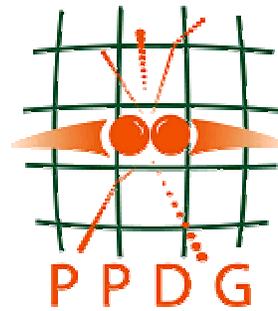


Particle Physics Data Grid Collaboratory Pilot

Quarterly Status Report of the Steering Committee, July - September 2002

31 Oct. 2002



1 Project Overview	2	3.4 . Wisconsin – Condor	22
1.1 Highlights	2	4 Appendix	24
1.2 Project Management and Organization	2	4.1 List of participants	24
1.3 Plans for the next Quarter	2	4.2 Meetings	26
1.4 Trillium Collaboration: PPDG, GriPhyN and iVDG	2		
1.5 Virtual Data Toolkit (VDT)	3		
1.6 Collaboration with EDG and EDT (European DataTAG)	3		
1.7 Pending Issues	3		
2 Common Service Areas	4		
2.1 Introduction	4		
2.2 CS-1, CS-2 Job Description Languages, Management and Scheduling	4		
2.3 CS-3 Information Services	5		
2.4 CS-4 Storage Management	7		
2.5 CS-5 Reliable File Transfer	8		
2.6 CS-6 Robust Replication	8		
2.7 CS-7 Documentation	10		
2.8 CS-8 Evaluations and Research	11		
2.9 CS-9 Security, Authentication, Authorization, Accounting	11		
2.10 CS-10 Experiment Grids and Applications	12		
2.11 CS-11 Grid Interface with Interactive Analysis Tools	18		
2.12 CS-12 Catalogs and Databases	20		
3 Single Collaborator Reports	20		
3.1 ANL – Globus	20		
3.2 SDSC – SRB	22		
3.3 STAR	22		

1 Project Overview

1.1 Highlights

In the last quarter there has been much concentration on continuing to deploy grid systems which can benefit the experiments. Since each of the PPDG teams is only a small fraction of the effort needed both by the CS groups and the experiments all the work has been in together with many other people both in the other US Physics and EU Grid projects, the larger Computer science project groups and the experiment data handling systems. We see a steady progression in the buy-in and commitment of the experiment management to grid systems and technologies both for the near and longer terms.

There was one news briefing¹ to report the significant achievements of the STAR use of HRM and GridFTP for production data replication and storage; there were also significant advances in the ATLAS, CMS, D0, BaBar and JLAB testbeds and grid use.

The SiteAAA project has done significant work to bring to the table many different areas of detailed discussion of Authorization – site, user, policy, experiment etc etc – issues. They plan to submit their report at the December Steering Committee and make recommendations of any future work.

1.2 Project Management and Organization

The weekly PPDG phone conferences have been restructured to address one CS (Common Services) area and one Experiment Activities area each week. These have been quite successful in allowing some focused discussion on developments, issues and concerns of the experiments (til the summer season hit!).

We have started to charge some of the funds allocated to the coordinators in the year 2 budget distribution:

- a) We welcomed Lorri St. Claire at LBL who will help with some of the administrative work for PPDG, in particular we will have her follow up on getting documents and papers related to PPDG submitted and posted on the web pages. There is a new email list (documents@ppdg.net) that will serve as an inbox for incoming material.
- b) The PPDG web pages were revamped thanks to the LBL design group.
- c) A summer student at Fermilab mentored 2 students from the University of Texas Brownsville in learning about VDT and grid components. He also contributed to the GLUE schema testing with DataTAG and helped commission a test and demonstration grid at Fermilab (the Handyman cluster) which is being use for the IST an SC demonstrations.
- d) The cost of the Trillium newsletter is being shared between GriPhyN, iVDGL and PPDG.

We submitted a third year budget profile to DOE.

1.3 Plans for the next Quarter

We will participate in the iVDGL and GriPhyN activity planning meeting December 16-18th and follow this with a PPDG face to face steering meeting and ½ day focused on the SiteAAA projects final report.

1.4 Trillium Collaboration: PPDG, GriPhyN and iVDGL

Members of PPDG are contributing to the WorldGrid interoperability demonstrations, coordinated by Rob Gardner the iVDGL coordinator. This is also a collaboration with DataTAG and the other US Physics Grid Projects. Fermilab hosted a successful demonstration preparation week which made significant progress towards this milestone. <http://www.ivdgl.org/demo/>

Ruth coordinated the upcoming Trillium newsletter that should be available for SC2002 also.

¹ <http://www.ppdg.net/docs/news/news-update-star-hrm-25sep02.pdf>

1.5 Virtual Data Toolkit (VDT)

PPDG should complete the contributions from US CMS UCSD manpower to the VDT testing suite. US CMS and US ATLAS are now using VDT. We are continuing the collaboration with DataTAG on the interoperability of EDG services with services provided by the VDT Client and VDT server.

1.6 Collaboration with EDG and EDT (European DataTAG)

The GLUE collaboration² continued to progress with work proceeding on all sides. Two new sub-groups have been proposed and endorsed by the High Energy Physics Intergrid Coordination Board:

- a) Meta-packaging and distribution. This will involve discussions covering EDG RPMs, LCFG, PACMAN etc. There will be contributions to this working group from the LHC computing Grid project as well as DataTAG and Trillium.
- b) Testing and validation. This is a collaborative effort by LCG, EDG and VDT teams on testing and validation procedures for deployment of interoperable grid middleware. It is anticipated that the LCG will lead and do most of the work for this project.

1.7 Pending Issues

1.7.1 CDF request to join PPDG

CDF have joined D0 in a joint data-handling project to use SAM as well as extend SAM to use standard grid technologies and start to merge the SAM services with other emerging components. CDF has submitted a letter of intent to join PPDG. We are very positive about such an addition to the project towards the goal of promoting commonality and collaboration across experiments for the development and deployment of common grid services. However, there are no available funds to support this activity in the project at this time.

Both ALICE-USA and PHENIX have shown a verbal interest in participating in PPDG but there have as yet been no official request from either of these experiments.

1.7.2 Support of the Grid Middleware for Production Services

There has been no concrete action to date from DOE to our presentation made in June, which raised the issue of support for the grid middleware being deployed in production service in our collaborating experiments. Once software is deployed in production there is a need to clarify expectations for support and problem resolution.

1.7.3 Availability of Cross-Project Technical Resources

There are several areas of common service specification and development that would greatly aid our collaboration with the European Data Grid project and our ability to promote common services across the experiments. The structure of PPDG in teams makes our ability to deploy such efforts difficult: examples of needs are: a) Testing of the EDG WPI resource broker and user interface in a PPDG environment; b) Specification of service interface to replica catalogs to allow integration of the different implementations into a common framework. b) Evaluation across the various monitoring tools and frameworks - at the moment there are several independent designs and implementations; c) Computer science group collaborative work on the CS-11 Analysis Tools sub-project.

1.7.4 Production and Facility Services

While we proposed a working group on production and facility requirements and services, a lack of available effort stalled this group from getting off the ground. In the year 2 project plan we proposed a

² <http://www.hicb.org/glue/glue.htm>

focus on these issues. While this is happening in each of the experiment test beds, the cross-cut effort to date is happening in the Site AAA project.

2 Common Service Areas

2.1 Introduction

We continue to encourage cross-experiment and computer science exchange of technical ideas and designs. Given the distributed and extremely matrixed nature of PPDG we find there is steady but slow progress.

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

2.2.1 Collaboration with EDG WP1

The collaboration with EDG WP1 has proceeded with a second joint meeting with WP2, WP1 and Condor before the EDG retreat in September. Francesco Prelz, WP1 manager, visited Fermilab in October and discussed requirements and design with D0 and JLAB. A member of BaBar grid team – Gilbert Grosdidier is now the EDG testing team manager and we are proposing to work with him and WP1 on testing the EDG resource broker with VDT client. A joint document was written between WP1, D0 and Condor on the status of the job scheduling and planning implementations to date.

2.2.2 Common Job Description Language

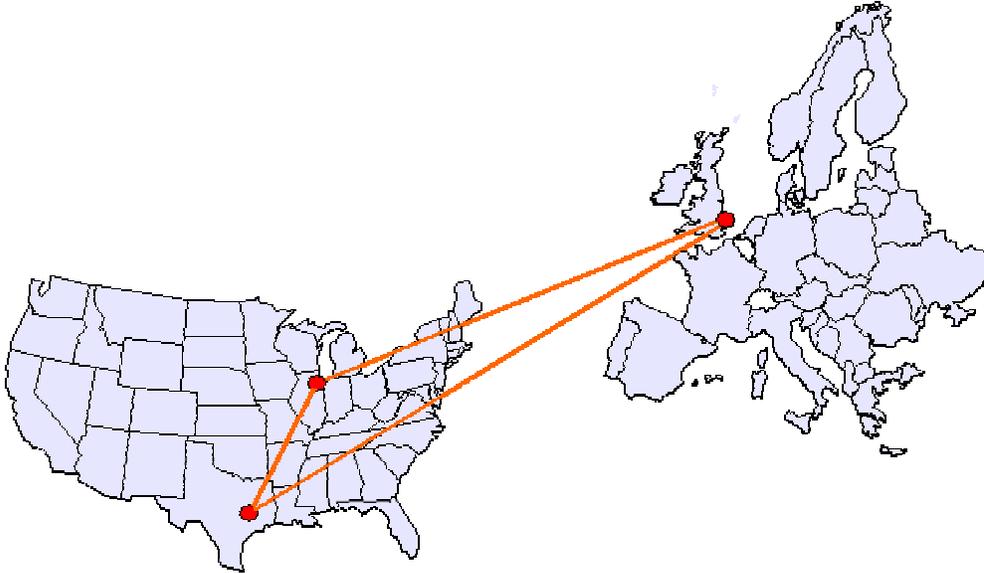
A PPDG mail list ppdg-jdl@ppdg.net and working group has been formed led by Andy Kowalski at JLAB. A draft requirements document has been started. The issue of whether the “Language” word is appropriate in the title is in discussion. Topics include whether or not a common JDL is required, how a common JDL would be used, and comparisons between existing JDLs. While INFN Grid is encouraging a common JDL across the EU and US experiment applications, more detailed technical discussions have raised whether this is in fact an appropriate goal. Andy Kowalski, acting as coordinator, distributed a draft requirements document to spur further discussion. This is a work in progress.

2.2.3 SAM Job and Information Management (JIM)

The work with the Job and Information Management (JIM) project has gone very well in the last several weeks. The collaboration with the Condor team, especially Todd Tannenbaum and Alain Roy, has been very productive. Two UTA - CSE students (funded through PPDG), Siddharth Patil and Abhishek Rana, and our Finnish Coop student Hannu Koutaniemi, have contributed a great deal to the project and it is well its way to being a working system. Work is ongoing with the Condor team to adapt the submission layer and the Match Making Service they are providing, and integrate them into the system. We have tested and integrated the changes delivered so far, and on October 10th Igor Terekhov and the SAM-Grid team, including the students, Gabriele Garzoglio, and Andrew Baranovski, presented a working demo at the DZero Collaboration Meeting.

The demo involved three testbed sites: FNAL, Imperial College London, and University of Texas Arlington. Each site advertised the availability of resources for job execution (1 cluster per site). One machine at FNAL and one at Imperial College acted as submission points. A simple Job Description Language, based on the attribute-value pair paradigm, is employed to start a job from any of the submission sites with a simple “samg submit” command. The Job Broker determines the best location for the job to execute; currently, the brokering algorithm queries the SAM replica catalog and dispatches the job to the resource that has the maximum number of files wanted by the job that are already available in the caches.

Job status as well as resource information can be accessed via the monitoring service: <http://samadams.fnal.gov:8080/prototype>. Information relating to the system at a particular site can be displayed by clicking on the map of the SAM Grid (see figure below); other views of the system, like the list of jobs submitted from a certain submission site, are also available.



The JIM products have been packaged via Fermilab's UPS/UPD and instructions are available for deployment at additional sites. Some of the products dependencies include the Globus Toolkit (GRAM, MDS, and GSI in particular) and various modules of Condor-G. We are soliciting additional DZero sites to be included in the testbed. Our next site will be MSU where we have close associations with Chip Brock (co-lead of DZero Data Handling with Lee Lueking) and Tom Rockwell who will do the actual installation and testing. We are using doesciencegrid and UK-HEP certificates for authentication. We especially thank Rod at IC, and Jae Yu and Mark Sosebee at UTA, for their assistance making the remote execution sites a success.

2.3 CS-3 Information Services

2.3.1 Joint PPDG, GriPhyN iVDGL Monitoring Project

Meetings of this group are in abeyance to allow concentration on the Glue³ Schema work. At the CS-3 PPDG phone conference in October information was exchanged about MDS 2.3, MDS in GT3 plans. Mona-Lisa from CMS includes in its goals historical/strip chart displays of the information. Hawkeye from Condor is focusing on providing sensors and analysis for system problem identification and notification. Other work which is of general interest and use are:

- Additional Information Providers for MDS from Globus

- Adaptors to allow information to flow between monitoring systems – e.g. mona-lisa to mds/ldif.

There is progress on several fronts and a general increase in understanding of the necessary components and capabilities in practice necessary for grid and fabric monitoring and information.

2.3.2 Globus MDS and Glue Schema 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. The CE-schema has been completed and initial testing of the information providers is progressing. Globus released their version as part of the MDS 2.2 release. Work for the storage element (SE) schemas continues. Brian Tierney of the DTMF project has also contributed significantly to this work and continues as coordinator of the glue-schema phone conferences. Alternate week conferences are now led by Cristina Vistoli of DataTAG and involve iVDGL as well as PPDG collaborators.

³ <http://www.hicb.org/glue/glue.htm>

2.3.3 Collaboration with Other HENP Grid Projects

As part of the collaboration through GLUE and the WorldGrid demonstrations a GIIS hierarchy has been established across many US HENP grid sites with a root of giis.ivdgl.org. With specific information providers, information across these sites can be viewed through Ganglia and Nagios.

2.3.4 Collaboration with IEPM, Network Performance Monitoring⁴

2.3.4.1 Web Services

Warren continued to work on web services. He attended a PPDG conference call to discuss web services. Warren also continued to work on user authentication using X509 certificates. Warren also discussed how to web services could be useful with the team working on setting up a resource broker.

2.3.4.2 PingER

There is a new PingER site at CAIA at Swinburne University in Australia. Les is reviewing the beacon sites. He cleaned up some troublesome PingER beacon hosts (Waikato, Spain) and is trying to get ping access to a host in Bombay. He also discussed setting up monitoring at Stanford medical school. Les looked at the transoceanic links used by PingER are like for Sally Floyd.

2.3.4.3 IEPM-BW

Warren discussed setting up IEPM-BW target sites with contacts at University of Utah and University of Michigan. Les and Jerrod worked with someone in the Czech Republic to set up an IEPM-BW remote site. Les is also working with UIUC to get an IEPM-BW remote host there.

Les talked to Richard Hughes-Jones in order to get web space for the monitoring at Manchester. Les is also looking into ssh problems at IN2P3, JLab and working with BNL to get a new host to monitor. Les is conducting a survey from I2 E2E folks on the iepm-bw toolkit/installation at their site. Les worked with Connie and Paola to get a real time display of throughput from the IEPM-BW bandwidth tests. Les worked with Connie to get the bandwidth tests to work in flood mode, i.e. run all tests simultaneously. Les is working with Sylvain Ravot of Caltech (stationed at CERN) to see how high throughput we can get from CERN to SLAC. Warren implemented some patches at SLAC. Les worked with Rice and LANL put together an INCITE poster submission for the DOE SciDAC booth at SC2002.

Les exchanged emails with Rob Thomas concerning his DNS monitoring page. Les worked on the networking panel for SC2002.

2.3.4.4 iGrid2002

Les worked with Paola, Jiri, Jerrod and Anthony from NIKHEF to get demos ready for iGrid2002 and to get access to the iepm-bw remote hosts from iGrid2002.

2.3.4.5 Beluga

Les is evaluating the new Beluga ping/traceroute monitoring tool from CAIDA.

2.3.4.6 Web100

Warren reinstalled Antonia and the modified web100-enabled kernel following tips from the web100 discussion list to attempt to make Antonia more stable.

2.3.4.7 Admin

Warren repaired a problem with the AMP machine by replacing the hard drive sent from SDSC. Warren upgrade the IPEX machine by replacing the hard drive sent from Telcordia. Warren installed and configured two machines (Hebe and Iphicles) ready for SC2002.

⁴ <http://www-iepm.slac.stanford.edu/>

2.4 CS-4 Storage Management

There was progress in several areas in the application of the SRM interface specification and GridFTP protocol to provide interoperable access to storage resources. PPDG continues to work closely with the SciDAC SRM and EDG projects to continue and increase the functionality and commonality in interface and protocols. The GGF working groups and the upcoming SRM meeting in December at CERN provide the forums for continued collaboration.

2.4.1 LBNL-SRM Development

People involved: Junmin Gu, Alex Sim, Alex Romosan, Arie Shoshani

2.4.1.1 Debug and setup the File Monitoring Tool (FMT) to track File Replication

The File Monitoring Tool (FMT) was developed to provide dynamic web-based graphical information on the progress of file replication. The new version developed this quarter under the SRM middleware project was deployed in NERSC for the file replication experiments. Tracking the progress of large number of multi-file replication requests was done using this tool. In these tests the FMT scaled well to 100s of files.

2.4.1.2 Development of a WSDL gateway for HRM version 1.0

The goal of this activity is to have a single agreed upon WSDL-based interface for Storage Resource Managers (SRMs)) working across all types of SRMs. Our first goal is to use the new WSDL gateway to HRM-HPSS developed under the SRM Middleware project. This is done in collaboration with Fermilab, who have developed the same capability for Enstore, and Jlab who have developed the same capability for JasMine. We have been working with these two groups to agree on a version of WSDL, and to develop client software at Fermilab and LBNL to access files from any of these three mass storage systems: HPSS, Enstore, and JasMINE. We plan to have a demo of this capability at SC 2002.

2.4.2 SDSC – SRB

The activities at the San Diego Supercomputer Center in support of the PPDG have focused this quarter on the demonstration of data sharing environments for the Tier-2 high-energy physics site at Caltech. In collaboration with Conrad Steenberg conrad@hep.caltech.edu, the Storage Resource broker has been installed on the Tier-2 site that is distributed between UCSD and Caltech. A Unix Python interface has been written for the SRB to support the integration of the Clarens system on top of the SRB data handling environment.

The Python interface was developed by Bing Zhu, and details of the functionality are provided in document PPDG-22.

2.4.3 Jlab-SRM

Jefferson Lab has continued to develop and deploy web services based Storage Resource Management software (server and client) on two fronts, one using the lattice QCD project as the customer, and the other using experimental physics as the customer.

2.4.3.1 JLAB-QCD

The lattice data grid is moving from testing to operational use, with much of the time in this quarter spent in adding some additional components related to layering the J-SRM software (java SRM) above a standard unix file system (i.e. not above JLab's Jasmine or other SRM; intended as a light weight SRM at remote sites). Also during this quarter a number of bugs were revealed during beta testing and these have been corrected.

The first large scale production testing of J-SRM and the jparss parallel file transfer server was performed. In this test, the GFM (Grid File Manager) java client requested a 3rd party file transfer of an entire directory tree (single click) between the J-SRM at MIT and the J-SRM at Jlab. The reliable file transfer service at JLab negotiated the file transfer protocol (as per the SRM functional spec), which resulted in jparss being selected. Jparss recursively transferring 5157 files totaling 146.7 GBytes from MIT to JLab. The transfer

completed during the night after the GFM client had disconnected. Bandwidth achieved was 2.44 MB/s, which is essentially the limit of what can be achieved between these two sites (mostly limited by MIT's T3 ESnet connection).

User testing of the software between MIT and Jefferson Lab revealed some additional features would be needed, particularly for moving trees from one site to another where the destination tree already exists (recursive merge capability).

2.4.3.2 JLAB-Experiments

The second project at JLab targets experimental physics and integration of existing computer center software. The experimentalists make heavy direct use of the Jasmine storage system (site disk and silo management), and so the primary goal of this part of the development is to extend the capabilities of Jasmine to a remote site using web services. This is being done by adding a PPDG standard SRM interface to Jasmine. The standardization of the interface will allow interoperability of tools with non-Jasmine clients and back ends.

This quarter the SRM interoperability testing passed an important milestone. The Fermi SRM client successfully moved files from the JLab SRM server. Mutual authentication was done using GSI and DOE Science Grid certificates. Work is under way to do the inverse: to let Jefferson Lab's SRM client interact with Fermi Lab's server. During this period, network problems between JLab and FSU were overcome. Several hundred GB per test were transferred to FSU using the SRM interface.

2.5 CS-5 Reliable File Transfer

Experiments continue to use GridFTP for high throughput file transfers and several issues, mainly involving TCP and network connectivity errors, were identified and investigated by the Globus team

2.5.1 Globus RFT

The Reliable File Transfer Service (RFT) will be Globus's first OGSA service, completely up to the current specification. 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. During this quarter we extended this service to allow for parallelism and changing the buffer size.

An alpha-release, and solicitation of alpha testers, is planned for Fall. The schedule of these releases, which is also the schedule for the Globus Toolkit 3.0 technology releases, is:

- Technology Preview 4: Oct 31, 2002
- Technology Preview 5: Early December 2002
- GT3 Alpha: GlobusWORLD, Jan 13, 2003
- GT3 Beta: End of April, 2003
- GT3.0: End of June, 2003

Additional information on RFT is available at http://www.mcs.anl.gov/~madduri/ogsa_docs/reliable_transfer.html.

2.6 CS-6 Robust Replication

2.6.1 GDMP

As new versions of GDMP are released by the EDG they are included in the VDT and deployed on the US CMS testbed. It is clear that GDMP will only be supported for the short term and work needs to be done to identify the replacement of choice that can be more common across the PPDG experiments. We expect to do this in collaboration with EDG WP2, Globus ISI, SRB, and the PPDG experiment data management groups.

2.6.2 BaBar Database Replication (BaBar-SRB)

Work continues with the integration and deployment of SRB based file catalog and data replication between SLAC, IN2P3 and RAL. The goal is to have this in production by early 2003 reliable replication between SLAC and in2p3. Replication from SLAC to RAL should be production ready before the middle of the year. Issues of collection management, interface to the various MSS systems in use, robustness and completeness are being addressed.

2.6.3 Globus ISI RLS work

During this quarter, we continued our development and testing of the Replica Location Service. Features added to the RLS this quarter include: 1) Packaging the RLS with the Globus toolkit; 2) Greatly improved documentation of the RLS server using the Doxygen system; 3) Implementation of bloom filter compression to summarize the state of LRCs, thus reduce the size of soft state updates and reducing the associated memory and network requirements for performing updates; 4) Implementation of a Java version of the client API that wraps the existing C client; 5) Implementation of a gridmap file for the RLS servers that supports GSI authentication. In addition to continued functional and performance testing, the RLS has seen wider use this quarter, being deployed in the European DataGrid testbed and the Earth Systems Grid testbed. We continue our testing efforts with Work Package 2 of the EDG project, and the two projects plan to participate jointly in a demo at the upcoming SC2002 conference.

2.6.3.1 New Initiative to Turn the Replica Location Service into an Open Grid Services Architecture Service

In August and September 2002, we began discussions internally and with IBM on the possibility of turning the Replica Location Service into a Grid Service compatible with the Grid Services Specification of the Open Grid Services Infrastructure working group of the Global Grid Forum.

2.6.3.2 Plans for next quarter

Continued development and testing of Replica Location Service

We plan to continue development and testing of the RLS in the coming quarter. However, to some extent, the current set of features has frozen until we get additional feedback from our user community about whether additional functionality is required, so our efforts will be focused more on testing and debugging the code. Planned development for the coming quarter includes making the RLS GT3 security compatible, which requires that the RLS accept the new GSI certificates that will be developed as part of the Globus Toolkit version 3.

Widespread deployment of Replica Location Services

We will focus on wide-spread deployment of RLS in the coming quarter. We will deploy a fairly large RLS testbed as part of a demonstration of RLS for the SC2002 conference. This testbed will include approximately twelve replica location index nodes and a larger number of local replica catalogs. Development for the testbed includes creating specialized information providers for the MDS information system to reflect the state of RLS catalogs in the testbed.

In addition to the RLS testbed, we will continue to work with PPDG to test and deploy the RLS.

Development of Replica Location Grid Service Specification and GGF Working Group

In the coming quarter, we plan to develop a first draft of a specification document for an Open Grid Services Architecture Replica Location Service. This draft specification will be developed under a proposed OGSi Data Replication working group of the Global Grid Forum.

2.6.3.3 Papers Published or in Progress

Giggle: A Framework for Constructing Scalable Replica Location Services. Ann Chervenak, Ewa Deelman, Ian Foster, Leanne Guy, Wolfgang Hoschek, Adriana Iamnitchi, Carl Kesselman, Peter Kunszt, Matei Ripeanu, Bob Schwartzkopf, Heinz Stockinger, Kurt Stockinger, Brian Tierney. To appear in Proceedings of SC2002 conference. Final copy submitted July 26, 2002.

<http://www.isi.edu/~annc/papers/chervenakFinalSC2002.pdf>

2.6.3.4 Presentations Given

22 July 2002: Presented Giggle Replica Location Service to Replication Research Group at Global Grid Forum Meeting in Edinburgh, Scotland.

2.6.4 JLAB-Replication (JLAB-SRB)

Developments on this project in this quarter were limited to improving the underlying SRM-like functionality of the two sites' software (reported elsewhere in this document).

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-23	Giggle: A Framework for Constructing Scalable Replica Location Services. (submitted to SC2002)	7/02
PPDG-22	SRB Extensions in support of BaBar replication	10/02
PPDG-21	GridMonitor : Integration of Massive Facility Fabric Monitoring with Meta Data Service in Grid Environment, RHIC, D.Yu, R. Baker	10/02
PPDG-20	Planning on the grid: a status report - draft , EDG WP1, D0-Grid, Condor-Project.	10/02
PPDG-19	Grid Service Requirements for Interactive Analysis	9/02

Talks and presentations:

Presentations & Publications	
October 2002	Large Site AA Research Group at GGF6; Talk to the LHC Computing Grid Project Grid Deployment Board and to the HICB ; Report at D0 Collaboration meeting , Lee Lueking. SiteAA ESSC
September 2002	Grids: the US Atlas Perspective , John Huth; Report on US Grid integration at CMS week; Greg Graham. SRB at BaBar Grid meeting, Wilko Kroeger, Adil Hasan.
August 2002	ESSC SiteAA
July 2002	Talk to BaBarGrid; Ruths PPDG talk and Igors D0 talk at the GGF DataGrid BOF. SiteAAA talk

2.8 CS-8 Evaluations and Research

PPDG evaluation and research is done as part of the ongoing activities of the experiments and computer science groups. – well.. . maybe it needs better definition and we could have a weekly phone conference to discuss this.

2.9 CS-9 Security, Authentication, Authorization, Accounting

2.9.1 Certificate/Registration Authority

The managers of the DOE Science Grid certification authority have decided that the domain name doegrids.org is a better reflection of the use and scope of this CA so that www.doegrids.org is now the primary URL. As you can see by visiting that URL, the community being served by this CA is considerably larger than that for the DOE Science Grid project. As of today there are 479 valid certificates issued by this CA (more than just PPDG) and a total of 53 certificates have been revoked (usually for accidental destruction of the associated private key).

At present the most significant technical issue is the question of having a type of service certificate that is not tied to a particular host. This usage can arise in from monitoring or other operational procedures at computing facilities where client processes need to authenticate but there is not really a single individual identified with the process but there is a responsible group associated, like operators, or sysadmins, etc. This is under active discussion on several related mailing lists and will be considered at the December meeting of the EDG CA managers at CERN.

2.9.2 Site-AAA

The Site-AAA activities progressed well on individual site activities, and status is summarized below. A significant fraction of the group coordinated effort was spent communicating intermediate consensus results with the Globus, Condor, and EDG developers and working toward presenting final results useful in the GGF context. Along that line, the GGF Site AAA Research Group was formed, as there was general encouragement of this requirements work at GGF5. All PPDG Site-AAA sites are participating in that effort and plan to contribute to the first working meeting at GGF6.

Conference call meetings continue to be 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 are proving fruitful with a number of sites choosing EDG components as part of their integration efforts. 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. Two sites have full drafts of policy requirements.

Brookhaven's project of integrating local user account registration with the Atlas framework has begun. BNL maintains the Atlas VO membership service and has modified the EDG gatekeeper maintenance software to permit easier integration with site requirements. US Atlas, iVDGL and EU-DataTAG are using this for their upcoming SC 2002 demo.

JLab SRM integration with GridFTP as the file transfer mechanism works for read access and in progress for write. An initial design of Grid batch service access has been completed and is currently under review. Draft requirements and lessons learned documents are circulating.

FNAL has their KCA proxy generation service running and is dealing with operational concerns of redundancy, finalizing certificate format, and CA operational requirements of the DOE Science Grid and EDG PMA. D0 and CMS are using Grid Resources based on this service now, with production tests planned for the next quarter. FNAL has a prototype implementation of a Site Authorization Service (SAZ) based on the EDG LCAS model and has modified its dCache Storage Resource interface to utilize this service. In the process FNAL implemented a Python GSI-GSS library and has begun talks about giving this

to the COGS project. With this site authorization service in hand, implementation of external CA request acceptance checks can begin. The issue of authentication strength is the remaining major issue of principal.

LBL has developed an authorization and authentication plan for access to the NERSC supercomputers and the PDSF. Tests of the EDG LDAP VO service and the myproxy server have been performed.

SLAC has developed a proposal for a proxy generation service based on the EDG VoMS service. Negotiations are in progress with the DOE Science Grid CA as to whether this use is acceptable to their policy.

A major issue identified by this project was the individually authenticated, site wide 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 EDG LCAS ideas seem to address this idea but are not yet implemented as services. Consensus is this is the key authorization issue from the site perspective. At GGF6, principals from the Globus, EDG, and FNAL authorization service projects reached agreement in principal on how to use a common interface. Detail specification is now in progress.

Two other major issues, that do not have generally acceptable solutions proposed are: a) authentication (and re-authorization) for long-running jobs, and b) accommodating requirements for supplemental information by authorizing agents (further authentication, information not presented with the request, etc.). For b), the answer may be that all information needed for authorization must be presented with the GSI credentials or else fail. In that case, the protocol for describing information needed for authorization, and method of transmitting that to authorizing agents needs to be specified.

Details of the activities in this PPDG cross-cut project (including detailed site reports for this quarter) are described at <http://www.ppdg.net/pa/ppdg-pa/siteaa/>.

2.9.3 Globus Site-AAA work

Von Welch continues to actively lead ANL participation in the PPDG Site-AAA group. In order to satisfy site authorization and identity mapping needs, we have designed a callout interface from the Globus Toolkit. We have presented the general outline of this design to the AAA group and to members of EDG at a meeting at GGF6 in Chicago. There was general acceptance of our proposal and we have presented a draft design document so that it can be evaluated in detail against PPDG (and EDG) requirements.

2.9.4 Globus CAS

The second release of our Community Authorization Service (CAS) prototype was made available in September. This release included significant re-engineering in order to meet requirements expressed by the PPDG Site-AAA group. The release was announced to PPDG but to our knowledge no alpha testing by PPDG users has begun, although the LBNL Site-AAA group is planning to start testing CAS in October.

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)

During this period of time, Magda was tried out for the first time in the testbed production for the Atlas Data Challenge 1 (DC1), and was proved to be a useful and feasible tool. 'globus-url-copy' and 'globus-job-run' were integrated into magda_putfile. magda_putfile can be used to do third-party transfer, put files to BNL HPSS directly, and register files to the Magda database. 'globus-url-copy' was also used in the bulk data replication between BNL and CERN. Another trans-Atlantic transfer engine used is 'bbftp', which worked very well also. A rate of 300 GBytes/day was observed from BNL HPSS to CERN castor, and about 4 TBytes data had been copied for the Atlas DC1. The file spider worked diligently as usual. 35K DC1 files are available through Magda.

Magda servers were moved and split in this period. The database now is on 'magda.usatlas.bnl.gov', which is a dedicated MySQL server. Significant performance improvement had been reported by Kaushik De - no more MySQL timeout error occurred to him after the transition. The main Magda web site now is <http://www.atlasgrid.bnl.gov/magda/dyShowMain.pl>

The command `magda_getfile` was developed and gridified. It does file copying economically: it figures out which file instances to retrieve, and fetches remote instances only if it is necessary. `magda_putfile` automatically determines which remote host to contact if the remote host is not provided as an argument and it is needed; if it is provided, use the provided host as a means to access storage site, should be hidden from users if possible.

A new attribute was added to the Magda location. The file spider checks that attribute to decide whether or not it needs to crawl a location. This was found to be very useful. In Magda, for a logical file, only one prime instance is allowed, while multiple replica instances can be distributed at different locations. If two primary instances exist for some reasons, the file spider can be disabled to scan one of the locations until problems are reported and fixed.

The interfaces for the deleting of file records had been developed both on the web and on command line. A protection of empty input from the web form was added.

The metainfo proposed by the Grappa developers was saved to the Magda database, can be retrieved with file name from both the web and the command line.

The integration of Magda with GDMP was tried in Milan: 'gdmp_register_local_file' and 'gdmp_publish_catalogue' were put into `magda_putfile`; it seems working as expected. The strategy of managing files which are distributed on local disk of each individual node of a Linux farm, was discussed and will be implemented soon. A farm can be seen a special 'Magda' site, and the location associated with it will be a directory path preceded by the host name of the node. The Magda web services were implemented with perl SOAP::Lite module. They can be invoked to query the database and register files. With web services, Magda provides one more option to interact with the database. Magda commands are available by default on BNL Linux farm. Shared disk cache is supported: users could share the same instance if they need the same file. This should reduce the traffic with the mass storage.

We finally got a workable machine with root privilege. The installation of software has been started to do the RLS test.

A poster about Magda and live demonstrations of file transferring with Magda are in preparation for SC2002 demo. The STAR experiment has Magda servers running and is evaluating it.

2.10.1.2 Evaluating the Capabilities of GRAPPA:

Jerry Gieraltowski devoted the month of August to evaluating the current Athena capabilities available through the GRAPPA web-portal, and creating a user's guide for the existing Athena capabilities: <http://www.hep.anl.gov/grappa/athena>

This provided an easy user's guide for members of the US-ATLAS grid to execute Athena "atlfast" and "atlsim" jobs using the GRAPPA web-portal. Jerry worked closely with Dan Engh (U of Chicago) to debug various flavors of executing atlfast and atlsim (i.e., locally, boxed and transmitted to a remote end for execution, or via the afs network), and also with Pavel Nevski of BNL to debug various releases of the atlsim software executing in the GRAPPA environment.

2.10.1.3 Inter-Grid Connectivity:

In early September, the GRAPPA web-portal was used to successfully submit an Athena fast simulation (atlfast) job to a site in the US-ATLAS grid and, in parallel, to a site in the EDG Testbed1 Production grid. The job generated 100 events. In the US-ATLAS Grid, the job was executed at Boston University on a Linux, RedHat 7.2 system. In the EDG Grid, the job was first submitted to a I/Gatekeeper server at Argonne, routed to the CERN Resource Broker and subsequently dispatched by that RB to the CERN compute resource (lxshare0339.cern.ch) for execution. GRAPPA was able to monitor the status of both jobs and successfully returned all output to a destination specified by the user. Further architectural changes were needed to allow GRAPPA to successfully submit boxed jobs to US and EDG nodes running

the Condor job scheduler. These changes were jointly created and tested with Dan Engh (Univ. of Chicago) and Shava Smullen (Univ of Indiana).

A comparison of the grid functionalities of the US-ATLAS, EDG, and NorduGrid grid networks was performed to assess the possibility of inter-grid connectivity. A report of these findings was presented by at the ATLAS Software Week in late September. The report can be found at:
<http://www.hep.anl.gov/gfg/us-edg-interconnect/jerryg-intergridcomparison.ppt>

Jerry Gieraltowski, Rob Gardener, Flavia Donno, and Saul Youssef have worked closely to define inter-grid connectivity requirements and ensure that they are implemented in the US-ATLAS, US-CMS, and DataTAG servers.

2.10.1.4 SC2002 Demo of Grappa

A demonstration of the Grappa web portal is being prepared for SC2002 in conjunction with iVDGL as an example of inter-grid connectivity.

2.10.1.5 US ATLAS Grid Testbed

Work on PPDG/SC2002 Grid Monitoring Project. Deployed US-ATLAS Map Center for SC2002 ATLAS demo. Grid Map Center has been designed to logically and graphically represent all elements, applications and services running over grids. It polls grid entities and services (GridFtp, MDS, Gatekeeper), check their status and builds aggregated views of difference types of grid entities.

The Web site can be found at: <http://www.atlasgrid.bnl.gov/mapcenter/>

Testing GLUE Schema. Deploy Condor-G on a BNL Grid testbed. Support GridFtp, Replica Catalog and GDMP tools which can be used by Magda- a distributed data manager prototype

VO server management and automatic grid-mapfile generation: Maintenance of the US-ATLAS Virtual organization server, add new users to ATLAS VO server. Maintenance of the VO server software package so that it can download the certificates from DOE-certificate authority on-line repository and modify the format of certificate, make them to be GLOBUS compatible. Enhancement of the grid mapfile software package developed by EU datagrid.

Traditionally, if a grid user wants to use the grid resource at a specific site, he/she has to contact with the site administrator for authorization, then the site administrator manually updates the gridmap file to authorize the user to use the local resource. This model is not scalable. The EU datagrid developed the gridmap package which could download the user certificates from its virtual organization and generate the mapping from user's identify to local accounts. But the model does not satisfy our site requirement and needs site administrator's manual intervention. Dantong Yu added several functionalities: such as merging the existing mappings with new mappings obtained from VO server, automatically updating the grid mapfile at daily basis The software package can be distributed by pacman. To install it, just type:
 pacman -get BNL-ACF:edg-mkgridmap.

2.10.2 BaBar

2.10.2.1 Collaboration with EDG WP2

Andy Hanushevsky attended the EDG retreat and workshop to participate in design discussions for the RLS and file replication services for EDG release 2 and beyond.

2.10.2.2 Deploying EDG software for BaBar use:

BaBar is working closely with the EDG to provide an end-to-end grid that spans the collaboration sites. The status was reported the BaBar grid meeting in September. Three machines are available at SLAC for the EDG install: bbr-gate01-03. EDG-1.1.4, previously installed on both CE & SE was replaced with EDG-1.2.0, following the upgrade of the RB at Imperial College in the U.K., and has been running since 22/08/2002. UI distributed via afs from centrally managed location to Linux front-ends at SLAC (nories). Three days of stress testing from SLAC towards SLAC show good stability and reliability (> 95%, much higher than before). The RB "hangs" that plagued the 1.1.4 installation appear to be no longer present within 1.2.0. Mild stress-testing of the system was achieved by passing several 10's of requests to the RB

with a 10sec interval between requests from the same UI and targeting the same CE. This is a significant improvement over the previous version where requests had to be staggered with many minutes interval.

However, there's a significant amount of testing that needs to be done before the system can be considered production ready.

2.10.2.3 Additional PPDG BaBar Team Lead

Tim Adye, from the Rutherford Lab in England, has agreed to participate in PPDG as an additional team lead for BaBar. This will provide for better integration of the BaBar grid efforts spanning the collaboration in the US and Europe.

2.10.3 CMS

The CMS PPDG team continued the quest to deploy a stable, productive grid for CMS simulations. By the end of the quarter an Integration Grid Testbed was constructed which within a few days was running fairly reliably and returning resulting ntuples to the CERN Tier 0 for use across CMS. Anzar Afat is the coordinator of the Integration Testbed. The CMS simulation MOP master scripts was installed and operated from Fermilab as well as at Wisconsin.

Work continues on the development testbed to integrate MOP with the MCRunJOB scripts which provide workflow management for the complete simulation process. Work started on designing the changes needed for MCRunJob, refdb and other components towards the CMS DC04 data challenge milestones in the spring and summer of 2003. The US CMS grid participants are working closely with the CMS grid and production coordinators at CERN to form an integrated team.

Design and architecture meetings were held with the Chimera project to understand the requirements and integration of virtual data components with the CMS production simulation programs.

The US CMS Software and Computing WBS was revamped. One change was to explicitly show dependencies and reliance on deliverables from the grid projects including PPDG.

UCSD PPDG continued to contribute to the VDT testing scripts. This responsibility will be regarded as complete at the end of this quarter. UCSD and Fermilab are contributing to the Trillium WorldGrid CMS demonstrations -working with iVDGL and DataTAG towards MOP executions across US and EU sites and to support the EU CMS simulation job submissions on US testbed sites.

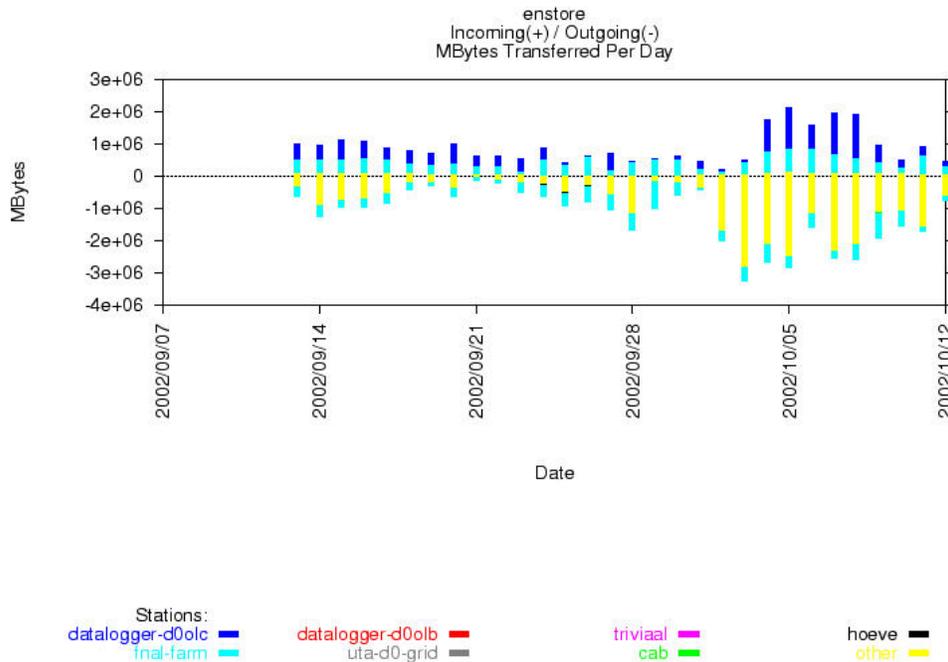
<http://www.ivdgl.org/demo/worldgrid/>

UCSD and Caltech continued to evaluate and test the use of SRB in the US CMS environment including an interface to SRB storage and metadata facilities through the Clarens analysis server. There is interest in the version of SRB using postgres and on the availability of the system in the public domain.

The mona-lisa monitoring system was deployed on 3 sites of the US CMS testbed. Several sensors have been developed together with improvements in the control and display interfaces. Monalisa is becoming an integrated and valuable tool for the US CMS grids.

2.10.4 D0

We have been 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 BaBar, CDF, DZero. In preparation for this we have developed monitoring tools that enable us to tally data transfers among all the active SAM stations. Shown in the chart below are daily transfers into (+) and out of (-) the mass storage system at Fermilab from and to several of the stations. There are nearly 3 dozen stations located at Fermilab and DZero collaborating sites.



We have submitted papers on SAM and SAM-Grid for the Beauty 2002 conference where Rod Walker presented, and the ACAT 2002 conference where Gabriele Garzoglio and Igor Terekhov presented the SAM and JIM projects. We will be presenting an overview of SAM and a case study of several cluster configurations within the system at the Large Scale Cluster Computing Workshop to be held at Fermilab the week of October 21st.

Over the summer we have increased the sophistication of the data transfers for SAM. The CRC for each file transferred is now verified which has eliminated a minor file corruption problem that has been observed in the past. This corruption was found usually to be related to problems with the raid disks, and not the actual network transfer, as was first suspected. GridFTP has been added as a transfer protocol for SAM and is being tested between Imperial College and the Central-analysis station at Fermilab. Gabriele and Rod have packaged the data handling bundles of Globus for both client and server, along with a security patch from Globus. There is a product called `sam_gsi_config` that checks that the configuration at the site is the correct one to handle transfer jobs. There are also utilities to keep the gridmap file up to date. There are problems with the multi-stream transfers to the IRIX version of the GridFTP server that causes transfers to frequently hang. This is being investigated and we hope there is a fix for this very soon. For now, we will only use single threaded gridftp and strongly encourage Globus to fix the parallel version.

SAM already has kerberized rcp (or scp), ftp, and bbftp enabled and each site can chose to set up their transfers as they like. Transfers using bbftp or GridFTP are limited extra-domain transfers for D0, although CDF has used bbftp for some of their intra-domain traffic as well. There is a new version of bbftp which now supports gsi security, and has bug fixes relating to timeout problems which have been observed in the past. See <http://doc.in2p3.fr/bbftp/doc.2.2.1.html> In fact it supports infinite security mechanism via plugins. This looks to be supported and bugs fixed and may be a good choice for data transfers.

We are moving toward using dCache as an intermediate transfer buffer for D0 to transfer data from enstore to remote sites. What exists now in dCache, and is being used by CDF, is a single stream non-secure ftp. Using GridFTP, even in single stream mode seems to be an improvement and we will discuss this with ISD, the group that built dCache.

2.10.5 JLab

Jefferson Lab has moved the J-SRM and related packages from development to early production use within the Lattice Hadron Physics Collaboration (LHPC, a part of the lattice QCD SciDAC project), as reported in the CS-4 section above. Early use is positive (after a few bugs were removed). At the end of this period the software is being tested between MIT and Jefferson Lab, and will soon be installed at the University of Maryland, and at ORNL to better support LHPC/SciDAC use of the ORNL supercomputers for lattice QCD. A machine at ORNL has been assigned as a J-SRM grid node, and will become operational as such in the next reporting period.

The Experiment side of the Jefferson Lab effort has focused on the maturation and deployment of our SRM layer and the commencement of the JDL project.

We have worked with Fermi and LBL, as described in section 2.4.3.2, toward interoperable client/server pairs based on the first SRM joint design specification. This interoperability effort has been useful both for debugging our independent implementations, and also as feedback into the upcoming second joint design, scheduled for December.

Florida State University uses the JLab SRM and disk cache to move CLAS data between their site and the JLab mass storage system. Our experience with this site, including the addition of components like GridFTP and a client-side disk cache, demonstrates the complexity in packaging and installing the SRM client tools. As a result of the growing client-side, we are embracing a sandbox-style installation that is more easily embedded within widely varying environments. The sandbox model allows rapid software distribution while maintaining a more regularized configuration. Current and upcoming deployments of the next release include Old Dominion University, The University of Glasgow, Louisiana Tech, and Carnegie Mellon University.

Andy Kowalski is coordinator of the Job Description Language effort described above in section 2.2.2.

2.10.6 STAR

During this quarter, the STAR experiment has concentrated efforts on finalizing and consolidating preceding projects and objectives but also has started to investigate new avenues and released Beta version of projects /tools started in the preceding quarter.

2.10.6.1 Deployment and tests of HRM-based File Replication

This project is currently lead by Eric Hjort and Doug Olson. Efforts by Dantong Yu, Eric and Shane Canon have been made to better understand some of the performance issues associated to using the HRM for STAR data transfer (between BNL and LBL). In early July, and after the internet connectivity upgraded from OC-3 to OC-12, Dantong Yu from the RCF demonstrated that the file transfer was currently limited by disk performance (see [OC-12.pdf](#)⁵ and [time.pdf](#)⁶ for example). However, we were saturating at a low 850 kB/sec per session with a maximum at 10.6 MB/sec for 14 sessions. Eric Hjort investigated this condition and we traced this to multiple factors amongst which, tuning of Sysconnect card on the respective Grid nodes. However, remaining network OC-3 connections should also be addressed.

The file replication software consisted of two HRMs installed at BNL-RHIC facility and at LBNL-NERSC facility, as well as a HRM-Command-Line-Interface (HRM-CLI) at NERSC. The HRM-CLI is the way a replication request is made. The actual file transfers from/to HPSS was done by the HRMs by staging and archiving file into/from their own disks, and using GridFTP for transfer between the HRM disks over the network. The main results of these long tests, running over many hours, are:

- the software was very stable performing without any crashes, and
- the bottleneck of file replication can vary depending on system and network loads on the source HPSS, the network, and the target HPSS.

⁵ <http://www.acf.bnl.gov/UserInfo/Facilities/Grid/OC-12.pdf>

⁶ <http://www.acf.bnl.gov/UserInfo/Facilities/Grid/time.pdf>

Regarding this last point, we observed that the HRM disks behaved as temporary buffers helping pre staging, and adjusting to varying network loads.

In addition, the SDM group had released a File transfer Monitoring tool supplied along with the HRM package. This has been extensively tested and feedback for possible improvements were provided to the developers. This work was the topic of a [PPDG news update](#)⁷.

2.10.6.2 Monitoring

a. Data transfer: Work with Eric Hjort on Star Data Transfer. We use GridFtp to Transfer data from BNL HPSS to NERSC HPSS. The overall data transfer rate is 8MB/Second. The bottleneck has been identified: there is only one tape drive involved in the data transfer and the average tape throughput is 10MB/second. BNL HPSS group upgraded the tape drive recently, and we expected that 30MB/second data throughput can be obtained after the upgrade.

b. Network performance tuning. After BNL's WAN connection was upgraded to OC12, we did lots of network test. Now we are able to transfer data from BNL to remote site at the aggregate speed of 50 to 60MB per second. We tested several gigabit network cards: 3COM and SysKconnect, intel, these cards are used in several experiments. Based on the test, we concluded that Intel gigabit network card has the best data throughput performance among these three types of cards, SysKconnect does not have good data throughput in LINUX environment, and 3COM always causes Kernel panic.

2.10.6.3 Job scheduling

The [STAR scheduler project](#)⁸, initially started and developed at Wayne State University is, since July, supported by Gabriele Carcassi at BNL. The goals presented in the preceding quarterly report are mainly reached.

- a) After integration of the job scheduler with our file catalog, we deployed a Beta version for the testers. In conjunction to our distributed data, we now feel we are on schedule as per the first objective of providing a completed tool for December/January.
- b) The deployment of the Beta version have allowed for useful feedback and further enhance the user interface. The next phase, deploying it to our PDSF site, will help further as we are seeking fast convergence and freezing of the interface. In addition, we are now in a good position to start the integration of job submission to the Grid.
- c) The software architecture and job description was further explained and presented at the [Grid PPDG-BNL site visit](#)⁹. According to suggestions, Gabriele Carcassi has since then joined the JDL group and provided feedback and comments on our approach and experience.

Our current plans for this project includes interfacing the scheduler with other queuing systems (Condor, PBS ...) and using in a Condor-G context to submit jobs on the Grid. The RCF have currently provided a Condor test bed and stress test have begun.

2.11 CS-11 Grid Interface with Interactive Analysis Tools

The final version of the [Grid Service Requirements for Interactive Analysis](#)¹⁰ was published as PPDG-19. Since that time the analysis tools prototyping has continued in parallel by three groups: CMS GAE, Slac JAS and Atlas DIAL.

2.11.1 Clarens

A Root-based analysis of simulated JetMET data for the CMS experiment was used to demonstrate remote analysis at the iGrid2002 conference held in Amsterdam.

⁷ <http://www.ppdg.net/docs/news/news-update-star-hrm-25sep02.pdf>

⁸ <http://www.star.bnl.gov/STAR/comp/Grid/scheduler/>

⁹ <http://www.star.bnl.gov/STAR/comp/meet/PPDGSept2002/>

¹⁰ http://www.ppdg.net/pa/ppdg-pa/idad/papers/analysis_use-cases-grid-reqs.pdf

This demonstration was meant to show the usefulness of high-bandwidth networking in a grid-based analysis environment. An on-screen display of analysis results based on data streamed from a remote grid site formed the mainstay of the demonstration. Data was streamed from a computing cluster located at the Starlight network point-of-presence in Chicago to the showroom floor using a 10-Gbit/s link using the Globus toolkit. Due to problems with the networking layer at the sending end, the actual transfer rate was limited to 100Mbit/s.

The concept of remote analysis was, however, successfully demonstrated and generated a positive response from conference-goers. The demonstration is part of a documentary being produced by the conference organizers.

In the past quarter development was focused on deployment and testing of the Clarens server and client software. To this end, the server has now been deployed at the following institutions: - Caltech (6 machines, including the testgrid and tier 2 machine) - Starlight POP (6 machines to be used in the SC2002 demonstrations)

- UCLA (1 machine)
- University of Florida (2 machines)
- Fermilab (2 machines)

The Clarens server was also deployed for the first time on a Sun Solaris-based machine located at CERN, Geneva.

Streamlining of the installation procedure was done in response to user input, resulting in a set of binary packages and an automated post-installation setup and configuration mechanism to be developed.

Deployments of the Clarens server on the WAN made testing of the efficiency of remote analysis possible, resulting in an enhancement to the Root-based Clarens client to include configurable caching and network traffic optimizations. These enhancements reduce the impact of network latencies in the analysis process, so that analysis of remote files become nearly indistinguishable from analysis of local on-disk files when the network bandwidth is comparable with disk I/O bandwidth.

The Clarens web pages at <http://clarens.sourceforge.net> were also expanded to include more documentation and usage information, as well as binary packages of the Root client for the first time.

Preparations for the Fermilab/SLAC demonstration at the SC2002 conference to be held in November were also carried out during this time.

2.11.2 ATLAS DIAL

David Adams PPDG-related efforts are focused on providing a collective view of event data and a framework for interactive analysis of such data.

The data collections are called datasets and are described at <http://www.usatlas.bnl.gov/~dladams/dataset>.

The first version of the system was completed. It includes an interface for datasets (Dataset) and implementations for ATLAS ROOT Monte Carlo files (TDatasetFileDataset) and merging events (EventMergeDataset).

The dataset interface is intended to be generic, i.e. applicable to any experiment, but my main interest is ATLAS and I did work to integrate this model with the data collections envisioned for ATLAS and LCG. I hope to have these issues resolved by the end of the next quarter.

A first release of the infrastructure for DIAL was completed (Distributed Interactive Analysis of Large datasets). The major components are now in place including a simple scheduler which starts tasks on the local machine. See <http://www.usatlas.bnl.gov/~dladams/dataset> for details.

The next steps are to distribute the scheduler so that tasks can be submitted to remote nodes and to add a component which makes use of GRID tools to locate data and compute resources and do matchmaking. We expect to complete a version of the former in the next quarter and begin on the latter early the next year.

2.12 CS-12 Catalogs and Databases

2.12.1 STAR file metadata catalog

In preparation for integrating STAR data analysis and processing with grid services there is work being done on the data management infrastructure, specifically putting a new file metadata catalog, developed by Jerome Lauret, into production. This [metadata catalog](#)¹¹ is based on MySQL. Perl modules and command line interface were provided to users, we gathered feedback, improved and extended features not thought of in the initial design and considerably improved the speed of most queries (by a factor of 5). We also worked on integrating the file catalog UI with the [DataCarousel](#)¹², a tool initially developed to satisfy user's need for restoring lists of files stored in HPSS to disk. The implementation was made to achieve three main objectives :

- a) Stress testing multiple and simultaneous insertion of entries in our file catalog. For this, we had a 20 nodes test bench with an average of 50 hits per minutes to the catalog (with a worst case scenario picking at 20 simultaneous hits). Our initial observation made us revisit our record insertion (using delayed or low priority insertions) since otherwise, MySQL mechanism of locking tables for rebuilding the index would become problematic. No further problems were encountered after modifications and we restored half of our last year data summary on distributed disk (about 1 TB of data spread over 50,000 files) . Independent MySQL stress test were made by Gabriele Carcassi (see [MySQL stress test web documentation](#)¹³) : the objective' was to make a scalability assessment as per using MySQL as the database software for our file catalog.
- b) Achieving a proof of principle as per the UI to the catalog and its ability to provide to users a way to unambiguously request files according to their physical location (node by name, by path, etc...), logical location (type of storage) or logical collections (data set, production condition, or any meta-data queries). We also tested our design implementation for supporting multiple sites from within the same database and did not find any problem.
- c) Providing a backbone for the deployment of our job scheduling project and, in general, a base tool for any further Grid tool deployment relying on the ability to locate data and logical data sets on the Grid via an independent UI or package.

Further file catalog developments and perspectives includes implementing our current catalog to our LBL site and using it in production mode for our next run period. The goals and objectives of this apparently non-Grid approach are numerous for STAR and we would like to stress again its main motivation : the possibility we have now to further develop and enhance tools such as job scheduling, up-to-date file transfer cataloging and distributed analysis tools .

Finally, we are also investigating the integration of our catalog with MAGDA along with help from Wensheng Deng. We hope to have this investigation and implementation completed in the next quarter and to benefit from this interaction as per the considering features foreseen by other groups.

3 Single Collaborator Reports

3.1 ANL – Globus

3.1.1 Coordination and Support

NOTE: RLS, MDS and RFT work are listed in the Common Services sections.

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

¹¹ <http://www.star.bnl.gov/STAR/comp/sofi/FileCatalog.html>

¹² http://www.star.bnl.gov/STAR/comp/sofi/carousel/data_carousel.htm

¹³ <http://www.star.bnl.gov/STAR/comp/Grid/fileCatalog/MySQLStress.html>

General PPDG support

- Participation in general PPDG meetings, including
 - August - D0 Meeting at Fermi
 - August - DOE High Perf Networking Meeting - DC
 - August - HPSS/GridFTP integration design meeting
 - August - Phone call with IBM HPPS development team to work out HPSS/GSI integration issues.
 - August - NorduGrid Meeting, Copenhagen
 - July, Trillium management meeting, Fermi
 - July - GGF5, Edinburgh
 - July - GLUE schema meeting, Edinburgh
 - June - Bill Allcock meeting with Arie Shoshani to discuss GridFTP issues
 - June - Bill Allcock meeting with Reagan Moore and SRB team to discuss GridFTP integration with SRB
 - Architecture planning calls with Arie, Miron

ATLAS- continuing support

- Weekly atlas testbed calls
- Approx. 6 ATLAS-CHANL meetings in this period
- US Atlas Grid Workshop, BU – June
- Meeting with SC demo coordinators, August, Boston
- Meeting with John Huth – Atlas Grid Planning, September, Boston

CMS- continuing support

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

SiteAAA work

- Bi-weekly phone calls

3.1.2 Release of Globus Toolkit 2.2

This quarter Globus released its latest version of the Globus Toolkit, 2.2. This release includes GRAM 1.6, an updated version of the GRAM protocol, a rewrite of the GASS-cache, and MDS 2.2, which includes information providers that meet the GLUE-schema specification. In addition, major bug fixes are included for GridFTP, MDS and buffer overflow repairs to the GSI code.

Details can be found at <http://www.globus.org/gt2.2/release-notes.html>.

3.1.3 Planned development work

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

- Work towards the authorization interface for CAS that is now being discussed
- Finish evaluation of the GridFTP performance

- Continued RFT development to stay in agreement with the still changing OGSA spec
- Extended support in RFT for queueing service (next 6 month time frame)
- Additional information providers for MDS 2.2
- Demos of GT 2.2, GT3, RFT, and CAS are planned for SC '02 (November)

3.2 SDSC – SRB

A description of work on SRB carried out in collaboration with Caltech is given above under storage management (section 2.4).

SDSC continues to provide SRB support for the BaBar experiment.

Additional infrastructure development is supported by both the NSF GriPhyN project, the NSF NPACI project, and the NSF Distributed Terascale Facility project. These efforts are developing technology that will be used within the PPDG data handling environment. The tasks include:

creation of a GridFTP driver for the SRB. This makes it possible for the SRB to access data that is located on a storage repository which supports a GridFTP daemon. The SRB functions that are interfaced to the GridFTP daemon were developed by Bing Zhu. The functions allow a SRB daemon to use its long-lived GSI certificate to store file(s) and retrieve file(s) from a remote grid-ftp daemon. The functions are listed in document PPDG-21 i

- Porting of the SRB metadata catalog to the PostgreSQL database. The port is completed and is in being tested for functionality. The port was done by Craig Lee of the Aerospace Corp. for NASA.

Development of WSDL interfaces to the SRB data handling system. This is an ongoing activity, with first interfaces expected in 1QFY03.

Papers 11th HPDC conference in Edinburgh, Scotland, "MySRB & SRB - Components of a Data Grid", A. Rajasekar, M. Wan, R. Moore, July, 2002.

A data grid was demonstrated across four United Kingdom sites at the HPDC conference.

3.3 STAR

Most STAR work is described above in sections 2.4 and 2.10.6.

We mentioned working with Atlas for investigating the use of MAGDA in STAR. The investigation is too premature to bring conclusions to this report but we would like to mention the good response and collaboration between the two participating experiment.

In early May, we also investigated the possibility to have MySQL databases replicated over the Grid (i.e. Either using Grid tools or authentication mechanism). At the time, MySQL development was emerging (version 4.x) with support for SSL and X509 authentication. The initial test we did, did not lead to success. However, in the light of recent independent feedback, we will resume this work in the next quarter. Currently, STAR has developed a tool to monitor remote MySQL slave servers.

We are considering investigating the implementation of a Ganglia base monitoring approach in the next quarter. Experience in Ganglia is already strong at BNL and we hope to benefit from a unified approach.

We have tried to initiate discussions and communications across experiments and the Rhic Computing Facility. Our main interest was to initiate opened discussions on frontier technologies approaches and tools in use in other groups and have a regular meeting where issues can be discussed and/or presented. In early September, the [Technology Meeting](#) was instated at BNL (many thanks to the RCF and Atlas Computing group for making this possible).

3.4 . Wisconsin – Condor

The Condor-project is working in various areas in support of PPDG experiments:

- Support for D0 Planner through extensions to ClassAds and Condor-G.
- Coordination of PPDG-20 - joint document on status of grid planning between EDG WP1, D0 and Condor project.
- Work to make the US CMS testbed a more production service and successfully submitted >3,000 jobs (600 jobs each with about 5 daqs?) with more than 95% running to completion - albeit with retries, automatic resubmittal etc. numerous investigations of failure modes.
- Work on CMS MOP extensions and operations of the CMS testbed. Identifying problems in Condor and globus layers and working through these to get a working system. Also took vacation.
- Working on extensions and fixes to DAQMAN for higher level planners in Chimera.

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	N	GriPhyN coordinator					x					x		
	Jenny Schopf	Y	GriPhyN, iVDGL, Globus team liason, ATLAS-CS liason			x				x	x		x		
	William Alcock	Y							x		x		x		
	Von Welch		CAS									x			
	Stu Martin				x									x	
ATLAS	John Huth	N	ATLAS Team lead											x	
	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								x	
	BaBar	Richard Mount	N	PPDG PI, BaBar Team co-Lead											
Tim Adye		N	BaBar Team Co-Lead												
Robert Cowles		N											x		
Andrew Hanushevsky		Y						x	x						
Adil Hassan		Y						x	x						
Les Cottrell		N	IEPM Liaison			x									
Wilko Kroeger		Y						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
	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	Y	job execution, grid monitoring		x	x									

Globus/ISI	Carl Kesselman	N	Globus/ISI lead																
	Ann Chervenak	Y								x									
SiteAAA (if not listed above)	Dane Skow	Y	FNAL, Co-Lead															x	
	Bob Cowles		SLAC, Co-Lead																
	Booker Bense	Y	SLAC															x	
	Tomasz Wlodek	Y	UT Arlington															x	
	Igor Mandrechenko	Y	FNAL															x	
	Steve Chan,		LBNL															x	
	Steve Lau		LBNL															x	
	Craig Tull		LBNL															x	
	Matt Crawford		FNAL															x	
	David Bianco		JLab															x	
	Hao Wang		U Wisconsin															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 Meetings

Jul 3	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> URL: http://www.ppdg.net/mtgs/phone/020703/default.htm
Jul 11	12 p.m. - 2 p.m.	<u>PPDG-Site-AAA bi-weekly meeting</u>
Jul 31	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> URL: http://www.ppdg.net/mtgs/phone/020731/default.htm
Aug 1	12 p.m. - 2 p.m.	<u>PPDG-Site-AAA bi-weekly meeting</u> URL: http://www.ppdg.net/pipermail/ppdg-siteaa/2002/msq00409.html
Aug 7	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> URL: http://www.ppdg.net/mtgs/phone/020805/default.htm
Aug 14	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> URL: http://www.ppdg.net/mtgs/phone/020814/default.htm

Aug 15	12 p.m. - 2 p.m.	<u>PPDG-Site-AAA bi-weekly meeting</u> URL: http://www.ppdg.net/pipermail/ppdg-siteaa/2002/msg00482.html
Aug 21	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> URL: http://128.3.182.66/mtgs/phone/020821/default.htm
Aug 28	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> URL: http://www.ppdg.net/mtgs/phone/020828/default.htm
Aug 29	12 p.m. - 2 p.m.	<u>PPDG-Site-AAA bi-weekly meeting</u> URL: http://www.ppdg.net/pipermail/ppdg-siteaa/2002/msg00500.html
Sep 4	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> URL: http://www.ppdg.net/mtgs/phone/020904/default.htm
Sep 12	12 p.m. - 2 p.m.	<u>PPDG-Site-AAA bi-weekly meeting</u> URL: http://www.ppdg.net/pipermail/ppdg-siteaa/2002/msg00518.html
Sep 18	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> URL: http://www.ppdg.net/mtgs/phone/020918/default.htm