoring group, Fermi, March
- Alliance HEP talk for PPDG, GriPhyN and iVDGL, UIUC, May
- Allcock visits to SRM and SRB, California, June
- Plans to meet with D0/CDF in early August
- Architecture planning calls with Arie, Miron

ATLAS- continuing support

- Weekly atlas testbed calls, bi-weekly atlas sw calls
- Atlas SW week, CERN, March,
- US ATLAS Core/Grid Software Workshop, BNL- may
- US Atlas Grid Workshop, BU – June
- UC/ANL Atlas Software and Computing meeting, July

CMS- increased effort and involvement this quarter

- Stu Martin (senior Globus developer) assigned to help with CMS testbed (see below)
- (Almost) weekly CMS testbed calls
- CMS/D0 monitoring meeting, Fermi, May
- CMS Grid Production Tools workshop, CERN, June

3.1.2 Planned development work

As stated above, for the next quarter our planned development work includes:

- Hardening of RLS code for easier alpha testing, and alterations after testing
- CAS prototype, Alpha 2

- Full evaluation of the GridFTP performance
- Continued RFT to stay in agreement with the still changing OGSA spec
- Extended support in RFT for buffer size tuning, parallelism and queueing service (the last is in the next 9 month time frame)
- Demos of RFT are planned for GGF6 (October) and SC ’02 (November)
- MDS 2.2 beta release in July, to be followed by a final release later in the quarter

3.1.3 Planned alpha-deployments with experiments

We have also identified the following, more focused development and deployment plans

1) RFT

Alpha testing for RFT will take place as a part of the larger OGSA alpha testing effort. We expect this to begin in approximately the September timeframe, and PPDG will be notified at this time.

2) CAS

Continuing discussions with CMS (Rick Cavanaugh) will take place, as well as soliciting of additional alpha testing from the ATLAS community.

3) RLS

Discussions with ATLAS (begun, although no response), CMS and possibly a third application

3.2 NERSC – SDM

A. File replication

A.1 File replication using HRMs was heavily tested between BNL-HPSS and NERSC-HPSS.

A.2 Various test were conducted on the number of concurrent griFTPs, number of concurrent files that can be supported, etc.

A.3 Additional logging capabilities were added to both TRM and DRM. This capability facilitates the generation of Netlogger style graph that help identify the transfer bottlenecks.

C. NERSC-HPSS security requirements added

The new security requirements at NERSC invalidated the previous version of HRM. We have developed the capability of interacing with HPSS's pftp usinf the new security procedure.

3.3 SDSC – SRB

The activities at the San Diego Supercomputer Center in support of the PPDG have included:

- BaBar support
- web services interface definition with Chip Watson at JLAB
- demos of web services to broaden usage of the technology

Presentations were given on the MCAT technology at PPDG's Interactive Analysis Tools meeting at LBNL. The purpose of this presentation was to provide detailed information about functionalities within SRB's MCAT metadata catalog, and to promote the usage of SRB's MCAT as we have done in the BaBar-SRB collaboration. A presentation on the management of distributed data collections is being given at the 11th HPDC conference in Edinburgh, Scotland, "MySRB & SRB - Components of a Data Grid", A. Rajasekar, M. Wan, R. Moore, July, 2002. A talk on data grid management system and SRB web services was presented at the Grid Web Services Workshop at Indiana University. Concepts were shared among researchers working on similar ideas. At a Grid Reliability Workshop at ISI on July 10, 2002 a talk on

semantic descriptions of exceptions on grids using XML was presented by Arun Jagatheesan. A poster paper on virtual services in data grids is being presented at HPDC-11 (A. Jagatheesan, R. Moore, A. Rajasekar, B. Zhu). Frank van Lingen gave a talk on SRB and how it could be used in CMS that was attended by CMS collaborators across Europe and USA using the VRVS (http://documents.cern.ch/AGE/current/fullAgenda.php?ida=a02757).

3.4 . Wisconsin – Condor

.The Condor project has been heavily engaged in the deployment and operation of the US CMS test grid and its use for CMS production simulation. Progress has been made on the hardening of many components of the Condor and Globus middleware used in the system, but much work remains.

A partial rewrite of the Condor negotiator to support matchmaking and include determination of the gatekeeper to which to schedule jobs has been done for the D0 SAM-Grid project. Extensions to the ClassAds library to support callouts to user functions was also developed. The code is in test at D0.

Extensions to the DAGMAN library were made to support the CMS testbed and GriPhyN virtual data work. Other extensions are being done in collaboration with the EDG WP1.

The support for GDMP transitioned from the US CMS PPDG team to the Condor project. Significant effort was expended in building and preparing the software for distribution through the Virtual Data Toolkit. Use of GDMP in the US CMS testbed resulted in identification and fixing by the WP2 support team of several bugs.

The Condor project continued its collaboration with the EDG WP1 and with the Globus developers in a number of areas. .

4 Appendix

4.1 List of participants

TEAM	Name	F	Current Role CS	1	2	3	4	5	6	7	8	9	10	11	12
Globus/ANL	Ian Foster	Y	Globus Team Lead, GriPhyN PI, iVDGL, GriPhyN						x	x					
	Mike Wilde	Y	GriPhyN coordinator					x					x		
	Jenny Schopf	Y	GriPhyN, iVDGL, Globus team liason, ATLAS-CS liason			x				x	x		x		
	Willliam Alcock	Y							x		x		x		
	Von Welch		CAS									x			
ATLAS	John Huth	N	ATLAS Team lead										x		
ATLAS	Torre Wenaus	N			x			x							
	L. Price	N	Liaison to HICB, HICB Chair												
	D. Malon	N													
	A. Vaniachine	Y													
	E. May	N						x					x		
	Rich Baker	N													
	Alex Undrus	Y													
	Dave Adams	Y													
	Wengshen Deng								x						
	G. Gieraltowski	Y									x		x	x	
	Dantong Yu	Y	Monitoring			x									
STAR	Jerome Lauret	N	STAR Team Lead		x								x		
	Gabrielle Carcassi	Y			x								x		
	Eric Hjort	Y					x	x					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	
	Koen Holtman	N	GriPhyN collaborator												
	Iosif Legrand	N	CS-8:Monitoring Tools								x				
	Vladimir Litvin	N	GriPhyN collaborator	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		
	Eric Aslakson	?	job execution, grid monitoring		x	x									
	Edwin Soedarmadji	?	web services prototyping											x	
	Saima Iqbal	?	web technology evaluation											x	
	Suresh Man Singh	?	grid deployment										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				x			
D0	Lee Leuking	N	D0 Team Lead		x	x									
	Igor Terekhov	Y		x	X	x									
	Sinisa Veseli	Y				X									
	Gabriele Garzoglio	Y							X						
	Andrew Baranovski	Y				X									
HRM/LBNL	Arie Shoshani	y	SRM Team Lead. GriPhyN collaborator				X								
	Alex Sim	Y					X								
	JunminGu	Y					X								
	Alex Romosan	Y					x								
SRB/UCSD	Reagan Moore	Y	SRB Team Lead. GriPhyN collaborator					X		X	x				
	Bing Zhu	Y	CS-8: Web Services					x			X				
SLAC	Richard Mount	N	PPDG PI, BaBar Team Lead												
	Robert Cowles	N										X			
	Andrew Hanushevsky	Y						x	X						
	Adil Hassan	Y						x	X						
	Les Cottrell	N	IEPM Liaison			X									
	Wilko Kroeger	Y						x	x						
JLAB	William Watson	Y	JLAB Team Lead				x	x	x		x				
	Ian Bird	N					X								
	Andy Kowalski	N					X								
	Bryan Hess	Y	CS-8:Web Services				x				X				
	Ying Chen	Y	CS-8: Web Services				X	x			x				
	Walt Akers	Y	CS-8, Web Services					x			x				
Condor/U.Wisconsin	Miron Livny	Y	PPDG PI, PPDG Coordinator. GriPhyN collaborator	x	x	X	x		x		x				
	Paul Barford	Y				X									

	Peter Couvares	Y		X	x							x			
	Rajesh Rajamani	N			x						X				
	Alan DeSmet	Y			x							x			
	Alain Roy	N			x										
	Todd Tannenbaum	N			X										
Globus/ISI															
	Ann Cvernek	Y							x						

CS-1 Job Description Languages

CS-2 Job Management and Scheduling

CS-3 Information Services

CS-4 Storage Management

CS-5 Reliable File Transfer

CS-6 Robust File Replication

CS-7 Documentation

CS-8 Evaluations and Research

CS-9 Authentication and Authorization

CS-10 Experiment Grids and Applications

CS-11 Analysis Tools

CS-12 Catalogs

4.2 Appendix - PPDG Meetings

Apr 10

12:30 p.m. - 2:30 p.m.

PPDG weekly phone meeting

Wednesday, April 17, 2002

12:30 p.m. - 2:30 p.m.

PPDG weekly phone meeting

Thursday, April 18, 2002

4 p.m. - 5 p.m.

PPDG IDAT Wshop pre-meeting call

Tuesday, April 23, 2002

8 a.m. - 10 a.m.

GDMP internal review

Friday, April 26, 2002

PPDG SiteAAA meeting, ANL

Wednesday, May 1, 2002

12:30 p.m. - 2:30 p.m.

PPDG weekly phone meeting

Thursday, May 2, 2002

2:30 p.m. - 4 p.m.

PPDG-GriPhyN-iVDGL Joint Monitoring Working Group

Friday, May 3, 2002

1:30 p.m. - 2:30 p.m.

PPDG IDAT Wshop pre-meeting call

Tuesday, May 7, 2002

1:30 p.m. - 3:30 p.m.

JLAB-Replication internal review

Wednesday, May 8, 2002

10:30 a.m. - 12:30 p.m.

MAGDA internal review

Thursday, May 9, 2002

7 a.m. - 9 a.m.

D0JobManagement internal review

10 a.m. - 12 p.m.

STAR-DDM internal review

12:30 p.m. - 2:30 p.m.

CMS-MOP internal review

Friday, May 10, 2002

11 a.m. - 1 p.m.

BaBar Database Replication internal review

1:30 p.m. - 2:30 p.m.

PPDG IDAT Wshop pre-meeting call

Monday, May 13, 2002

PPDG Job Scheduling, FNAL

Tuesday, May 14, 2000

9 a.m. - 12:30 p.m.

PPDG Job Scheduling, FNAL

Thursday, May 16, 2002

2:30 p.m. - 4 p.m.

PPDG-GriPhyN-iVDGL Joint Monitoring Working Group

Friday, May 17, 2002

1:30 p.m. - 2:30 p.m.

PPDG IDAT Wshop pre-meeting call

Thursday, May 23, 2002

12 p.m. - 2 p.m.

PPDG Site-AAA Working meeting

Friday, May 24, 2002

1:30 p.m. - 2:30 p.m.

PPDG IDAT Wshop pre-meeting call

Wednesday, May 29, 2002

12:30 p.m. - 2:30 p.m.

PPDG weekly phone meeting

Thursday, May 30, 2002

2:30 p.m. - 4 p.m.

PPDG-GriPhyN-iVDGL Joint Monitoring Working Group

Friday, May 31, 2002

1:30 p.m. - 2:30 p.m.

PPDG IDAT Wshop pre-meeting call

Wednesday, June 5, 2002

12:30 p.m. - 2:30 p.m.

PPDG weekly phone meeting

Friday, June 7, 2002

1:30 p.m. - 2:30 p.m.

PPDG IDAT Wshop pre-meeting call

Thursday, June 13, 2002

12 p.m. - 2 p.m.

PPDG-SiteAAA ldap VO meeting

Tuesday, June 18, 2002

PPDG - Analysis Tools & Requirements, Berkeley

Wednesday, June 19, 2000

PPDG - Analysis Tools & Requirements, Berkeley