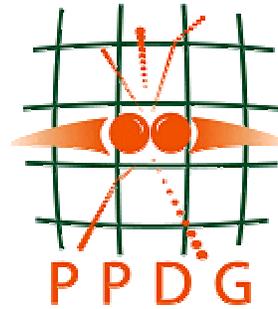


**Particle Physics Data Grid  
Collaboratory Pilot**

**Quarterly Status Report of the  
Steering Committee,  
October - December 2001**

31 Jan 2002



1 Project Overview.....	2	4.4 D0.....	22
1.1 Highlights.....	2	4.5 Jlab.....	23
1.2 Project Management and Organization ....	3	4.6 STAR .....	23
1.3 Plans for the next Quarter.....	3	4.7 ANL – Globus.....	24
1.4 Summary of progress on Common Services Development, Integration and TestBeds .....	3	4.8 NERSC – SDM.....	25
1.5 Year 1 Status .....	6	4.9 SDSC – SRB.....	26
1.6 Interactions with other Projects and Activities.....	10	4.10 . Wisconsin - Condor.....	27
2 Project Activities.....	11	5 Appendix.....	28
2.1 GDMP (CMS-DataGrid-Globus).....	11	5.1 List of participants.....	28
2.2 D0 Job Management (D0-Condor) .....	12	5.2 SuperComputing 2001 demonstrations related to PPDG .....	30
2.3 CMS-MOP (CMS-Condor).....	12	5.3 International HENP Grid Coordination and Joint Development Framework .....	31
2.4 STAR-DDM (STAR-LBNL/SDM) .....	13	5.4 Appendix - PPDG Meetings .....	35
2.5 JLAB-Replication (JLAB-SRB) .....	13		
2.6 ATLAS distributed data manager, MAGDA (ATLAS-Globus) .....	14		
2.7 BaBar Database Replication (BaBar-SRB) .....	15		
3 Cross-cut Activities and Collaborations.....	16		
3.1 SuperComputing 2001 Demos.....	16		
3.2 Certificate/Registration Authority.....	16		
3.3 Monitoring.....	17		
3.4 Collaboration with IEPM, Network Performance Monitoring.....	17		
4 Single Collaborator Efforts and End to End Applications.....	18		
4.1 ATLAS.....	19		
4.2 BaBar .....	19		
4.3 CMS.....	21		

# 1 Project Overview

## 1.1 Highlights

These highlight summaries of PPDG work were included in the News Update<sup>1</sup> of 29 January 2002.

### **CMS-MOP**

MOP is the MOnTe carlo distributed Production for the CMS experiment that manages the generation of simulated data for CMS physicists. It was demonstrated at SC2001 running jobs at FNAL, SC2001, UCSD, Caltech and U. Wisconsin. It is accepted as a fundamental component of the US-CMS Grid Testbed. It relies upon Condor/G, the Globus Toolkit and GDMP, and is an application that will integrate additional job management functionality as this is developed within PPDG.

### **ATLAS-Magda**

Magda is the grid data manager for the ATLAS experiment that permits physicists to browse, access, replicate and publish data around the world (US-Europe). It is deployed for the Data Challenge 0 which began in December 2001. Magda utilizes portions of the Globus Toolkit now (GSI, GridFTP) and integration with additional features of Globus (replica catalog, replica management services) is planned.

### **X.509 PKI RA – DOESG CA**

As the datagrid efforts of the physics experiments in PPDG were getting underway we became painfully aware of missing an acceptable public key infrastructure (PKI) for issuing the X.509 certificates necessary for all grid activities. The certificate authorities (CA) used by the grid middleware developers were not acceptable to the production computing sites the experiments are using in the US and Europe. ESnet was approved to set up and operate a PKI in the doesciencegrid.org domain which provides an acceptable level of security. Agreement was reached in December 2001 with members of the EU-Datagrid project that certificates issued under doesciencegrid.org would be accepted, an issue of critical importance to all the HEP experiments, whose grid dimensions span the Atlantic.

PPDG has set up a prototype registration authority (RA) to handle the approval process for certificates issued to the US HENP community, see [www.ppdg.net/RA/](http://www.ppdg.net/RA/). As of January 2002 the doesciencegrid.org certificate authority has begun issuing the first certificates. The exercise of having one of these certificates installed in Europe for a transatlantic test is underway.

### **GDMP**

The Grid Data Mirroring Package is a joint project between PPDG and the European Data Grid (EDG). It extends the Globus replica management to support publish-subscribe data file replication. GDMP is part of the integrated EDG TestBed 1.

### **SC2001 Demonstrations**

There were numerous examples of grid enabled applications shown at SuperComputing 2001 in addition to CMS-MOP mentioned above. The complete list demonstrations of PPDG participants is listed in the table in the appendix, Section 5.2, and also at [http://www.ppdg.net/docs/presentations/sc2001\\_demos.htm](http://www.ppdg.net/docs/presentations/sc2001_demos.htm).

---

<sup>1</sup> <http://www.ppdg.net/docs/news/news-item-29jan02-b.pdf>

## 1.2 Project Management and Organization

The ramp up in staffing the PPDG positions continued both in the experiment and computer science groups. Lee Lueking, co-project leader of the D0 SAM data access system, replaced Ruth Pordes as the PPDG D0 Team Lead and joined the Steering Committee (PPDG-SC). The SC mail list was extended to include our DOE sponsors and liaisons to the other major US HENP grid projects. This is a help in ensuring good ongoing communication without us remembering to have to do it! PPDG phone meetings became more or less weekly - with Doug Olson providing reliable postings of booking, agendas and minutes (<http://www.ppdg.net/cgi-bin/we4.0/webevent.cgi?cmd=openca&cal=cal2>).

The Steering Committee held 2 phone meetings, and the executive team held regular discussions on the progress of the project.

## 1.3 Plans for the next Quarter

During the next quarter we need to start to plan for the next years activities and deliverables. This will be a focus of the executive team, the steering committee and part of the collaboration meeting. The PPDG effort should now be ramped up to full complement and first year deliverables as defined in the proposal well underway

## 1.4 Summary of progress on Common Services Development, Integration and TestBeds

PPDG projects and tasks are categorized by CS number (“Categories and Subjects” or “Common Services” or ...). The details of the projects and deliverables are maintained by the project leaders of the various activities, however a central plan and strategy needs to be maintained. Following up on the specific Project Activities – their design, implementation, deliverables and schedule – is an area that PPDG coordination has not been sufficiently proactive on. We will be exploring ways of improving our professionalism in this area over the next 6 months.

The list of CS topics from the proposal has been augmented based on the activities thus far with the addition of CS-8,9,10:

CS#	From Proposal
1	Job Description Language
2	Scheduling and Management of Processing and Data Placement Activities
3	Monitoring and Status Reporting
4	Storage Resource Management
5	Reliable Replica Management Services
6	File Transfer Services
7	Collect and Document Current Experiment Practices
8	R&D , Evaluations
9	Authentication, Authorization and Security
10	End-to-End Applications and TestBeds

#### **1.4.1 CS-1 Job Description Language**

No formal Project Activity, Project Effort or Cross-Cut Project has been started for this work. Individual experiment and CS groups have working implementations at different stages of sophistication and development:

CONDOR – Class-Ads, Data Placement (DaP) Jobs

Globus – RSL, GRAM

SAM – “SAM submit” interfaced to several different batch systems

CMS – RES – fault tolerant features of job production

#### **1.4.2 CS-2 Scheduling and Management of Processing and Data Placement Activities**

Nearly all experiments are currently developing or have working prototypes of distributed job production using extensions of existing batch and experiment specific distributed production software which incorporate initial components of common middleware. Details of the work is reported in the individual experiment sections. We plan to have a focus meeting on Job Description Languages and Distributed Job Production in the near future. This will give us an opportunity to review the existing middleware, the development path of the computer science groups and the plans of the experiments for the next year, and understand any opportunities for commonality.

#### **1.4.3 CS-3 Monitoring and Information Services**

As reported below in the Globus report, PPDG is participating in a joint GriPhyN/PPDG Monitoring working group. In fact both co-leaders of the group are funded by PPDG. This project has gathered use cases from a number of experiments and is using these as a basis from which to start developing an understanding of the requirements – functionality and performance. The group is also working on a roadmap which will be discussed at the PPDG collaboration meeting in February.

Many groups DOE, NSF and other research groups are developing and deploying networking monitoring and prediction capabilities. Such services are essential for achieving the high throughput, scalability and robustness required by the PPDG experiment applications. We must position ourselves to take advantage of these developments by understanding clearly our needs, the deliverables and schedules of each of these groups, and ensuring that there is a common framework through which the various monitoring information can be made accessible for prediction and analysis.

Additionally, the PPDG monitoring effort must include the capture and dissemination of monitoring information from all components of the compute fabric such as storage elements, compute systems, code and meta-data repositories.

#### **1.4.4 CS-4 Storage Resource Management**

The collaboration between the Storage Resource Management SciDAC project and PPDG has shown results in the delivery of a document which starts to define a standard interface to a Storage Resource from the Grid fabric (see PPDG-9). The JLAB, LBL and SRB teams are working together to extend and refine this document. It is clear that we need to understand the synergy between this document and the GridFTP RFC. The PPDG executive team plans some phone conferences to discuss this in the next quarter.

#### **1.4.5 CS-5 Reliable Replica Management Services**

The Globus Replica Management services have been released as part of Globus V2.0 and are incorporated into the European Test Bed V1.0 release, as well as the expected V1.0 release of the GriPhyN Virtual Data Toolkit. The SRB replication system has been used in prototype mode by BaBar. The SAM file replication system has been further enhanced to handle multi-stage routing of data files as they are stored or delivered.

### 1.4.6 CS-6 File Transfer Services

ATLAS, CMS and JLAB are using gsiFTP – the pre-released version of GridFTP – for data movement for simulation production. STAR is using GridFTP for the transfer of data files from BNL to LBL. BaBar's bbscp has been extended to include more sophisticated algorithms for obtaining the highest throughput end-to-end transfer of data files.

### 1.4.7 CS-7 Documents, Reports and Meetings

#### 1.4.7.1 CS7-1 Meetings and Workshops

There were no PPDG face to face meetings this quarter although there were significant numbers of PPDG participants attending GGF3 held at INFN Frascati in October and Supercomputing 2001 in November.

A list of the teleconference meetings is given in the Appendix, section 5.4.

The first "Focus Meeting" on Robust File Replication is scheduled for January 10<sup>th</sup> 2002. A second Focus Meeting on Grid Job Scheduling is planned for the end of the first quarter of 2002. A monitoring Focus Meeting will be scheduled in the second quarter in collaboration with GriPhyN. The February PPDG collaboration meeting will be held in Toronto immediately following GGF4.

#### 1.4.7.2 CS7-2 PPDG Document Series

The PPDG management plan was published. The PPDG Steering Committee agreed to a joint Data Grid Reference Architecture document in collaboration with GriPhyN.

PPDG-11	Robust File Replication Focus Meeting Report	
PPDG-10	Numeric Requirements for the Replica Catalog Service	V0.2
PPDG-9	Common Storage Resource Manager Operations	V1.0
PPDG-8	Data Grid Implementations - Comparison of Capabilities, R.Moore et al	V6
PPDG-7	PPDG Management Plan	<a href="#">V1.00(10/10/01)</a>
PPDG-6	<a href="#">Joint GriPhyN-PPDG Data Grid Reference Architecture</a>	
PPDG-5	<a href="#">CMS long term GRID requirements document for EU DataGrid, GriPhyN, and PPDG</a> , K. Holtman	
PPDG-4	<a href="#">Sam and the Particle Physics Data Grid (doc)</a> , V. White	9/01
PPDG-3	<a href="#">Year1 Project Plan (doc)</a>	9/01
PPDG-2	For SciDAC: <a href="#">Overview</a> , <a href="#">Application</a> and <a href="#">Collaboration</a> tool surveys	8/01
PPDG-1	<a href="#">PPDG Update</a> at DG coordination meeting, Rome ( <a href="#">pdf</a> , <a href="#">ppt</a> )	6/01

#### 1.4.7.3 CS7-3 Quarterly Reports

You are reading one.

#### 1.4.7.4 CS7-4 Architecture

PPDG is collaborating with GriPhyN on documenting the Data Grid Architecture with the current revision of the document being 2.09 (DGRA, PPDG-6 [http://www.ppdg.net/docs/documents\\_and\\_information.htm](http://www.ppdg.net/docs/documents_and_information.htm)). Extensions/addenda to this that address the PPDG scope of end-to-end applications may be necessary. It is expected that a technical comparison of the EDG and GriPhyN/PPDG architectures may be useful in the next year or two.

#### 1.4.7.5 CS7-5 Project Web Pages

Extensions to the project web pages continue, group web sites were modified to have some modicum of consistency.

#### 1.4.8 CS-8 Evaluations and Research

#### 1.4.9 CS-9 Authentication, Authorization and Security

While discussions of authorization and authentication are never really separate, the mechanisms for implementing them are distinguished and one can address the implementations separately, if not independently. In the future we expect PPDG to get involved with the technology for implementing authorization while to date PPDG effort has only been addressing the authentication process, as described below.

##### 1.4.9.1 CS9-1 Establish Certificate Authority for use by PPDG

There were a sequence of discussions over the past year with DOE Science Grid people about using a certificate authority (CA) in the doesciencegrid.org domain. In September 2001 ESnet received approval and funding to set up and operate a PKI infrastructure which PPDG will be able to use for getting X.509 identity certificates. Tony Genovese and Mike Helm of ESnet are setting up this infrastructure and the CA should begin initial operations in January 2002.

##### 1.4.9.2 CS9-2 Establish PPDG Registration Authority

Part of the responsibilities of a certificate authority can be delegated to an entity called a registration authority (RA). The main role of an RA is to carry out the actual identity check for individuals who request a certificate. Following discussions with Tony and Mike, it was decided that PPDG would set up and operate an RA to handle these requests from the PPDG community. The implementation plan for this is described below in detail and this should begin initial operation along with the CA in January 2002.

##### 1.4.9.3 CS9-3 PPDG Certificate Policy

An essential factor in determining whether or not someone trusts the certificates issued by a CA is the policy by which the CA operates. The Certificate Policy and Certificate Practice Statement (CP/CPS) for the doesciencegrid.org is available at <http://envisage.es.net>.

#### 1.4.10 CS-10 End-to-End Applications and TestBeds

All experiments in PPDG are maintaining working end to end simulation and/or data processing and analysis frameworks. ATLAS and CMS have deployed prototypes and demonstrators over the last quarter using Grid components of distributed data replication and job submission. They continue their work to move these into the production versions of the US and eventually the experiment wide data handling systems.

### 1.5 Year 1 Status

#### 1.5.1 Status of deliverables from Year 1 Project Plan

##### Project Activities

Project Activity	Experiments	Yr1	Status 1/1/02
------------------	-------------	-----	---------------

CS-1 Job Description Language – definition of job processing requirements and policies, file placement & replication in distributed system.			
CS1-1 Job Description Formal Language	D0, CMS	X	
CS1-2 Deployment of Job and Production Computing Control	CMS	X	Prototype demonstrated at 4 sites
CS-2 Job Scheduling and Management - job processing, data placement, resources discover and optimization over the Grid			
CS2-1 Pre-production work on distributed job management and job placement optimization techniques	BaBar, CMS, D0	X	CMS – prototype demonstrated as above. D0 – SAM in use by the collaboration.
CS-3 Monitoring and Status Reporting			
CS3-1 Monitoring and status reporting for initial production deployment	ATLAS	X	Initial sensors deployed and MDS installed.
CS3-2 Monitoring and status reporting – including resource availability, quotas, priorities, cost estimation etc	CMS, D0, JLab	X	CMS – unfunded effort working on sensing agent infrastructure. D0 – SAM in use with initial implementation.
CS-4 Storage resource management			
CS4-1 HRM extensions and integration for local storage system.	ATLAS, JLab, STAR	X	STAR. JLAB – working with HRM group at LBL on implementation ATLAS – ? some words here?
CS4-2 HRM integration with HPSS, Enstore, Castor using GDMP	CMS	X	First versions working with LBL HRM/DRM implementation and EDG WP2
CS-5 Reliable replica management services			
CS5-1 Deploy Globus Replica Catalog services in production	BaBar,	X	Prototype demonstration at SC2001
CS5-2 Distributed file and replica catalogs between a few sites	ATLAS, CMS, STAR, JLab	X	Atlas, CMS - simulation production using initial file catalogs and GDMP. JLAB – grid portal prototype implemented. STAR – using LBL HRM interface, gridFTP and own file catalogs
CS-6 File transfer services			
CS6-1 Reliable file transfer	ATLAS , BaBar, CMS, STAR, JLab	X	

CS-7 Collect and document current experiment practices and potential generalizations	All	X	Several documents have been delivered.
--	-----	---	--

**Project Efforts:**

Collaborator	Title	Schedule	Description
Atlas	Tool for distributed data services - MAGDA	8/1/01 <i>Done</i>	Transition to a Project Activity with Globus - integration of Globus replica catalog , GDMP and Globus replica management services.
BaBar	Intra-site data replication	10/1/01 <i>Done – demonstrated at SC2001</i>	Prototype a file replication system tailored to the BaBar objectivity dataset files. Initially this application will be deployed local to SLAC for performance and robustness tests. This work will follow on from tests with the SRB MCAT catalog and file replication service, and file request redirection techniques using HTTP redirection and the Globus LDAP replica catalog. Following this the effort will transition to a Project Activity to develop Inter-site replication between SLAC and IN2P3.
CMS/Caltech	Remote data analysis prototype	<i>Done – demonstrated at SC2001</i>	Remote data analysis using JAS and GSI authentication
D0	SAM Information Services and Test Harness	12/1/01 <i>Test harness extended; performance analysis started.</i>	Provide common status and information services throughout the SAM system. Present and analyze information to understand performance and availability aspects of the system. Complete Test Harness application to allow simulation of system, stimulation of error conditions, and configuration of system parameters to the boundary conditions.
CMS/Caltech	Partitionable Execution Service for Distributed Processors - TQS	<i>Start of integration with MOP/Condor</i>	High Availability , Fault Tolerant Job Queuing service
CMS/Caltech	Optimized database query tags for Tier 2 data set	<i>Prototype done</i>	In collaboration with GriPhyN - Simulate the grid environment with Belief Desire Intention (BDI)-based software agents. Test different algorithms for optimising query planning, execution and long-term load balancing of the data grid
LBNL	Disk Resource Manager – DRM – IDL and prototype implementation	12/1/01 <i>Done as part of the Project Activity with STAR</i>	In collaboration with the Resource Management for Data intensive Grid Applications SciDAC project.

**Documents:**

Document	Schedule	Editor
Year 1 Project Plan (this document)	Review: 8/15/01; Final: 10/15/01 <i>DONE</i>	Ruth Pordes
Replication Requirements		
Replication Use Cases	In progress	Mike Wilde
Storage Management Experiment Use Cases	Comparison of data grid products. <i>DONE</i>	Reagan Moore

## 1.5.2 Current Issues and Concerns

### 1.5.2.1 Planning for the Second Year of the Project

During the next quarter the Steering Committee needs to start defining the goals and scope of PPDG for the 2<sup>nd</sup> and 3<sup>rd</sup> year. The impact and success thus far of cross-project efforts is something that we can hope to build on for the future.

### 1.5.2.2 Those “Extra” deliverables – education, outreach and vendor contacts

Summer 2002 is a time when we need to implement part of the education and outreach tasks in the PPDG proposal. Planning for this will start at the Steering Committee meeting at the February collaboration meeting.

### 1.5.2.3 Meeting and Schedule Overload

PPDG regards as important participation and communication at the many HENP Grid, experiment meetings, cross-cut and coordination meetings, workshops and conferences. These contributions do however, take significant effort from the PPDG leadership. This year the project has been sparing in the number of face-to-face meetings we have arranged. We need to ensure this is not to the detriment of establishing the best cross-experiment and cross-CS group common work and developments. Starting this next quarter we plan to arrange more “cross project technical focus meetings” to try and mitigate this risk.

### 1.5.2.4 Effort

The PPDG annual funding (~\$3.0M) supports 15-20FTEs, of which 2 x .5 FTEs are the coordinators, 2-3 FTEs for each of 4 experiments, 1 FTE for each of 2 experiments, 1 FTE for each of 3 computer science groups, and 2-3 FTEs for the 4<sup>th</sup> computer science group = 16-21 FTEs. Many of the individuals on PPDG are also of course contributing directly to the experiment or computer science group of which they are a member, or on other grid projects such as GriPhyN. This reduces the risk of divergence and increases the possibility of good communication across the projects.

As always we have underestimated the amount of effort required for integration and deployment of end-to-end applications, installing and configuring hardware and software. The complete set of PPDG effort could with justification be deployed to work with the data handling groups to bring Grid middleware to production use in the participating experiments. Fortunately the PPDG teams have brought to the collaboration unfunded effort which is fully integrated into the project and allows some opportunity for the project to develop extensions to the middleware, and understand and code the numerous “glue” components that are always necessary to integrate heterogeneous components and make an end-to-end application that in practice works.

The list of participating individuals (table in Appendix 5.1) shows clearly that most individuals are not working full time on PPDG as funded resources:

## 1.6 Interactions with other Projects and Activities

### 1.6.1.1 SciDAC PI Meeting

Miron gave a presentation <http://www-fp.mcs.anl.gov/fl/accessgrid/doenc-ppts/livny.ppt> at the SciDAC National Collaboratories PI's roundtable meeting in November. This was a useful meeting in giving an overall perspective of the set of SciDAC collaborative projects and understanding of their overlap and differences. Following this meeting we submitted the following list of contacts from PPDG to the other Collaboratory projects:

Earth System Grid II	Arie Shoshani	<a href="mailto:arie@lbl.gov">arie@lbl.gov</a>
Collaboratory for Multi-Scale Chemical Science	Richard Mount	<a href="mailto:richard.mount@slac.stanford.edu">richard.mount@slac.stanford.edu</a>
National Fusion Collaboratory	Doug Olson	<a href="mailto:dlolson@lbl.gov">dlolson@lbl.gov</a>
DOE Science Grid	Doug Olson	
Pervasive Collaborative Computing Environment	Miron Livny	<a href="mailto:miron@cs.wisc.edu">miron@cs.wisc.edu</a>
Reliable and Secure Group Communication	Miron Livny	
A High-Performance Data Grid Toolkit	Ian Foster	<a href="mailto:foster@mcs.anl.gov">foster@mcs.anl.gov</a>
Middleware Technology to Support Science Portals	Miron Livny	
CoG Kits	Ian Foster/Mike Wilde	<a href="mailto:wilde@mcs.anl.gov">wilde@mcs.anl.gov</a>
Scientific Annotation Middleware	Reagan Moore	<a href="mailto:moore@SDSC.EDU">moore@SDSC.EDU</a>
Storage Resource Management for Data Grid Applications	Arie Shoshani	
Middleware to Support Group to Group Collaboration	Harvey Newman	<a href="mailto:newman@hep.caltech.edu">newman@hep.caltech.edu</a>
Distributed Security Architectures	Doug Olson / Mary Thompson	
Security and Policy for Group Collaboration	Doug Olson / Von Welch	
National Computational Infrastructure for Lattice Gauge Theory	Chip Watson	<a href="mailto:watson@jlab.org">watson@jlab.org</a>

A meeting of all SciDAC PI's was held in January 2002 and both a poster and presentation were prepared for this meeting. These are available as:

[http://www.ppdg.net/docs/ppdg\\_poster\\_OL.pdf](http://www.ppdg.net/docs/ppdg_poster_OL.pdf) (poster) and  
[http://www.ppdg.net/docs/presentations/mount\\_scidac\\_jan02.ppt](http://www.ppdg.net/docs/presentations/mount_scidac_jan02.ppt) (presentation).

### 1.6.1.2 SciDAC DOE Science Grid

This collaboration is detailed in the Certificate/Registration Authority section below. Here it will suffice to say that PPDG acknowledges the DOE Science Grid project for their energy and commitment to the development and support of these services, and for working with the European Data Grid to allow the PPDG experiment end-to-end applications security infrastructure to interoperate with their European colleagues.

### 1.6.1.3 Internet End-to-end Performance Monitoring (IEPM)

PPDG and IEPM (<http://www-iepm.slac.stanford.edu>) have established a collaboration to provide an ongoing forum for working together. IEPM will be invited to participate in PPDG collaboration meetings and report on their efforts which will benefit PPDG in the quarterly reports. IEPM monitoring and development work will take account of the needs of PPDG end-to-end application and monitoring requirements.

### 1.6.1.4 Logistical Networking Project

Initial contacts were made with Micah Beck to understand whether there are mutually beneficial tasks that could be initiated.

### 1.6.1.5 High Energy and Nuclear Physics Intergrid Management and Joint Technical Boards (HICB and HICB-JTB)

The HICB met in association with Global Grid Forum 3 in Rome in October. The International HENP Grid Coordination and Joint Development Framework plan (See Appendix 5.3) was accepted. Experiments presented their initial plans for international testbeds. Peter Couvares, Doug Olson, and Ruth Pordes were asked to represent PPDG on the Joint Technical Board. Other members of the board represent the European Data Grid, PPARC, DataTag, CrossGrid, the French and Italian Grid projects, GriPhyN, iVDGL, and Asian HENP grid projects. The JTB plans monthly phone conferences and is focussing initially on understanding the scope and efforts required for experiment and infrastructure joint testbed activities between the three continents. Larry Price is the chair of the HICB, Ruth and Peter Clarke - PPARC and DataTag – were asked to serve as joint co-chairs of the JTB.

### 1.6.1.6 GriPhyN

ATLAS and CMS are collaborators on the GriPhyN project. Their experiment GriPhyN application plans include the use and extension of PPDG deliverables. The personnel and plans of the two projects work together to provide worth for the experiment data handling systems. PPDG collaborated with GriPhyN on the grid project sponsored SC2001 CMS simulation data production demonstration.

### 1.6.1.7 iVDGL

PPDG is collaborating with the iVDGL project whose mission is "to be an international laboratory for the development and testing of virtual data grid middleware and the data grid applications that uses this middleware, in association with a number of experimental projects in physics and other disciplines. The U.S. part of iVDGL (US-iVDGL) is funded by an award from the 2001 NSF ITR program. The full international laboratory will be built on resources from US-iVDGL and a number of international partners, as outlined in the iVDGL proposal to NSF." The PPDG Steering Committee will hold some meetings with the directors of iVDGL to define more details of how we will proceed with the joint GriPhyN/PPDG collaboration on iVDGL. ATLAS and CMS are currently members of US-iVDGL and BaBar and D0 are exploring ways of working with the collaboration through Memoranda of Understanding. In particular the iVDGL foci on testing and support, and on Tier-2 data analysis centers, complements well the PPDG focus on end-to-end application development and Tier-A/Tier-1/Regional Center services.

## 2 Project Activities

### 2.1 GDMP (CMS-DataGrid-Globus)

The main aim of this quarter was to port GDMP to Globus 2.0 Alpha and GDMP 2.0 was released in October 2001. The software release has then been tested and used in the European DataGrid (EDG) testbed. Work has been done to help the integration work package of EDG to integrate the software with other EDG components.

A few new features have been and bug fixes have been added as regards GDMP version 2.0alpha.

Another main achievement was the usage of GDMP within MOP at Supercomputing 2001 in Denver.

Change the message passing system b/w GDMP client and server and force the implementation to send a reply for each request. After this new message system we authenticate the client once per connection only, not once per message.

Add the functionality so that GDMP server can use any certificate/key (any key which was created with nopass option). Now we can also use the host certificate/key to use as GDMP server certificate.

Now no need to publish the GDMP install path and files root directory path. Remote sites are free to change the installation paths of GDMP and files root directory paths. Sites can switch to new root directory as long as the file relative path under new root directory remains same.

Also now we maintain a list of host to which we had subscribed.

Still working on the way that one installation of GDMP can be shared by multiple users. For that I changing GDMP client tools ( gtmp\_publish\_catalogue, gtmp\_get\_catalogue, gtmp\_replicate\_gte, gtmp\_ping, gtmp\_register\_local\_file etc) in a way that these tools connect to the local GDMP server and then GDMP server connect to remote GDMP servers on behalf of them. By doing so the remote GDMP servers has to just authorize our GDMP server and not each local client. Local GDMP server will first authenticate local clients ( who has valid grid proxy and allowed to use gtmp ) and then talk to remote gtmp servers on behalf of them. Another advantage of this is that the authorized user can also use the GDMP remotely e.g. if a "usera" is allowed to use GDMP server "GDMPA" and if he/she has a valid grid-proxy then he/she can use this GDMPA from remote machines too.

All these functionality's will be available in GDMP 3.0, which is going to be released in start of Feb.

## 2.2 D0 Job Management (D0-Condor)

The D0 ppdg effort continues to define the Job Management problem within the context of SAM. Use cases have been developed to better understand the issues of remote job submission including resource brokering and remote environment setup requirements. We are working to coordinate with other D0 grid interests, including NIKHEF, Lancaster, Imperial College and U Texas Arlington, to define the requirements and understand existing Grid technologies that can be utilized. A system for defining and managing Monte Carlo processing requests has been completed and is being rolled out for use by the Dzero MC processing sites. This addition provides a flexible scheme to record parameters which define simulation datasets before the data is generated. As the data is produced and inserted into the system its metadata is verified against the initial parameters. Users can subsequently search for datasets based on these parameters. Experience with this may help eventually to better define the job control language needed to characterize jobs submitted to the system.

Work continues with the existing SAM deployment. Local job submission is used and great progress has been made improving the local fabric function on a distributed Linux cluster, including a reconstruction farm with 90 nodes, and a distributed analysis cluster that we expect will grow to over a hundred nodes in the next several months.

We anticipate a new hire to start in the next four to six weeks. This person will be dedicated to core Dzero SAM work and this will greatly increase availability of manpower for additional work on ppdg activities.

## 2.3 CMS-MOP (CMS-Condor)

The MOP project started the quarter with a successful demonstration at SC2001. Since then, the project has focused on integrating MOP into the mainstream production efforts of CMS. Integration of MOP into the long-term Grid strategy for CMS is ongoing.

The SC2001 demonstration of MOP included CMS Monte Carlo production jobs running at Fermilab, the University of Wisconsin, the University of California at San Diego, Caltech, and the SC2001 show floor itself. The machines on the show floor acted as a master site to submit jobs and collect output, in addition to running some of the production jobs. Many people contributed to the SC2001 demonstration. At Fermilab, James Amundson managed the MOP software itself in addition to doing the installation on SC2001 machines. Liz Quigg and John Weigand contributed to the graphical presentation, examples of which are available at <[http://www-ed.fnal.gov/work/sc2001/mop-animate-90\\_bothswf.html](http://www-ed.fnal.gov/work/sc2001/mop-animate-90_bothswf.html)> and

<[http://home.fnal.gov/~amundson/mop\\_pictures/fnal.png](http://home.fnal.gov/~amundson/mop_pictures/fnal.png)>. Greg Graham worked on the production software on the SC2001 machines. Conrad Steenberg demonstrated visualization of the production output as seen in <[http://home.fnal.gov/~amundson/mop\\_pictures/caltech.png](http://home.fnal.gov/~amundson/mop_pictures/caltech.png)>. GDMP support was provided by Shahzad Muzaffar. Support for the remote sites was provided by Peter Couvares, Rajesh Rajamani, Ian Fisk, Koen Holtman, Suresh Singh and Takako Hickey. The demonstration itself was very successful. Experiences from running the MOP demonstration were fed back to the Condor team.

Since SC2001, the focus for the project has been to integrate MOP with the mainstream efforts of CMS. Greg Graham and James Amundson worked to plan the re-integration of the MOP code into the current version of IMPALA. Greg Graham and Peter Couvares are currently working on the implementation. James Amundson presented a MOP overview and status report, <<http://home.fnal.gov/~amundson/mop12042001/>> at a CMS Grid integration meeting. Members of the Fermilab team have had meetings with both the Condor team and the CMS GriPhyn team to plan out future cooperation between MOP and other CMS Grid projects

Finally, MOP is a fundamental component in the upcoming US CMS Grid Testbed, a joint project between PPDG and GriPhyn. The baseline requirements for the testbed are to run MOP and the GriPhyn SC2001 demonstrations. Since the testbed will use Globus 2.0, MOP is being ported to Globus 2.0.

## 2.4 STAR-DDM (STAR-LBNL/SDM)

The HRM version 3.0 that was completed last quarter was installed on a PDSF NERSC machine and a machine at BNL. Both HRM have been configured to communicate with the local HPSS systems at BNL and LBNL. They were configured to work with Globus 1.1.3 alpha, since we did not have experience working with the Globus GT2 beta version yet. We had to overcome firewall limitation at BNL. This is basically accomplished by requesting to "pull" the files out of BNL using a designated port.

A successful test for File replication between BNL and NERSC was performed. It works as follows. A client program which can (reside anywhere) makes a request to the HRM at NERSC (on PDSF) to move a set of files. HRM-NERSC allocates space for each file, and makes a request to HRM-BNL for each file. HRM-BNL connects to HPSS-BNL to stage each file (using PFTP get). When the file is staged it notifies HRM-NERSC. HRM-LBNL then issues a globus-url-copy to move the file over the net. When this is done it notifies the client program that the file was transferred and it is now in the HRM cache. It then schedules the file to be archived (using PFTP put). When this is done the client notified that the file is now on HPSS.

This work will be reported in the January 10th meeting at Jlab. The test included monitoring the file transfer rate using large window with globus-url-copy. This information is available from the HRMs logs.

## 2.5 JLAB-Replication (JLAB-SRB)

Jefferson Lab and SRB continue to make progress in defining a web services interface to a file replication system as part of a strategy to provide a common layer to multiple storage systems. The goal in the first year is to define and implement a common interface to replication services provided by SDSC's SRB software and Jefferson Lab's JASMine software. An important step in this task is to carefully identify the capabilities to be provided by each web service, using as input the capabilities of existing data management systems as well as planned enhancements.

An initial comparison of the capabilities of five data grids that are used to support high energy physics has been completed (Storage Resource Broker (SRB) data grid from the San Diego Supercomputer Center, the GDMP data replication tool (a project in common between the European DataGrid and the Particle Physics Data Grid, augmented with an additional product of the European DataGrid for storing and retrieving meta-data in relational databases called Spitfire), the Globus data grid, the Sequential Access using Metadata (SAM) data grid from Fermi National Accelerator Laboratory, and the JASMine data grid from Jefferson Lab; document PPDG-8). Over 120 different features organized into 11 different categories are being supported by at least one of the data grids. Over 75% of the features are present in at least two of the data grids, with 50% of the features present in the majority of the data grids.

All of the data grids are implemented through a logical name space that is independent of the local storage system name space, with extensions to Unix commands based upon additional attributes managed in the logical name space. Extensions include latency management functions, discipline specific attributes, and attribute based discovery.

An important step in this project is to define which capabilities will be provided by each web service (high level decomposition). Two documents relevant to this step have been produced during this reporting period: (1) PPDG-9: Common Storage Resource Manager Operations, (<http://sdm.lbl.gov/srm/documents/joint.docs/srm.v1.0.doc>) and (2) Web Services Data Grid Architecture (draft, <http://www.jlab.org/hpc/datagrid/WebServicesDataGridArch.pdf>). The first of these defines the capabilities of the storage resource itself, and hence of the web service interface to a single site storage resource. The second provides a list of capabilities for additional web services including replica catalog service, replication (copy) service, and smaller services related to keeping the catalog up to date. This was further refined in discussions at the January replication workshop.

An important milestone will be the specification of the web services using Web Services Definition Language (WSDL), which is analogous to Corba's IDL (Interface Definition Language). Specifying this WSDL document requires agreeing on a large number of tag names and function names. Towards this end, the SRB group provided Jefferson Lab with a list of the attribute names and definitions used within SRB. Jefferson Lab has similarly provided SRB with the XML tag names used in its initial prototypes. In the coming quarter an initial subset of functionality (and names) will be selected as a first version of the WSDL for data grids.

## **2.6 ATLAS distributed data manager, MAGDA (ATLAS-Globus)**

The principal goal of the Magda project for this period was the completion and deployment of a version capable of production deployment in the ATLAS Data Challenge 0 commencing in December. This was achieved, with a DC0-ready version deployed and announced on December 7. Magda was adopted by international ATLAS as the file cataloging and replication tool for DC0 and by the end of the period was in use cataloging the DC0 data generated to date.

The most important new functionality implemented during the period was the completion and deployment of command-line tools providing a file access interface to production jobs. The `magda_findfile` command searches the catalog on the basis of LFN, LFN substring, location, etc. and reports results in a parsable format. The `magda_getfile` command retrieves files from any accessible location, making them available locally either as a local copy or a soft link to a replica in a managed location. Usage counts of files in managed locations and caches are maintained, with usage decremented when `magda_releasefile` is used, such that files can be pinned while they are in use. The `magda_putfile` command archives files in managed store locations and registers them in the catalog. These command line tools provide all the capability currently needed by ATLAS jobs to exploit Magda, so the direct integration of Magda into the Athena framework continues to be deferred until manpower for this more exploratory work is identified.

Integration of GDMP with Magda was identified as the highest priority in further integrating Grid toolkit components with Magda. An integrated deployment of Magda and GDMP is foreseen in the ATLAS Data Challenge 1 commencing in Spring 2002, permitting ATLAS to draw on both PPDG and EDG WP2 data management efforts in a coordinated way. Towards this end, the GDMP design and feature set was reviewed with a view to Magda integration, and an integration plan begun. Problematic issues in the integration were identified and gathered for discussion at a PPDG data management meeting in early Jan.

ATLAS/PPDG has been instrumental within international ATLAS in planning and coordinating a coherent approach to replica and metadata management for the ATLAS Data Challenges, integrating the plans and deliverables of PPDG and EDG.

Magda deployment was completed or initiated at several new sites during the period, including Indiana University (completed), IN2P3 and UT Arlington (underway). Magda-based replication of ATLAS data between CERN and BNL continued, with ~300GB of data now replicated. Magda now catalogs files representing more than 6TB of data.

Near term plans include exercising Magda in a production setting in DC0 and feeding experiences back into the development cycle; integration with hybrid (RDBMS+object streaming) event stores; integration with application metadata catalogs; integrating GDMP in preparation for DC1; and further integration of Globus tools, particularly remote command execution for more flexible Magda usage at testbed sites.

During the period we developed (primarily off-project) a design and description of a 'hybrid' persistent data model consisting of data files plus a data management and metadata layer, the latter to be implemented using a combination of grid toolkit components and higher level metadata services. The hybrid data model is a proposal for managing the event data in an HEP experiment. It explicitly recognizes that the data is stored in files and separates the largely grid-based management and tracking of those files from the management of event data objects within the files. It addresses the problem of maintaining persistent references between event objects. The management of physics data collections (called datasets) is also discussed. Most of the work thus far is in design work directed at file-level management of and access to distributed event data, directly applicable to our PPDG program in distributed data management development. For details see <http://www.usatlas.bnl.gov/~dladams/hybrid>.

## 2.7 BaBar Database Replication (BaBar-SRB)

Work on the first BaBar file replication prototype using SRB was completed in November 2001. The intention was to extend the MCAT schema to include BaBar-specific attributes. Two test federations were setup and information was stored in the MCAT. SRB was then used to query the MCAT and replicate databases from one test federation to the other using bbcp. The prototype was demonstrated during SC2001 and succeeded in replicating data from source to target federation. The first prototype, although successful, uncovered a number of potential problems that could affect the implementation in a production system at BaBar (the main concern was the strong coupling between the experiment specific schema and the SRB). This led us to start work on a second prototype. The intention is to store BaBar specific metadata information in relational database tables whilst keeping file-specific metadata information in the MCAT.

Arcot Rajasekar, Mike Wan, Bing Zhu had a collaboration meeting at SDSC with Adil Hasan from SLAC who works on the BaBar project.

The meeting has the agenda of designing and implementing one of the user scenarios for BaBar collection data movement. The scenario was to provide a simple method for 'bundling' collections in BaBar, based on user restrictions, and copy the data to users workspace. We were able to accomplish that goal within two days.

BaBar has a complex data model, based on Objectivity technology, and has a lot of legacy code. The aim was to use the existing 'methods' in BaBar but provide a means for it to provide data movement and data sharing capability based on the SRB.

As a result of the two day intensive hands-on meeting, we were able to provide:

- (a) a means to export required object/collection metadata from Objectivity to a relational database
- (b) expose this relational database and its querying capability through SRB
- (c) use BaBar's own pftp-based method to stage data from HPSS using the SRB (a new driver for doing this was written and tested within one day)
- (d) implement methods to use the Objectivity metadata to register HPSS files into SRB, thus exporting them to the SRB space,
- (e) "connect" the metadata in the relational DB with the files in the HPSS through the SRB,
- (f) Bring up a new SRB server to interface with the HPSS using the new driver and making this SRB server part of a SRBspace based on MCAT running at SLAC.
- (g) Define and implement a remote proxy operation to be run under SRB which performs access and bundling operations.
- (h) write a simple client to provide an interface to the user. (this was put together later by Adil).

Apart from this, during the meeting, we discussed the current data model within BaBar, SRB software, and other collaboration areas between BaBar and SRB.

We anticipate to have a fully functional second prototype system that's useable by BaBar collaborators before the next quarter.

## 3 Cross-cut Activities and Collaborations

### 3.1 SuperComputing 2001 Demos

SC2001 provided an excellent opportunity for demonstrations of both end-to-end GRID applications and selected components of the ongoing work of PPDG. A full list is given in Appendix A and is available at [http://www.ppdg.net/docs/presentations/sc2001\\_demos.htm](http://www.ppdg.net/docs/presentations/sc2001_demos.htm). The many visitors to the ANL, Caltech, Fermilab/SLAC and LBNL booths showed much interest in working software and future developments.

In conjunction with the CMS distributed production demonstration a joint handout with GriPhyN was developed which is available at <http://www.ppdg.net/docs/presentations/handout.pdf>.

### 3.2 Certificate/Registration Authority

At that end of September 2001, ESnet received approval and funding to run a certificate authority (CA) for the doesciencegrid.org domain. PPDG has had numerous discussions and teleconferences about setting up the policy (CP/CPS) and how to implement and operate a registration authority (RA) in conjunction with this CA. These discussions included ESnet personnel, members of the DOE Science Grid SciDAC project (DOESG), and people from the community of computing sites participating in the PPDG-related physics experiments. An email list, with web archive, was set up called [ppdg-cara@ppdg.net](mailto:ppdg-cara@ppdg.net). Links to list membership and the archive of messages can be found at [www.ppdg.net](http://www.ppdg.net). A primary motivating factor for this effort is to establish an operating CA/RA in the US which is able to issue X.509 certificates to users (and computer hosts) that have a sufficient level of identity checking so that computer administrators around the US and Europe (EU-DataGrid project) are willing to accept these certificates in the authentication step of allowing users access to the computing resources.

In December, Tony Genovese and Mike Helm of ESnet attended a meeting of the EU-DataGrid (EDG) testbed security working group at CERN to discuss that status of the DOESG certificate authority and draft policy document. A summary of this meeting as well as status of the CA implementation and plan is available at <http://envisage.es.net/>. A successful result of this meeting was that the EDG testbed sites will accept the user certificates from the DOESG CA. This was a critical step in the effort to achieve and demonstrate interoperable datagrid activity between the US and Europe.

The implementation plan has a certificate authority beginning operations in January 2002 and to have the registration authority functions delegated to people in each of the projects for which certificates will be issued. There will be one registration authority for PPDG with Doug Olson as primary contact, as well as two additional registration authorities (initially) for the DOE ScienceGrid and the Fusion Collaboratory projects with Mary Thompson as the primary contact.

A description of how the RA for PPDG will operate is included as part of the CPS (Certificate Practice Statement) available at <http://envisage.es.net/> but a brief summary is provided here. The RA will have a defined group of people within PPDG who can carry out the identity check step for people who request a certificate. This group is initially defined as the steering committee members. The identity check will either be face-to-face in the case that the individuals do not already know each other, or it can be handled by telephone or digitally signed email for the case that the individuals are already well know to each other. This check will be communicated to Doug Olson who will then use the RA interface on the secure web server operated by ESnet for approval or rejection of the certificate request. This RA operation should also get started in January 2002.

We expect that there will be an assortment of issues that arise when actually operating a CA/RA for the PPDG community and that details of this operation will evolve over the next several months. It is anticipated that some of the physics experiments participating in PPDG may decide it is most effective to

run their own RA function and both BaBar and D0 have indicated interest in this. In the long run we expect that experiments and/or significant computing sites (such as NERSC, SLAC, Fermilab, etc.) will run their own RA or CA functions.

As mentioned briefly above, the role of the CA and RA is only for issuing identity certificates that are then used in the authentication step of accessing computing resources. A larger issue with less well-developed technology and understanding at this point is the topic of “authorization” for using computing resources on the grid. Today, of course, each site has its own well-established and functional mechanisms for authorizing use of resources. The effort from the grid community is to establish mechanisms for defining membership and rights in virtual organizations (VO ref. here) and to be able to grant access to resources based upon this membership and rights defined by groups within one of these VOs (a physics experiment is one example of a VO). We expect that PPDG will become involved in this area of authorization in the coming months. There are three projects developing technology in this area today, the Globus CAS project, the Akenti project, and the EU-DataGrid WP6 project at INFN. PPDG will work on testing and deploying one or more of these technologies.

### 3.3 Monitoring

Initial steps in organizing a monitoring effort were taken during this period. A web site was created, available at <http://www-unix.mcs.anl.gov/~schopf/pg-monitoring/>, as were a mailing list and archives.

The first goal for this group was to define use cases for requirements gathering. We did this by first defining a template, and then by requesting use cases from the experimentalists involved in this effort. To date we have 19 of these covering a wide range of examples from testing a network for stability to evaluating the progress of an application. Jennifer Schopf will present this work at the Internet2 End-to-End Performance Initiative Measurement Workshop in January 2002.

We are in the process of defining a set of sensors to be deployed in the various application testbeds, and identifying extensions to them that are needed for a Grid environment, for example capturing summary data for farms of machines. Once an initial set of sensors are defined, we will need to define schema for related values so that they can be interfaced to a Grid Information System. Once schemas are defined, we will deploy the sensors and interface them to the Globus Toolkit MDS 2.1. The goal is to have an initial set of sensors selected, with well-defined schemas and interfaces to the MDS for the VDT 2.0 release in July 2002.

Individual groups have also been working on monitoring internally. For example, ATLAS has been evaluating and installing sensors to capture the needed data for their testbed facilities, and determining what information should be shared at the Grid level, and the best ways to do this. A farm monitoring system has been developed which can monitor up to 300 Linux nodes, although further scalability is also being investigated. This system consists of three modules: several data collection modules which collect system sensor data and push the data into local data server, a database module which stores system status data and provides data services, and an information provider module which pulls data from the database server, summarizes it and publishes the data into the Globus MDS. An initial deployment is in prototype on the BNL compute farm.

### 3.4 Collaboration with IEPM, Network Performance Monitoring

Contact: Les Cottrell, SLAC

We put together a project for the SC2001 Bandwidth Challenge. See <http://www-iepm.slac.stanford.edu/monitoring/bulk/sc2001/> for more details. It included over 25 collaborating sites (including all the PPDG sites) to which we sent large amounts of bulk throughput from the SLAC/FNAL booth at SC2001. We achieved over 1.6Gbps/sec throughput simultaneously to about 17 sites in 5 countries. We also demonstrated the effectiveness of QBSS for very high speed links (2Gbits/sec). Following SC2001 we extended and ruggedized the infrastructure put in place for the SC2001 bandwidth challenge and tentatively name the project IEPM-BW (Internet End-to-end Performance measurement - BandWidth). We now have about 30 sites in 8 countries, and are making regular measurements with ping, traceroute, bbcp (both memory to memory and disk to disk), bbftp and pipechar.

We are starting to analyze the data from these measurements. We made presentations on Achieving High throughput performance at the ESCC meeting at ANL, and the Inaugural Internet 2 HENP networking working group at Ann Arbor Michigan. We made presentations on High throughput network performance measurements to the ICFA / Standing Committee on Interregional Connectivity (SCIC) at CERN, and at the Babar collaboration meeting at SLAC. We have also made presentatins on QBSS, PingER futures, and Grid Monitoring at ESCC and the Virtual Internet 2 Members meeting.

Early results from the IEPM-BW project indicate:

- Reasonable estimates of throughput can be obtained with 10 second iperf or bbcp measurements. This is typically much shorter than it takes to make a pipechar measurement.
- In many cases it is not sufficient to simply increase the window size to achieve high throughput, multiple parallel streams are also critical.
- Careful attention to window sizes and parallel streams in necessary. Improvements of between 5 and 60 times have been observed for the optimum window and stream settings compared to using a single stream and the default maximum window size.
- It is also observed that there is an optimum window\*number parallel streams beyond which performance does not increase, or may decrease, while cpu, packet loss increases.
- Throughput can vary by an order of magnitude with time of day or day of week etc.
- Roughly speaking one needs 1 MHz to provide 1 Mbit/sec on today's cpus and OSs.
- The bbcp file copy rates from memory to memory are about 60+-20% of the iperf rates.
- File copy rates disk to disk are typically about 90% of the memory to memory rates, for rates below 60Mbits/s, but can vary dramatically depending on disk performance, caching etc. Uncached disk performance typically tops out at between 4 and 8MBytes/sec.
- In some cases (e.g. SLAC to CERN for BaBar Objectivity data) compression can improve throughput by over a factor of 2 on a reasonably high performance host (Sun 336MHz cpus).
- When running high throughput applications, the RTT for other users can be noticeably increased, e.g. for SLAC to CERN the average increases from about 160 msec. to about 260 msec.
- The impact of high throughput applications, on other applications requiring low latency, may be reduced by applying lower than best effort priority (Scavenger Service) to the high throughput applications' packets.

We are in the process of improving the analysis and reporting/graphing/table tools. We are also building tools to facilitate and automate the infrastructure management. This includes downloading of code, checking whether measurements are successful, gathering the remote configurations parameters (OS, cpu speed, code versions), understanding disk performance, verifying windows and streams are set correctly.

We plan create a web site organized to provide easy access to all aspects of this project. We will also measure the impacts of compression, add and understand gridFTP and other bandwidth measurement tools, and compare and contrast the various measurements. Following this we will select a representative minimum subset of tools to make measurements with, improve the reporting/graphing/table tools and make the data available via the web. We also hope to tie together the measurements being made in the UK with the SLAC measurements so they appear more integrated to the user.

## 4 Single Collaborator Efforts and End to End Applications

## 4.1 ATLAS

### 4.1.1 US ATLAS Grid Testbed

Testbed sites continued to deploy and test additional grid infrastructure components, including Magda, the pacman package manager, GDMP, and Globus 2.0 (GridFTP, replica catalog, etc.). Work is underway to support the automated distribution via pacman (developed within ATLAS GriPhyN) of all components needed for deployment of a Grid-integrated testbed site capable of running ATLAS software. Regression tests distributed by pacman to validate various grid services (Globus, GDMP, Magda, etc.) on the testbed are also under development. We are becoming involved (Jerry Gieraltowski at ANL) in the ATLAS grid validation activities taking place within EDG. Network performance monitoring and tuning activities between BNL and US ATLAS grid testbed sites continued during the period.

### 4.1.2 Monitoring

ATLAS continued its involvement in the PPDG Grid Monitoring Project, developing use cases, defining the scope of the project, designing and developing linux farm monitoring using MDS. A prototype Linux farm monitoring tool using MDS was developed, based on the existing BNL farm monitoring system. The system can monitor up to 300 nodes, and the scalability will improve in future versions. Currently the system can answer limited questions which Grid users might ask via MDS, eg. “give me 40 least loaded nodes with Linux Kernel 2.4”.

### 4.1.3 Distributed job management

Distributed job management activity focused on ATLAS DC0-directed development and deployment of a job management infrastructure integrating the use of distributed data management (Magda) and application metadata management (under development by Grenoble ATLAS) tools. Development of distributed job management proper will begin post-DC0, since DC0 does not involve distributed production. ATLAS DC1, commencing spring 2002, will involve distributed production, and we anticipate developing, deploying, testing, and iterating the development of distributed job management tools before and during the ~6 month duration of DC1. We will look closely at existing tools for adoption, in particular the MOP package of CMS/PPDG.

### 4.1.4 Data signature

During the quarter a preliminary design was developed for the ‘event data history’ classes that will constitute part of the data signature required for data equivalency tests or on-demand data regeneration. The goal of event data history is to record the history of data at the level of individual event data objects (EDO's), i.e. the collections of physics objects (clusters, tracks, electrons, ...) that comprise the event data of high energy physics. We require that the history be sufficient to reproduce the data at that level. We identify three levels of history objects:

1. Algorithm history
2. Job history
3. EDO history (includes pointers to its job and algorithm histories)

Classes describing each of the above were developed and can be found in the ATLAS repository under Control/AthenaHistory. For details see [http://www.ustalas.bnl.gov/~dladams/data\\_history](http://www.ustalas.bnl.gov/~dladams/data_history).

## 4.2 BaBar

A significant grid event for BaBar this quarter was the “Distributed Analysis at Tier A Centres Workshop”, <http://www.slac.stanford.edu/BFROOT/www/Computing/Offline/BaBarGrid/meetings/011215/agenda.html>.

Below is a summary of this meeting.

Present: Tim Adye, Roger Barlow, Dominique Boutigny, Fredric Brouchu, Cristina Bulfon, Bob Cowles, Sridhara Dasu, Serge Du, Pete Elmer, Guiseppa Della Rica, Alessandra Forti, Gerald Adil Hasan, Mark Kelly, Miron Livny, Doug Olson, Steve Playfer, Douglas Smith, Roberto Stroili, Artem Trunov.

Miron Livny: Experience from Condor-G

Miron outlined the architecture of Condor-g. Miron's interesting talk provokes the following questions: How should we handle errors, log files/output, how should this information be communicated to the user? What happens if the connection between the Tier-C and Tier-A disappears, should batch log files be kept locally at the Tier-A and shipped to the Tier-C at the end?

Condor maintains a central queue - this queue is the one that users know about and query jobs through. This allows users easier control, permitting them query, cancellation rights. Should we be thinking about a similar approach? What are the problems we face if we don't do this?

The resources that a batch job needs are handled by the DAG-man, this process has the ability to prevent the job from running until the resources needed by the job are ready (eg files are staged in, etc). How could we deal with such problems?

Ralph-Muller Pfefferkorn: BaBar Grid tests at Karlsruhe

Ralph report that Karlsruhe is now a Tier A centre. They have started to do some grid tests at Dresden using 1 node and are awaiting a German CA. BaBar will be the first user of the Karlsruhe center, initially MC production will be done there.

Bob Cowles: Security Issues

Bob's interesting talk on security issues provoked some rather worrying concerns such as:

The grid needs a trusted time source in order to be able to revoke/expire certificates after a certain time. How should we handle the problem of certificate expiration?

Should we be thinking about restricting resources at Tier-A centers such that BaBar users only use specified resources. How could we handle that?

The problem of issuing certificates came up. Should we have group certificates? How would we find the users responsible for a set of jobs?

Roger Barlow: Intergrid and liason with grid projects

Roger indicated that there was interest in trying to involve BaBar grid efforts within the wider grid efforts going on. Roger has communicated BaBar's intentions to InterGrid. I believe that we want to keep intergrid informed on what we're doing and what we want/require (from these grid efforts).

Roger Barlow: Requirements for becoming a BaBar Grid

Roger presented a list of requirements that were thought important in order to achieve a BaBar mini-grid. Miron encouraged us to look at Condor for information and examples. At the moment people in the US (and elsewhere) can get certificates from CNRS in France.

One of the major concerns was the requirement for afs: this is mainly because the BaBar software doesn't currently lend itself to a distributed computing model, requiring ascii files to be read at runtime from afs. To get around this Roger offered 2 possibilities, shipping the necessary files with the job or run afs at the Tier-C center. This needs further investigation as it was unclear whether coupling the job running at the Tier-A to the files at the Tier-C center was a good idea. It is clear that re-thinking how the BaBar s/w loads the information necessary for analysis jobs would be a good thing.

Serge Du: Globus submission and analysis job tests

Serge reported on tests performed between in2p3 and RAL. The tools built to carry out these tests are available in CVS (GanaTools). There's a clear need for coordination to make sure that each site can do what's advertised. Serge and Stephane's tests brought up valuable questions that need to be folded back into the design of the distributed batch system.

Fergus Wilson: MC production

Fergus reported that MC production is already a distributed system. However, the system doesn't appear able to benefit from the distributed batch system due to the need to prepare a considerable amount of infrastructure at each production site (eg a run generated at site A cannot be reconstruct at site B - unless all the necessary dbs etc are copied from A to B). There is interest in trying to see how grid-tools could be applied to MC production to get around some of the problems encountered. Should we be thinking about a pool of BaBar resources that changes over time and that can be used instantly when the need presents itself.

Adil Hasan: BdbServer++, Distributed Objy catalog

Adil reported that Dominique and people at in2p3 are currently manually extracting Stream17 for Caltech. The steps that are involved in this process will help to shape the design of BdbServer++. BdbServer++ is a crucial component of the new data model allowing users to extract events of interest and ship them to their site. The extraction would read from one set of dbs and write to another set allowing users to the ability to ship different levels of data.

One of the key ingredients for distributed batch and data distribution is a catalog containing the list of collections and databases associated to those collections. Work is currently going on to produce a prototype prod setup in which background trigger databases and collections would be registered in the Storage Resource Broker allowing users to extract collections through the SRB and ship them to their external site.

**4.3 CMS**

CMS related effort in PPDG concentrated at Fermilab, Wisconsin and Caltech.

Work at Fermilab ws dominated by the MOP demo at SC2001, and by software development for GDMP. The very successful demo of distributed and remote job submission of CMS Monte Carlo production using MOP, and recent developments on MOP are described in section 2.3 in the report. This work is being done together with the CS team at the University of Wisconsin.

The other CMS-PPDG activity at Fermilab is software development on GDMP, in collaboration with WP2 of the EDG project. The newly implemented functionality will be available in GDMP 3.0, which is going to be released beginning of February.

This work includes a change in the message passing system between the GDMP client and server. In the new implementation the server is sending a reply for each request. In this new message system GDMP authenticates the client just once per connection, instead of once per message in the previous implementation.

Functionality was added so that the GDMP server can use any certificate/key (that is any key which was created with a nopass option). The host certificate/key can now also be used as the GDMP server certificate.

To ease installation at remote sites, the previous requirement to publish the GDMP install path and root directory path for files was removed. Remote sites are free to change the installation paths and the files root directory paths. Sites can switch to a new root directory as long as the relative path of files under new root directory remains same.

GDMP now maintains a list of host to which was being subscribed. Also, work is progressing on enabling one installation of GDMP to be shared by multiple users.

The GDMP client tools (gdmp\_publish\_catalogue, gdmp\_get\_catalogue, gdmp\_replicate\_gte, gdmp\_ping, gdmp\_register\_local\_file etc) are being modified in a way that these tools connect to the local GDMP server and that server then connects to remote GDMP servers on behalf of the clients. By doing so the remote GDMP server has to just authorize the local GDMP server and not each local client. The local GDMP server will start with first authenticating local clients (who have a valid grid proxy and are allowed to use GDMP ) and then will communicate with the remote GDMP servers on behalf of the local

clients. One advantage of this new scheme is that an authorized user can also use a GDMP service from remote.

At Caltech, Iosif Legrand continued to work on the prototype for distributed services. Most of the effort was put in the developing a flexible monitoring service. The design of this monitoring framework is described in [http://clegrand.home.cern.ch/clegrand/CMS\\_Monitor/MonitorTool.doc](http://clegrand.home.cern.ch/clegrand/CMS_Monitor/MonitorTool.doc).

He developed a set of APIs to allow JINI services to very easily establish a peer to peer connection. This is done using the JNI lease mechanism so that the network of services using these APIs are automatically updated and each unit is a listener for events generated by the peers and the lookup discovery services in case one unit fails.

The basic JINI services (discovery mechanism, Javaspace and the transaction manager services) were successfully tested over the wide area network between CERN and Caltech. This has been done by creating a network of peer to peer services providing simple monitoring tasks and continuously updating the content, see [http://clegrand.home.cern.ch/clegrand/CHEP01/chep01\\_10-010.pdf](http://clegrand.home.cern.ch/clegrand/CHEP01/chep01_10-010.pdf)

A GUI was developed to control and configure large scale monitoring systems. This GUI was done to allow to be dynamically exported as a complex proxy (marshaled and serialized) and registered as a service "attribute" with the lookup discovery services. A component factory is used to provide dynamically the classes the client needs to run it. These scheme allows to provide a GUI which is well adapted for client configuration.

Modules using SNMP (get, walk and trap) can now be dynamically added into the monitoring system. The task scheduling is done in a multithreaded framework which controls the execution and acts in case of network failure to recover the unfinished tasks. This system was successfully tested at CERN using ~ 400 linux nodes for several days. It also allows to monitor routers and switching units and this part was also tested at CERN.

Other PPDG activities at Caltech included work on the Robust Execution Services (RES) by Takako Hickey. Access to the Caltech prototype Tier-2 facility has been enabled from GRAM. An interface for submitting jobs from the MOP scheduler via GRAM to RES is now working, and the implementation of the interfaces for killing jobs and getting job status will follow. For the MOP demo at SC2001 the GRAM/PBS software was installed at Caltech, for running jobs via PBS. The use of RES will be tested in the coming quarter. The system will be used from MOP, where RES replaces the previous (non-robust) job execution system PBS.

Conrad Steenberg continued his work at Caltech on the prototype for the remote analysis environment in CMS (Clarens). The server side of Clarens was rewritten to improve scalability and robustness. This change allows in-server execution of plug-in modules to give clients transparent access to different services, installing new services without server down-time. Also, clients are now authenticated using Globus certificates.

The Clarens data server is installed and is being tested on the Tier2 prototype at Caltech. Clarens was part of a remote data access demo at SC 2001, as part of the Fermilab/SLAC booth.

Koen Holtman at Caltech was involved in the planning, coding, installation, and performing of the CMS/Caltech demo 'Bandwidth Greedy Grid-enabled Object Collection Analysis for Particle Physics' at the Caltech CACR booth at SC2001.

He has also worked on architecture and requirements, and published a note "Views of CMS Event Data: Objects, Files, Collections, Virtual Data Products" (CMS NOTE-2001/047). He contributed to the CMS database selection milestone, investigating the impact on the requirements for grid middleware of CMS investigation and possible move from using Objectivity to a hybrid persistency model. He also worked on requirements for the replica catalog and on fault tolerance requirements for related grid-wide services.

#### 4.4 D0

The people working on this at FNAL include Gabriele Garzoglio, Igor Terekhov, Sinisa Veseli and Lee Lueking. Work has begun to organize the Dzero Grid effort. Regular bi-weekly meetings are held, and a task list with plans and deliverables has been put together and discussed. Igor is leading the group to

develop use cases for the job submission and system monitoring efforts. From this we will establish requirements and compare to the specifications for other planned, and existing, Grid software being built by other groups. Sinisa has been putting together the use cases for monitoring and is designing the needed architecture. He has installed Globus and plans to evaluate this, along with other potential Grid technologies for use in this application. He has helped develop the work plan and tasklist for the monitoring portion of our effort. Also, we are joining relevant phone cons including the bi-weekly ppdg monitoring meetings.

We worked on the demonstration for SC2001 until the conference in November. This included extracting file transfer information from the SAM log files and entering it into a relational database. This information was used as input to a display that showed the source, destination, and rate of data transfers throughout the currently deployed SAM system with facilities at FNAL, NIKHEF, Lancaster, Imperial College, and U. Texas Arlington. This demo represents a working example of a SAM/Grid monitoring tool useful not only for the demo, but also for monitoring and understanding the operational system. Gabrieli worked with two students on this project to get information from the database and produce the graphical presentation and Sinisa created the tools that populated the database. Other work has been accomplished by Gabriele to plot the access patterns of the SAM file access and these are available at <http://d0db.fnal.gov/sam/plots-and-stats.html> under the “production” link in the category “data access statistics”. Also, a SAM adapter was built for PBS (Portable Batch System) and it has now been debugged and is being put into use on one of the D0 distributed analysis Linux clusters.

In the upcoming months we plan to have a working meeting while many of the Dzero Grid developers are at FNAL for the Dzero collaboration meeting in early February. We will begin evaluation of GridFTP in the coming period, and also integration of CA/PKI authentication with kerberos authentication has been done and we will begin testing this.

#### 4.5 Jlab

In addition to the joint project with SDSC/SRB on defining a common web services interface to storage systems, Jefferson Lab worked with Arie Shoshani's group at LBNL and others to define the common capabilities of a storage resource manager (SRM). Some progress was made this quarter in planning for changes in the lab's back-end storage system JASMine to conform to this document, with additional effort going into web services at the next layer up.

The lab continued to develop a data grid web service (version 0.2) to the JASMine storage system, with a general interface corresponding to the SRM document. During this quarter, the initial XML prototype was converted to a standard SOAP interface. An initial Web Service Definition Language (WSDL) document for this service was produced ([http://lqcd.jlab.org/grid/gridService\\_wsdl.xml](http://lqcd.jlab.org/grid/gridService_wsdl.xml)). Additional discussion and collaborative improvement of the interface definition is expected in the next several months.

Jefferson Lab also started work on a Reliable File Transfer Web Service, defining the initial web service interface, which includes reliable 3rd party transfers. The initial prototype work is expected to be finished by the middle of February.

#### 4.6 STAR

The STAR activity at BNL focused mainly on updating the STAR grid node, [stargrid01.rcf.bnl.gov](http://stargrid01.rcf.bnl.gov), with the most recent versions of Globus and HRM. An important milestone was passed in performing a real-world test of HRM, transferring STAR data from HPSS at BNL to disk (and HPSS) at PDSF. Dantong Yu worked on installing the necessary software on the local node. Jérôme Lauret acted as coordinator of this effort between BNL and PDSF, in particular coordinating the efforts of Dantong Yu and Alex Sim (PDSF) and the interaction with the local HPSS group at BNL. He also provided the necessary environment with BNL cyber-security in applying for the required conduits in the perimeter firewall and following up on their implementation with the local cyber-security group. This turned out to be not as trivial a task as one might have wished and one concludes that some work has to be done in formalizing this kind of interaction to make a larger scale implementation of GRID software for STAR data-transfer feasible.

The current setup of the stargrid node is thus:

- Redhat Linux 6.2, NFS-mounted STAR data disks and user home-directories, openAFS 1.0.4, making this node so far *identical* to the main STAR analysis nodes rcasnn.rcf.bnl.gov. This is an important pre-requisite within the view of making GRID enabled file-transfer a standard tool of STAR data analysis in the future.
- Globus 1.1.3 alpha
- Globus 2.0 beta, including now
  - Replica catalog
  - Grid ftp

Progress has been made in hiring a professional software developer using the STAR-BNL share of our annual PPDG budget. We have identified a candidate and are in the procedure of opening a job requisition at BNL. It is expected that this person will start contributing 70% of his time to STAR-PPDG within the next month.

## 4.7 ANL – Globus

### 4.7.1 DGRA document

A new version of the Data Grid Reference Architecture Document was released by Ian Foster and Carl Kesselman for comment. This is now a joint framework for PPDG and GriPhyN, and the document details the current understanding of the requirements that motivate the development of Data Grids and addresses the PPDG/GriPhyN approach to addressing these requirements. This document is available at the GriPhyN web site.

### 4.7.2 Globus Toolkit 2.0 work

A beta release of the Globus Toolkit 2.0 was made in the last weeks of November. This release includes new features from four previous alpha releases, including Data Grid tools, MDS-2.1, a significant repackaging of the entire software for easier installation, as well as numerous additional patches and features including GRAM 1.5 with increased job submission reliability.

More information about the GT 2.0 release can be found at <http://www.globus.org/gt2/>

### 4.7.3 Community Authorization Service

The Globus Project efforts towards the Community Authorization Service (CAS) in the last quarter have been directed preparing CAS for release to close collaborators. We integrated CAS with a actual scientific application, the Earth Systems Grid Visual Climate Data Analysis Toolkit (VCDAT), for a demonstration at SC'01. Work has continued since then improving and documenting CAS for a planned release to close collaborators at the end of January 2002.

More information on CAS can be found at <http://www.globus.org/security/CAS/>

### 4.7.4 Reliable File Transfer Service

Work for the Globus Reliable File Transfer (RFT) Service continued. An RFT Service is a service that allows byte streams to be transferred in a reliable manner. Reliability, in this context, means that problems of less than a certain, user defined magnitude are dealt with automatically. i.e. problems like dropped connections, machine reboots, temporary network outages, etc are dealt with automatically (usually via retry) until they either resume or meet some "ultimate failure" condition.

The RFT consists of following pieces

1. 1.The Transfer Service, which accepts the transfer requests

2. 2.Transfer Request Client GUI, to submit the transfer requests to service and to receive the status updates of the same.
3. 3.Transfer Client, is a C binary that actually performs transfers using GridFTP.
4. 4.Netlogger,to monitor and archive the performance of transfers.
5. 5.Database, to store the state of all the transfers.

A prototype built over current Globus tools was demonstrated at SuperComputing '01 (details available at <http://www-unix.mcs.anl.gov/~madduri/SC2001.html>)

Currently we are working on developing a prototype that exposes the Reliable File Transfer Service as a Web Service using SOAP. We are using Axis as our SOAP engine and Apache Tomcat as our Webserver. We are also working on delegating the user's proxy along with the Transfer Request so that the transfers are done basing on user's credentials.

Detailed information is available at <http://www-unix.mcs.anl.gov/~madduri/RFT.html>

#### **4.7.5 Replica Location Service**

In terms of easier deployment for PPDG users of replica catalogs, we are working on setting up a Globus Replica Catalog test scenario for users, and then writing a straightforward guide or HOW-TO that shows users how to easily setup and administer a replica catalog.

In terms of research into replica selection, Sudharshan Vazkudai, Jennifer Schopf and Ian Foster had a paper accepted at IPDPS detailing a predictive technique for GridFTP for replica selection decisions. A copy of the paper is available at <http://www.globus.org/research/papers/Prediction-Paper-249.pdf>

#### **4.7.6 Monitoring**

Jennifer Schopf continued her role as co-lead of the joint PPDG/Griphyn.iVDGL monitoring group. This work is detailed in the section on Monitoring.

### **4.8 NERSC – SDM**

People involved: Junmin Gu, Alex Sim, Arie Shoshani

There were 3 major activities during the last quarter that we can report progress on:

- 1) The HRM version 3.0 that was completed last quarter was installed on a PDSF NERSC machine and a machine at BNL. This is reported above in the section on the STAR-DDM project activity.
- 2) We initiated a joint functional design effort between PPDG and EU Data Grid. First, Arie Shoshani visited CERN on October 11-12, 2001, meeting with people from WP2 (Wolfgang Hoschek, Peter Kunszt, Heinz Stockinger, Kurt Stockinger, Brian Tierney) and WP5 (Jean-Philippe Baud). This was followed by a second meeting at LBNL on December 2-3, 2001, with participants from JLAB (Bryan Hess, Andy Kowalski), Fermi (Don Petravick, Rich Wellner), and LBNL (Junmin Gu, Ekow Otoo, Alex Romosan, Alex Sim, Arie Shoshani).

We wrote a document summarizes the conclusions reached for the functional specification of Storage Resource Managers (SRMs) by the participants of two recent meetings. The participants are people involved in the PPDG and EDG projects who are interested in SRM technology and who either developed or are in the process of developing SRMs. This document reflects the common wisdom and experience of people from both PPDG and EDG. It is intended as a guide for a joint PPDG-EDG document on the SRM functional specification. The document was written by Arie Shoshani, and will be submitted and presented at the GGF4 meeting in February.

In our discussions, we had the benefit from people's knowledge of four different archival systems: Fermi has experience with their own Enstore system, JLAB has its own home grown mass storage system, JASMine, LBNL has developed SRMs for HPSS, and CERN has developed their own system, CASTOR.

Our goal is to achieve the generality of providing the same SRM interfaces to all these system, any disk cache systems, or any future storage systems.

3) We have deployed an HRM command-line client that was developed for another grid project (ESG) to the needs of the PPDG project. This was used to invoke the replication test mentioned. This command-line client will be used initially by the STAR team, and later may be used for streamlining the replication process. The HRM command-line module is a client-side module that can be used directly by clients to get/put files into an HPSS archive using the HRM as an intermediary grid component. In addition, the command line client module facilitates a third party "copy" from any HRM-HPSS site to another HRM-HPSS site.

The advantage of using HRM-HPSS as an intermediary is that HRM utilizes its own disk cache for fast transfer of file over the grid (as fast as the network connection will permit), and queues the staging and archiving from/into HPSS at the rate it permits. It monitors that transfer of files into HPSS, and provides the status on the progress of the file transfer and archiving. It uses the latest GridFTP software available to take advantage of parallel FTP streams and large window sizes. The client does not have to be aware of these details. In addition, HRM insulates the client from transient HPSS failures - If HPSS temporarily fails to stage or archive a file, HRM keeps trying till HPSS recovered (we have seen this happen several times in the past, even with well run HPSS systems). Clients do not see this interruption, except for slower service.

HRM does not need HPSS to be grid-aware. It can run on a grid-aware machine independent of HPSS, and performs all communication with HPSS internally to the HPSS site. At the same time, HRM performs all transfer between sites in a secure grid-aware fashion.

The function implemented are using HRM to get/put single/multiple files from/to HPSS through HRM. In addition, a third party copy function for single/multiple files was implemented. Two administrative function: list and status were also implemented to find out the dynamic status of file being replicated. The functions are: hrm-get, hrm-mget, hrm-put, hrm-mput, hrm-copy, hrm-mcopy, hrm-ls, and hrm-status. A user guide describing the command-line interface was written.

## 4.9 SDSC – SRB

Significant enhancements have been made to the SRB data grid technology to improve performance. Of critical need is the ability for bulk manipulation of attributes associated with data entities that are being registered into the data grid. Basically, the utility of collections is determined by how rapidly new data entities can be registered, extracted, and manipulated. In a development effort that was also funded by the DOE ASCI Data Visualization Corridor project, the SRB data registration utilities were significantly improved. Through a single registration request, it is now possible to aggregate metadata attributes about data entities into bulk load commands into the MCAT metadata catalog. The bulk load is multithreaded, with a user-defined number of simultaneous streams of attribute loading. Registration rates of 250 files per second were measured on the production collection management systems at SDSC (Wan, Rajasekar).

This capability will be extended to support bulk metadata extraction from the MCAT catalog. This is of particular use when SRB containers are used to aggregate small files into a single physical file. By storing an XML file denoting the location and length of each data entity within the container, bulk manipulation of the container then becomes possible without having to reference the metadata catalog. This capability is of greatest use when the data entities are being streamed through an analysis platform.

The SRB data management system is being augmented with a web services interface. The interface will be based upon the common capabilities identified across the high-energy physics data grids. A first simple prototype has been developed at SDSC (Jagatheesan, Zhu, Cowart) that provides services for replicating files and adding files to a logical collection. We will work with Chip Watson to identify common parameters that should be used by each web service, and then implement the capability on top of the SRB infrastructure. This will provide a standard WSDL/SOAP interface for data and information management.

#### 4.9.1 SDSC - JLab Replica management interface

An initial comparison of the capabilities of five data grids that are used to support high-energy physics has been completed. The data grids include the Storage Resource Broker (SRB) data grid from the San Diego Supercomputer Center, the GDMP data replication tool (a project in common between the European DataGrid and the Particle Physics Data Grid, augmented with an additional product of the European DataGrid for storing and retrieving meta-data in relational databases called Spitfire), the Globus toolkit, the Sequential Access using Metadata (SAM) data grid from Fermi National Accelerator Laboratory, and the JASMine data grid from Jefferson National Laboratory. Over 120 different features organized into 11 different categories are being supported by at least one of the data grids. Over 75% of the features are present in at least two of the data grids, with 50% of the features present in the majority of the data grids.

All of the data grids are implemented through a logical name space that is independent of the local storage system name space, with extensions to Unix commands based upon additional attributes managed in the logical name space. Extensions include latency management functions, creation of discipline specific attributes, and attribute based discovery.

#### 4.9.2 GridPortal project at SDSC.

The Grid Portal developed at SDSC supports execution of jobs within the Globus grid environment. Jobs run under Globus typically generate local files, which must then be either copied into an archive or registered into a data collection. Effectively, three data management environments are needed: an execution environment tuned to support high-performance access to data residing on local systems, a data grid that provides mechanisms for latency management over wide area networks, and a persistent collection environment for publication or storage of results. These three data management systems can be coupled within a grid portal, automating access to each of the environments. The current version of the SDSC Grid Portal supports access to local files stored on any resource accessed by Globus and to files registered into a Storage Resource Broker collection. A demonstration has been done of the automated registration of output files into the SRB, and their replication into an archive.

The Grid Portal is being restructured to use WSDL services as the interface to the SRB collections. This will improve the ability of the system to support alternate implementations of persistent collections. Note that this will require defining services for the creation and maintenance of a collection, as well as services for registering output files into an existing collection and services for discovering data entities within a collection. This project is being supported by additional projects, including the NASA Information Power Grid, the NSF Distributed Terascale Facility, and the NSF NPACI program.

### 4.10 . Wisconsin - Condor

We assisted Jim Amundson with the PPDG-MOP demo at SC2001, and John Weigand with the development of a new DAG visualization tool used in the demo. We also participated in the SC2001 Bandwidth Challenge.

Continued collaboration with Greg Graham and Jim Amundson at Fermi to improve the interoperability of Condor-G with the IMPALA and MOP software, and to enhance the reliability and robustness of distributed CMS production. We held a meeting in December with a number of Condor and Fermi staff to understand the architectures of IMPALA, MOP, and Condor-G, and BOSS, and to identify specific technical areas for collaboration.

We explored distributed error-propagation and reporting issues, and prepared a talk for presentation at January's PPDG focus group meeting at JLab.

We participated in the Joint Technical Board meetings, and in the planning of a US-CMS grid testbed utilizing PPDG software.

## 5 Appendix

### 5.1 List of participants

TEAM	Name	F	Current Role	CS	1	2	3	4	5	6	7	8	9	10
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 collaborator				x							
	William Alcock	Y								x				
ATLAS	Torre Wenaus	N	ATLAS Team Lead.		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													
	Dantong Yu	Y	Monitoring				x							
STAR	M. Messer	N	STAR Team Lead											
	Eric Hjort	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											
	Tokako Hickey	Y	CS-8:Robust Job Scheduling, GriPhyN collaborator										x	
	Conrad Steenberg	Y	CS-8:Analysis Tools, GriPhyN collaborator										x	
	Koen Holtman	N	GriPhyN collaborator											
	Iosif Legrand	N	CS-8:Monitoring Tools										x	
	Vladimir Litvin	N	GriPhyN collaborator		x	x								
	Jim Amundson	Y				x								
Shazhad Muzzafar	Y							x						
James Branson	N	CMS Tier 2 manager												
Ian Fisk	N	CMS Level 2 CAS manager, iVDGL liaison												



## 5.2 SuperComputing 2001 demonstrations related to PPDG

CMS Simulation Production – IMPALA and GDMP	FNAL, Caltech	Demonstration of current CMS simulation production tools and GDMP replication tools
CMS Distributed Simulation Production (MOP)	Caltech, FNAL, Wisconsin, ANL, UCSD	Use of Condor-G/DAGMAN to automatically run CMS simulation production at multiple sites
Bandwidth Greedy Grid-enabled Object Collection Analysis for Particle Physics	Caltech, UCSD	Demonstration of the use of Grid tools and virtual data to support interactive physics analysis.
Reliable Transport	ANL	Extensions to the transport layer of GridFTP to support retry
Proxy Server Demo	ANL, SLAC	Demonstration of replica catalog proxy server
GriPhyN Virtual Data (CMS)	ANL, Florida	Generation of CMS simulation scripts from definition of physics parameters
Globus CAS prototype	ANL	Use of Community Authorization Service in Earth Sciences Grid
"Letting Scientists Concentrate on Science: Providing a Transparent View of Data on the Grid"	LBNL	<a href="http://gizmo.lbl.gov/~arie/sc2001.demo/slides/index.htm">http://gizmo.lbl.gov/~arie/sc2001.demo/slides/index.htm</a> <a href="http://gizmo.lbl.gov/~arie/sc2001.demo/poster.pdf">http://gizmo.lbl.gov/~arie/sc2001.demo/poster.pdf</a>
“Bandwidth to the World”	SLAC/FNAL	The "Bandwidth to the World" project is designed to demonstrate the current data transfer capabilities to about 25 sites with high performance links, worldwide. ( <a href="http://www-iepm.slac.stanford.edu/monitoring/bulk/sc2001/">http://www-iepm.slac.stanford.edu/monitoring/bulk/sc2001/</a> )
SDSC Grid Portals Architecture	SDSC	The SRB team has been working with Grid Portal Architecture group to use SRB in building Grid Portal services.

### 5.3 International HENP Grid Coordination and Joint Development Framework

Guy Womers/Draft 20CT01

{PRIVATE }

The HEPN Grid R&D projects (initially DataGrid, GriPhyN, and PPDG, as well as the national European Grid projects in UK, Italy, Netherlands and France) have agreed to coordinate their efforts to design, develop and deploy a consistent standards-based global Grid infrastructure. The guidelines for coordination and joint development by the projects are enunciated below. This collaborative effort can be referred to as INTERGRID.

#### Preamble

The consortia developing Grid systems for current and next generation high energy and nuclear physics experiments, as well as applications in the earth sciences and biology, have recognized that close collaboration and joint development is necessary in order to meet their mutual scientific and technical goals. A framework of joint technical development and coordinated management is therefore required to ensure that the systems developed will interoperate seamlessly to meet the needs of the experiments, and that no significant divergences preventing this interoperation will arise in their architecture or implementation.

To that effect, their common efforts will be organized in three major areas:

- An InterGrid Management Board (IGMB) for high level coordination
- A Joint Technical Board (JTB)
- Common Projects, and Task Forces to address needs in specific technical areas

#### A/ IGMB (Intergrid Management Board)

##### A.1 IGMB Role

- Information exchange on the status, plans and issues facing national and regional Grid initiatives

- Periodic review of key developments and directions in the Grid projects, with particular attention to maintaining convergence and interoperability, including review of the Common Projects
- Set up a legal framework for collaboration, covering intellectual property rights and associated issues
- Organizes Common Events (Workshops, Seminars, etc.)
- Proposes joint submissions of items to external bids, where appropriate
- Receives regular reports from the Joint Technical Board
- Approves the list of common projects and ad hoc task forces, proposed by the JTB

#### A.2 IGMB Composition

The IGMB is presently composed of the combined Management Boards of the DataGrid, PPDG and GriPhyN projects. It will be extended to represent new Grid projects as they come along.

#### A.3 IGMB Meetings

Three times per year, **synchronised as much as possible with Global Grid Forum meetings**

#### A.4 Chairmanship

The IGMB will elect a chairman , who will serve for one year.

### {PRIVATE }B. Joint Technical Board

---

#### B.1 Role

- Ensure compatibility and interoperability of Grid tools
- Clearly identify API, interfaces
- Launch task forces on specific issues (such as networking, architectural issues, security, ...)
- Reviews the common projects
- Reports to the InterGrid Management Board
- Ensures good contact with the various Grid forums, especially the Global Grid Forum working groups

#### B.2 Composition

6 members for European GRID projects, 6 for US GRID projects and 2 for Asian Pacific projects

## B.3 Chairmanship

One year term

## B.4 Meetings

At least 4 times per year, using teleconferencing as needed

## Common Projects

Common projects are specific well-focused joint efforts on a small number of key issues, or sets of issues.

### C.1 Scope

Common projects will normally take one of two forms:

- Joint development of specific Grid services or components targeted at one or more large HENP experiments involving US and Europe partners
- Realisation of dedicated transatlantic testbeds for software development, network tests, etc.

Testbeds will normally be linked to a well-specified development program with deliverables, and will be targeted at near or medium term goals of the targeted experiment(s)

### C.2 **Liaison Team**

A liaison team will be appointed for each project by the IGMB and the relevant partners.

- Role

The role of the liaison team will be to develop a reasonable work plan with precise milestones and deliverables for each partner (Grid consortium and HENP experiment), and manpower requests from each partner. The work plan will be reviewed by the Joint Technical Board and approved by the IGMB.

- Composition

The liaison team for a specific project will include one member from each Grid consortium involved, and if a HENP experiment is involved,

one European member and one US member of this experiment. The liaison team will designate its chair for interaction with the Technical Panel

### C.3 Reviews

The project will be reviewed at regular intervals by the Joint Technical Board.

## 5.4 Appendix - PPDG Meetings

Below is the list of PPDG meetings, which are all teleconference mode this quarter. Where web pages are available, the URL is shown. The complete list of PPDG meetings is maintained in the calendar at the URL <http://www.ppdg.net/cgi-bin/we4.0/webevent.cgi?cmd=openical&cal=cal2>.

Oct 3	12:30 p.m. - 2 p.m.	<u>PPDG phone conference</u> <b>URL:</b> <a href="http://www.ppdg.net/mtgs/phone/011003/default.htm">http://www.ppdg.net/mtgs/phone/011003/default.htm</a>
Oct 10	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> <b>URL:</b> <a href="http://www.ppdg.net/mtgs/phone/011010/default.htm">http://www.ppdg.net/mtgs/phone/011010/default.htm</a>
Oct 17	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> <b>URL:</b> <a href="http://www.ppdg.net/mtgs/phone/011017/default.htm">http://www.ppdg.net/mtgs/phone/011017/default.htm</a>
Oct 24	12:30 p.m. - 2:30 p.m.	<u>PPDG steering committee</u>
Nov 7	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> <b>URL:</b> <a href="http://www.ppdg.net/mtgs/phone/011107/default.htm">http://www.ppdg.net/mtgs/phone/011107/default.htm</a>
Nov 21	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> <b>URL:</b> <a href="http://www.ppdg.net/mtgs/phone/011121/default.htm">http://www.ppdg.net/mtgs/phone/011121/default.htm</a>
Nov 28	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> <b>URL:</b> <a href="http://www.ppdg.net/mtgs/phone/011128/default.htm">http://www.ppdg.net/mtgs/phone/011128/default.htm</a>
Dec 3	2 p.m. - 4 p.m.	<u>Robust Replication pre-meeting call</u>
Dec 5	12:30 p.m. - 2:30 p.m.	<u>PPDG steering committee</u>
Dec 10	2 p.m. - 4 p.m.	<u>Replication Focus teleconference</u>
Dec 12	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> <b>URL:</b> <a href="http://www.ppdg.net/mtgs/phone/011212/default.htm">http://www.ppdg.net/mtgs/phone/011212/default.htm</a>
Dec 17	1 p.m. - 3 p.m.	<u>Replication Focus teleconference</u>
Dec 19	12:30 p.m. - 2:30 p.m.	<u>PPDG weekly phone meeting</u> <b>URL:</b> <a href="http://www.ppdg.net/mtgs/phone/011219/default.htm">http://www.ppdg.net/mtgs/phone/011219/default.htm</a>
Dec 31	1 p.m. - 3 p.m.	<u>Replication Focus teleconference</u>