Grid3 or Grid2003

Version	Date	Author	Status
21	November 5, 2003	rwg	Add BTeV description from Dan Engh
20	October 15	rwg	Review of progress at 2nd Face to Face Meeting. See comments in **boldface blue.**
19	August 15th	RP	Enough is enough
18	July 25th	RP	Edits from LatB, RP
17	July 24th	RP	
16	July 22, 2003	Rwg	Incorporated text from Ian Fisk, Kaushik De, Greg Graham
15	July 20, 2003	rp	Schedule mods
14	July 15-17, 2003	ah, rwg, rp	RWG/RP: edits from Annes document.
13	July 14, 2003	lf, rwg	CMS section merged with Rich's v12; Monitoring section
12	July 11, 2003	rb	Reformat section 1 Condense section 3.2 Eliminate sections 3.6, 3.7, moving issues to 3.2 Remove section 3.8 Remove table in section 4 Added Section 10 Issues
11	July 10, 2003	rwg	Added SDSS

10	July 8, 2003	rp	Include Scotts changes and Add some names
9	July 7, 2003	rwg	Added Netlogger section from Brian Tierney.
8	July 7, 2003	rp, rwg	Grid shell section, minor updates.
7	July 3, 2003	Rwg	Minor updates.
6	July 1, 2003	rwg	Update experiment sections, effort table, management section, schedule section, sections from Leigh.
5	June 19	rp	Add first attempt at a schedule
4	June 5, 2003	rwg, latb, rp	Modifications of goals after steering meeting
3	June 3, 2003	rwg, latb	Metrics table restored; other minor edits
2	June 3, 2003	rwg	Revised, inserting questions where possible problematic statements were.
1.1-shank	May31	js	Edits before passing the token
1.1	May31	rp	Edits before passing token
1	May 30, 2003	rwg, latb, rp, js	Initial draft

Contents

1 Overview 3
1.1 Purpose and Goals 3
1.2 Requirements and Strategies 3
1.3 Grid3 and the LCG 6
2 Project Organization 6
2.1 Management and Coordination 6
2.2 Project Member Responsibilities 7
3 Design 7
3.1 Services 7
3.2 Site Installation and Configuration Procedures 7
3.3 Support and Operations 8
3.4 Monitoring and Information Services 8
3.4.1 Resource Publication and Discovery 9
3.4.2 Dynamic Resource Monitoring 10
3.4.3 Task, Work, and Workflow Monitoring 10
3.5 Virtual Organization (VO) Management 11
4 Application Milestones 12
4.1 ATLAS Challenge Problems 12
4.2 CMS Challenge Problems 14
4.3 SDSS Challenge Problems 16
4.4 LIGO Challenge Problems 17
4.5 CS Challenge Problems 18

4.5.1 VDT Scalability and Performance Experiments 18
4.5.2 Scheduling Algorithm Experiments 18
4.5.3 NetLogger and Grid3 18
4.6 BTeV Challenge Problems 19
5 Metrics and Throughput Parameters 19
6 Sites and Resources 20
7 Schedule 20
8 Effort 23
9 Dissemination 25
10 Issues 26
11 Next Steps 26
12 Appendix 26
12.1 iVDGL Architecture 26
12.2 iVDGL Year 2 proposal: Demonstrate Value in Application Experiments 27
12.3 Task Force 28
13 References 28

1 Overview

1.1 Purpose and Goals

The purpose of the Grid3 is a project to build a grid environment to:

- Provide the next phase of the iVDGL Laboratory
- Provide the infrastructure and services needed to demonstrate LHC production and analysis applications running at scale in a common grid environment
- Provide a platform for computer science technology demonstrators
- Provide a common grid environment for LIGO and SDSS applications

The goal of the project is to integrate, deploy and apply a functional grid across (at least the) LHC institutions, extending to non-LHC institutions and to international sites, working closely together with the existing efforts. We plan to demonstrate the functionalities and capabilities of this Grid, according to some well-defined metrics using specific ATLAS, CMS, SDSS, LIGO and CS applications. The application demonstrators are discussed in sections below. We expect to work within the existing VDT-based US LHC Grid efforts and federating with the LCG and other grid efforts, towards a functional and complete Grid.

The Grid3 project major demonstration milestone coincides with the SC2003 conference, and the Grid3 project itself would end at the end of 2003 with the expectation that the Grid3 infrastructure continues to be available for applications and scheduled iVDGL work. A follow-up work plan will be defined well in advance of the project's end.

1.2 Requirements and Strategies

The following initial list of requirements, components, constraints and strategies will be reviewed and revised throughout the project

- Experiments must be able to effectively interoperate and run their applications on non-dedicated resources. **Done**
- Applications must be able to install themselves dynamically, thereby imposing minimum requirements on grid facility managers. **Done**
- We will work within a grid architecture consisting of facilities (e.g., execution and storage sites), services (e.g., a prototype operations center, an information system for resource publication and discovery, and so on), and applications (such as those listed in section 4). This structure is roughly indicated in the iVDGL proposal (ref). done It will be necessary to detail the grid architecture and conduct a review by the Stakeholders (iVDGL, GriPhyN and PPDG Steering Committees and US ATLAS and US CMS software and computing managers). **Did not attempt.**
- The middleware will be based on VDT 1.1.9 or greater, and will use components from other providers as appropriate. Which components of the common middleware will be different for each ofthe applications. **Done, VDT 1.1.11.**
- Packaging, installation and configuration should be simple and easy for system managers - running a single Pacman script for example. Simple update capabilities should be available, without having to reinstall. **Done**
- An information service for resource publication and discovery based on MDS will be deployed. A team will collect requirements for schema definitions and information providers to ensure correct operation of all intended services. **Done**
- A simple monitoring service will be deployed based on Ganglia, version 2.5.3 or greater (supporting hierarchy, grid-level collections), with collectors at one or more of the operations centers. Additionally, Ganglia information providers to MDS may be deployed if required, and native Ganglia installations may report to other collectors corresponding to other (logical) grids. An example is a collector at Fermilab providing a Ganglia grid-view of the US CMS managed resources. **Done, through ML**.
- Other monitoring systems, such as MonaLisa or a workflow monitoring system, may be deployed. Guidelines for information providers, sensor installation, and pacmanized versions of these with clear instructions would be followed. **Done.**
- A consistent method or set of recommendations for packaging, installing, configuring, and creating run-time environments for applications will be provided, subject to review. This can include, for example:
    a) Pacman caches and instructions for pre-installation of application libraries. **Done**
    b) Use of the Condor grid shell, job-wrapping mechanism. Merge with GridLaunch in progress, still in development. **No longer a Grid3 issue**.
    c) Precise instructions for application environments. **Done**
    d) Use of the Chimera Gridlauncher. **No longer a Grid3 issue**
    e) Use of the WorldGrid project mechanism. **Did not attempt.**
- One or more VO management mechanisms for authentication and authorization will be used.

a) The VOMS server method as developed by the VOX project[1]. **Done**.

b) The WorldGrid project method as developed by Pacman[2]. **Not done**.

c) A fallback solution is to use LDAP VO servers, one for each VO containing the DN's of the expected application users[3]. **Not needed.**

d) An acceptable use policy to be signed by all participants. In progress.

- Support for replica location services as required by application components, such as Magda for ATLAS. An example would be an LRC, indexed to the experiment's RLS. To be discussed. **Moved into the domain of the VO application.**
- A simple user-support model, and other service requirements such as maintaining a trouble reporting system, liaison functions, etc., will be developed and reviewed for the operations center. **On-going**.
- A common reporting and event logging tool or framework, possibly with a graphical user interface, to be used by the application teams, will be used to collect information about task workflow performance, statistics, etc. This will be used for archival purposes, grid-viewing data displays, feedback to the CS teams, and will form the basis for input to computing reviews, e.g., SC2003. **In progress. Show demo, first discussion on Thursday, 1pm.**
- Data replication and data movement services, for example a set of sites may deploy a storage element based on SRM, Gridftp, and dCache, as required by the applications.
    - **Transdemo is exercising data movement, and it is intended to setup two SRMs (one at UCSD, one at FNAL).**
    - **Incorporation of Scott's Gridftp demo.**
- Grid3 will use existing and emerging services as far as possible. Given the aggressive milestones there is little time for development. **Followed.**
- Challenge problems will be deployed on Grid3 after being demonstrated to run effectively on the grid testbeds of the responsible groups. **Done.**
- Grid3 will attempt to deploy consistent and compatible grid services. Our strategy is to interface with existing computing installations on a few well defined interfaces: batch queues, information service providers, and storage elements, etc. It is our intention to have as little impact as possible on the computing resources. We will bring the environment with us whenever possible and assume as little as possible about the existing setup. **Followed. In progress.**

---

[1] From the VOX project

[2] from the PACMAN module of the iVDGL Core Software working team.

[3] Using Grid3 effort.

1.3 Grid3 and the LCG

Grid3 should federate with the LCG. We will have a working group within Grid3 to understand what is needed for basic interoperability, specifically submission of jobs and movement of files across the grids.

There will be other issues that may affect interoperability - consistent replica management, virtual organization support and optimizing of resource usage across federated grids. We do not have the effort to address all these during the Grid3 project itself. We will identify, discuss and document such issues for collaborative work with the LCG. Many of the working areas in Grid3 are already joint projects between the LCG and Trillium or the S&C projects. Additional collaboration in areas of monitoring and operations have been discussed over the past few months. As we proceed to better understand the technical plans the expectation is that we will propose further areas of joint work.

- **Initial plans to merge Grid3 and LCG schema in place.**
- **Agreed cooperation on US ATLAS and US CMS data transfer projects for Grid3. Further demonstrations for post-Grid3 discussed.**

2 Project Organization

The project management and organization will be as lightweight as possible. We plan to use existing management structures and organizations of the participating projects where possible.

2.1 Management and Coordination

A management team will need to be formed by consensus of the Stakeholders. The team will meet weekly to agree on tasks and deliverables, and publish issues and status to the participating organizations. Team members will act as liaisons and report to and from their wider project and collaboration. The team will act as the liaison to the LCG and other international partners, and will coordinate with the LCG as much as possible.

We recommend that the Taskforce (see section 12.3) carry over to become the management team, and thereby assume coordination responsibilities. This maintains the necessary representation from all Stakeholder organizations.

The project management and Stakeholders will meet as needed to reassess the project scope, goals, and progress, to endeavor to resolve any outstanding issues, and to arrange for any necessary additional manpower.

US ATLAS will contribute manpower through the Facilities subproject. Development and integration are managed through GTS (Grid Tools and Services) and Grid3 test deployments will be done using the US ATLAS Development Test Grid (DTG). Deployments onto the US ATLAS production grid facilties, and management of production services (generation, simulation and reconstruction) is managed through the US ATLAS Production sub-project.

USCMS will contribute R&D manpower through the DPE (Distributed Processing Environment) project. Effort will be tracked simultaneously by Grid3 management and DPE management using clearly defined DPE sub-projects. In this way, we can fully exploit the inherent synergies that exist between USCMS DPE and Grid3 while allowing DPE to fulfill its mission to USCMS.

- **All of the above in place and functioning well.**

2.2 Project Member Responsibilities

Members of the project will be expected to contribute their effort as agreed and give Grid3 the priority it deserves, as necessary, in order to achieve the goals and milestones. This will be especially important at the time of SC2003, and during other agreed upon demonstration or data challenge activities. Each project member is responsible for alerting his or her management team representative regarding issues of availability, priorities, assignments and/or expectations. Additional manpower needs will be submitted to the project management team.

Success of the project requires at least one face-to-face working week. We have scheduled one for August 18-22, at Fermilab. All members are expected to attend this and others as scheduled. Each institution will sponsor the travel for its participants.

- **Two meetings took place. Not all project members attended, though critical mass was achieved in both meetings.**

3 Design

The iVDGL grid design includes the concept of an iSite, a model for a heterogenous collection of processors and storage elements (Fig. 1, Appendix). Grid3 needs to determine and provide details for this design.

3.1 Services

The services envisioned in the design are:

Management services which enable applications to allocate and manage computational and storage resources at the sites. The allocation mechanism can be achieved through one of the VO processes listed above (to be reviewed).

Monitoring services which support discovery of the existence, configuration, and state of the grid resources. MDS with the GLUE schema, and other information as necessary at all iSites is planned as the default information system. Individual iSites, of course, may wish to publish their configuration to other (logical) grid GIIS's. In addition, dynamic monitoring will be provided by the Ganglia tool.

Control services which support global experiment data management and testbed configuration. A first step towards control services might consist of the LCAS callout in gatekeeper, and the VOMS services as specified by the VOX team.

3.2 Site Installation and Configuration Procedures

Grid3 will include deployment of uniform installation and configuration procedures for the sites. We need to determine procedures regarding VDT for the following:

- Installation procedures
- Configuration
- Post-installation testing
- Certification

Conventions and procedures will be documented and provided to site managers and operators to ensure a consistent, uniform Grid3 environment from site to site.

A plan needs to be devised which allows a given resource to participate in more than one grid application environment. At present, this is allowed by either equipping a Grid3 gatekeeper with multiple job managers which reference separate job queues and publish their availability to one or more GIIS's, or by establishing multiple gatekeepers which register to individual GIIS's (a Grid3 GIIS and a VO-specific GIIS). We need to formulate a set of requirements/procedures/options to provide to site managers for each of these options.

3.3 Support and Operations

The deployment and operation of the Grid3 environment will require some centralized support activities. We envision that operations groups from each VO will coordinate with the iGOC to lead this effort. Given our constraints in schedule and effort, we will have a simple and pragmatic support model where the application demonstrator teams will support their applications, the middleware technology providers will support their software, the site administrators will operate and manage all aspects of their grid sites, and ongoing effort to identify and address the end to end diagnosis and problem finding and resolution will be the responsibility of the Grid3 project as a whole. The scope of work and the issues to consider are the following:

- Create pacman repository to hold Grid3 packages. **Done**
- Create cluster/computer resource installation instructions. **Done.**
- Create and implement policy agreements with regard to acceptable use of a cluster or resource. **Done.**
- Track resource availability (including new hardware procurements at participating sites) and coordinate schedules for resource usage.
- Negotiate usage policies and priorities with site administrators. **In progress, through VO managers.**
- Create user forum at the iGOC in order to provide an informal method of consulting with grid users, grid resource providers and grid developers. **Email lists used extensively.**
- Create "weather map", a graphical representation of current activity at each site.
- Allow trouble tickets to be created by collaborators at the iGOC. **In progress; proposal to use project Savannah portal bugzilla, with summary digests emailed daily**.
- Monitor various grid services such as
    - VO server
    - RLS servers
    - GIIS servers
    - Web Server
    - Dynamic monitoring data gathering server
    - **➔ plan to use MonaLisa to monitor these services.**

3.4 Monitoring and Information Services

The monitoring and information services for Grid3 will be based as far as possible on existing or emerging software, which will require a suite of different monitoring tools to be used for various functions. As much as possible common tools will be candidates for inclusion in the common middleware deployments. We focus our attention on three primary areas:

1. An information infrastructure for resource publication and discovery. **Done**
2. "Dynamic" resource monitoring such as the Ganglia and/or Nagios monitoring system. As demonstrated at SC2002, VO-Ganglia, which collects information in a hierarchy and sorts by VO, provides a basis for a policy monitoring infrastructure. **Ganglia deployed. Job statistics collection agents for MonaLisa under development.**
3. Task workflow monitoring - including high-level workflow monitoring tools and frameworks to manage specific physics workflow tasks. This work will be greatly enhanced by use of an advanced Grid Shell environment provided by the Condor and/or Chimera teams. Other instrumented environments such as BOSS could be used to provide information to an overall monitoring/reporting framework. **No longer considered.**

3.4.1 Resource Publication and Discovery

The grid resources at each site in the Grid3 environment will be equipped to provide information about their environments, intended functions, capacities, etc., to a known set of information servers. Application clients will query these information servers to determine things such as the correct run time environment on a particular remote compute element. The VDT core software will provide the Globus MDS software service to each cluster and/or compute element at a site. The MDS component will be the basis for the collection of specific static information using the LDAP schema framework and the index information search and retrieval capabilities of that framework.

Review and summarize requirements for information services from application groups, and provide initial work plan for each site including plans for packaging, configuration, and testing of a system.

Service status, site certification, monitor tools and frameworks must be considered.

The requirements must address high level architecture of GIIS hierarchies, or reporting to multiple GIIS's at a site.

Two specific tasks are identified: (**both accomplished**)

- Generation of a Compute Resource Specification for MDS. This involves:

  1. Specification of base distribution VDT MDS schema.
  2. Specification of newest GLUE schemas for CE/SE and CE-SE Bind
  3. Specification of Integration Information Providers:
     - GLUE Information Provider
     - Fabric Monitoring Information Provider (for Ganglia)
     - Application schemas and Information Provider
     - Software Package Information Provider (Pippy)
  4. Registration for Clusters to multiple index services.

- Generation of a set of servers to provide indexed information services. This involves:

  1. Installation of GIIS service on reliable hardware

2. Specific configuration of web portal(s) whose function is to highlight the information pertinent to the applications and end users.
3. Configuration of LDAP browser for the index service.

3.4.2 Dynamic Resource Monitoring

Grid3 must include a discussion and decision on which tools from Ganglia, Mona Lisa, and Nagios monitoring systems in use by testbed groups will be migrated to Grid3. This will include Ganglia information providers to MDS. We have identified two sets of tasks in this area:

First,

1. Pacmanize installation of monitoring system (Ganglia 2.5.4 completed; documentation and instructions still needed.)
2. Deploy this installation on clusters and/or compute elements with collecting data source at the iGOC operations center (goc-mon.uits.iupui.edu).
3. Report the configuration to Tier 1 and other Testbed collectors.

And secondly,

1. Setup collection host (done).
2. Archive collected data.
3. Allow replay of collected data.
4. Display "weather map" of dynamic data.

3.4.3 Task, Work, and Workflow Monitoring

There is a critical need for a structured monitoring framework to collect, aggregate, archive, view and report sensor data from application- and middleware-levels of the grid execution stack. It is a requirement of the Grid3 project to collect and summarize performance results in order to gauge the grid's overall response to workload requests. As we've see from recent grid production runs, users are additionally demanding more tools to receive feedback about progress, errors, completion status, and other stateful information of the submitted tasks. Some tools such as the gahp shadow server mechanism of CondorG to some extent. Aggregation and reporting tools are needed to manage task execution from many perspectives (production manager, end-user physicist, middleware developer). Task, work, and workflow monitoring is discussed in a separate document, see http://www.ivdgl.org/grid3/Grid3_mon_v2.doc.

The Grid3 project must create a framework for monitoring, collecting, aggregating, and reporting information, from multiple remote sites/VO's/applications. For remote monitoring, we plan to use the Condor Grid Shell environment. The Chimera GridLaunch/Kickstart has been merged into this environment. This is a component of VDT which creates an environment in which channels are provided for sending remote monitoring semaphores and conditions back to the (end-user) submit host. The Condor grid shell environment is designed to capture stateful information for the user.

It is envisioned that an integration of this system with Monalisa framework is possible, and should be explored.

The initial version of the Condor Grid Shell is to be delivered Aug 2003 by the Condor team, and supported thereafter. Grid3 will utilize it as a "job wrapper" on execution CPUs. The wrapper will serve to generate and report generic job statistics (return value, cpu and i/o usage, etc.) and to transport this information back to the original submission site. We envision one delivery mechanism for this summarized information as being a web portal. Additional features, beyond the initial, core requirements, will be added in subsequent releases.

Job monitoring information will be divided into three broad categories, in decreasing order of priority:

1) Coarse, application-independent, job life-cycle events (e.g., submitted at submit site, submitted at execution site, started, completed, failed)

2) Finer-grained, application-independent job life-cycle events (e.g., input staged to execution site, job restarted, output staged back to submit site, etc.)

3) Fine-grained, application-specific job run-time events (e.g., individual physics event completion, etc.).

The grid shell will provide #1 initially and #2 roughly midway through the Grid3 project, and application-specific monitoring tools may provide #3.

3.5 Virtual Organization (VO) Management

The basic outline of work regarding VO management includes (see further references at the end of the document):

- Review VO solutions and consider candidate systems, configuration, and deployment for Grid3.
- Consider full end-to-end requirements of application users.
- Consider multi-VO issues, such as policy and authorization.
- Consider, for ATLAS and CMS, use of LCG VO registration infrastructure.
- Define a VO solution for Grid3. **Done**

The role of VO management in the Grid3 will likely encompass:

- Establishing that a user with a valid Certificate is part of an approved experiment.
- Enabling sites to have the ability to control who has access to resources, on a user-by-user basis.
- Allowing a mechanism to grant special access to resources for specified groups within experiments.

The current VO management solution uses an LDAP server in which “groups” define VO participants and identify them on the server. A resource needing to authorize a user then queries the LDAP server(s) using some version or variation of the mkgridmap script. The query to the LDAP server populates a static grid-mapfile.

VDT 1.1.9 includes software that allows a resource to increase the granularity of the user authorization process. It provides an authorization callout within the context of the Globus-gatekeeper that allows resources to implement additional local checking of an incoming user certificate proxy. Also included in VDT 1.1.9 are the "edg-lcas modules" which provide "callout stubs" for the globus-gatekeeper; this further simplifies use of the authorization callout by adding authorization modules for banned and allowed users, and for timeslots.

If the edg-voms client mechanism (edg-voms-proxy-init) is distributed to and used by application users to create an extended user proxy, and if an edg-voms server is implemented by the VO, then even more granularity may be obtained. The extended certificates are checked by the authorization callout to verify VO and other attributes, thus circumventing the static grid-mapfile dependency.

The set oftasks for the client and server sides will be discussed and decided upon during the project. [4]

4 Application Milestones

Grid3 will be aligned with the application milestones (such as US ATLAS and CMS data challenges) which are detailed below. Additionally the iVDGL has project milestones which will be met to demonstrate a working, multi-application Grid laboratory.

4.1 ATLAS Challenge Problems

ATLAS will continue official data challenge efforts in July-August to do final reconstruction of DC1 (Data Challenge 1) datasets for high level trigger (HLT) studies, leading to presentation of the HLT TDR in September. ATLAS will also generate new top samples for the ATLAS SUSY and Top physics groups in July-August. Pre-DC2 (Data Challenge 2) exercises are anticipated for the Fall, with focus on analysis of distributed datasets. Well defined analysis tasks will be designed using a number of interfaces designed for physicist users. Emphasis will be on using incremental LCG services, involvement of Tier 0, 1, 2 centers, dataset replication and transformations. Analysis challenges will involve SUSY searches, Top studies and other physics simulations.

A number of discrete steps are planned: production of Monte Carlo datasets (generation, simulation and pile-up steps), replication, reconstruction, and distributed analysis. The scale of the dataset is 10 million fully simulated, reconstructed, and perhaps "skimmed" events.

1) Monte Carlo production:
   - Implement large scale automated Monte Carlo production, primarily using Chimera-based tools. Some production may also be done using legacy Globus/Condor based tools. In addition to ATLAS testbed sites, we hope to extend production to non-U.S. sites, which may require adding LCG/EDG interfaces (JDL-like DAGs). The basic goal is to have two or more production activities serving as a source of simulation output datasets.
   - Demonstrate distributed VO user registration functionality.
   - Deploy a distributed infrastructure for resource discovery.

[4] The technical details are currently under discussion between the VOX project and Grid3.

- Use a distributed monitoring framework to keep track of resource usage, and real time status of grid production sites.
- Use the requisite packaging and configuration tools for software distribution to grid3 sites.
- Allow an event size for the simulation output of 2-10 MB per event.

2) Collect and archive MC data at BNL.

There has been much discussion about how this is best done. The most likely (workable) scenario is as follows.

- Output files produced in step one are registered with the local LRC (and thus become visible in the RLI running at BNL).
- MAGDA runs a cron job (spider) that looks for new entries in the RLI
- When new files are found, MAGDA copies them to BNL. This does not require or assume that the original output was in a MAGDA-managed location; this way it is compatible with non-dedicated resources.
- MAGDA registers the new replica with the LRC at BNL.
- Chimera keeps tabs on the file copy (this is a separate DAG node from the data production step) and deletes the original output when the copy at BNL is found in the RLI.
- A quality control daemon verifies and validates the data files registered in the BNL RLI. Production will be rescheduled automatically in case of problems with any of the output files. After successfull verification, the daemon will register the information in the ATLAS Metadata Catalogue (AMI).

This is perhaps the part that requires the most thought and development work. It means that we would add a new mode of operation into MAGDA. The jobs would not put the data into a “MAGDA Cache” location because we can't assume that the site actually has such a cache. The output is left on local storage and the location is published as the url in the LRC. MAGDA would need a new type of spider to search the replica catalog to find new files. Then, by policy, MAGDA would copy the file to BNL and register the new copy in the LRC. All of this is a modest change to MAGDA, but would very significantly expand the capabilities. The last step means that the DAG splits the MC job and the file cleanup into two (or more) separate steps in the DAG. After the MC file is produced, the next step is to check the RLI periodically until the new replica is found at BNL. When it is found, check for consistency and then delete the original.

3) Push MC data files to CERN.

This is something MAGDA can already do.

4) Reconstruction at CERN.

This should probably be done with LHC/EDG components. Output will be registered with a CERN LRC that feeds into the RLI running at BNL. Some of the reconstruction may have to be done at other Grid3 sites if sufficient resources are not available to reconstruct 10 million events at CERN.

5) MAGDA "spider" finds new reconstruction output in RLI and copies them to BNL.

Similar to step 2, but this may require interfacing to the EDG RLS, as well.

6) Data reduction at BNL, creating “collections” or “datasets” and skimming out n-tuples. This can be Chimera based, creating objects that are useful for DIAL, possibly distributing data to Tier 2 sites (using MAGDA). A new daemon would be implemented to guarantee consistency of multiple replicas, working together with Magda and RLS.

7) Analysis: Distributed analysis of collections and datasets.

At this time two scenarios are being discussed:

- Use DIAL to analyze distributed, combined n-tuples: DIAL description and timelines are found at (http://www.usatlas.bnl.gov/~dladams/dial/). Discussions have begun to interface the DIAL scheduler which interacts with datasets, applications, and “tasks” with the transformation catalog/virtual data catalog and execution environment provided by Chimera/Pegasus.
- Use Ganga for user-based production; see http://ganga.web.cern.ch/ganga/

4.2 CMS Challenge Problems

In July CMS will begin producing events for the 2004 data challenge (DC04). Fifty million events with minimum bias pile-up at a luminosity of $2x10^{33}$ are needed in the final sample. Creating CMS detector simulation currently consists of 3 steps: (1) event generation with Pythia, (2) event simulation with either a GEANT3- or GEANT4-based simulation application, and finally (3) reconstruction and digitization with the additional pile-up events. The U.S. is responsible for approximately 25% of the total sample. The number of events created within the Grid3 framework will be dependent on the availability of a stable working environment, which is as efficient as the system currently used.

The sample created will be used to validate the CMS software infrastructure by reconstructing events continuously at the rate of 25 Hz for a period of a month. The sample of simulated events will be accumulated at CERN for primary reconstruction, but will be distributed in real time to Tier1 and Tier2 centers for calibration and toy analysis. A number of discrete steps are envisioned to create the CMS sample within the Grid3 environment.

1. Monte Carlo production:

- Implement large scale Monte Carlo production using the US-CMS developed Distributed Production Environment.
    1. Generate Events with CMKIN (Pythia based selection)
    2. Simulate Events with CMSIM (GEANT3 based simulation) and OSCAR (GEANT4-based simulation). The ratio is expected to be 20% CMSIM and 80% OSCAR.
- Demonstrate distributed VO user registration functionality.
- Use a distributed monitoring framework for resource usage.
- Use the requisite packaging and configuration tools.
- Allow an event size for the simulation output of approximately 1.5 MB per event.

2. Transfer and Reconstruction
    - The simulated events will be transferred to FNAL for archiving and reconstruction using Globus-URL-Copy
        - The reconstruction and digitization phase of CMS simulated event production requires specialized infrastructure to serve the randomly sampled minimum bias pile-up events to the signal events. This is not expected to be installed on a generic grid node, therefore, the simulation needs to be transferred to a dedicated site for the final step.
    - As the Grid3 project progresses, US-CMS and the LCG will be working on the development of a storage element (SE). This will be based on the RLS and RLI for cataloguing and discovery, and on SRM for data transfer and virtualizing the storage interface. The SE may be integrated into the Grid3 environment after successful validation within the US-CMS testbed. This would replace the Globus-URL-copy infrastructure currently used.
3. Archiving to CERN
    - Once data is reconstructed and digitized it will be transferred to CERN in preparation for the data challenge.

The current Middleware and Packaging used in the US-CMS Distributed Production environment is defined below.

Master (submitting) Site

- Components are stored in cvs and configured with a series of scripts
- Software Suite includes
    - MC_RunJob a CMS tool for job specification, which specifies CMS Production jobs by reading input from a mySQL database of input parameters
    - MOP for Job Submission, which creates a DAG for the CMS production jobs and submits them through Condor-G

- VDT 1.1.8 (We expect to migrate to VDT 1.1.9)

Processing Site

- The client site is completely configured with a PACMAN cache. This requires only one command from the site administrator
- VDT 1.1.8 (We expect to migrate to VDT 1.1.9)
- CMS Application software packaged in PACMAN. There are efforts to create a second cache for the application software.

4.3 SDSS Challenge Problems

The SDSS Grid3 challenge project will be to construct SkyGrid, the prototype astronomical grid joining JHU and Fermilab The SDSS challenge requires that the environment contain the following components, consistent with Grid3:

- Pacman
- VDT 1.1.9
- Chimera/Pegasus/RLS
- DOE Science Grid certificates
- SAZ
- Pacman caches of application codes

The application codes consist of the SDSS Challenge Problems:

1. The Cluster Finding Challenge Problem: this explores a challenging catalog level problem, finding clusters of galaxies in the SDSS data. We have chosen to tie this intimately to Chimera as a test of the interrelated ideas of virtual data, provenance, and work flow. The result is DAGs of hundreds of thousands of nodes.
2. The Pixel Level Challenge Problems: this explores the problems of analyzing astronomical pixel data. Here the problem is moving around the hundreds of thousands of images needed to do large scale computations on wide area sky surveys. We have three problems in mind: the first is the NVO Galaxy Morphology Demo, in which cutouts of images about galaxies are sent to the grid for a computation, with the aim of adding more columns on an existing catalog. The second is the search for near earth asteroids in the SDSS data, which calls for transferring out complete SDSS images in search of highly elongated objects,. The third is the co-addition of images from the SDSS Southern Survey which is driven by the desire to go deeper than single SDSS images provide; this calls for the transferring of tens of images to the computation site for the re-mapping into another, expanded image, and the actual combining of these images into single final images.
3. The Photo-z Challenge Problem: this explores the combining of computation resources and databases. The exploration here is to extract of millions of objects from a database, and perform a computation on each (the estimation of the redshift of the object given its magnitudes, colors, and errors along with what is known about the distribution of spectral energy distribution in the SDSS data), and the result placed back into the database.

4.4 LIGO Challenge Problems

The LIGO Grid-3 demonstrator project will be to conduct an extensive, all-sky, blind search for continuous wave (pulsar) signals in the LIGO S2 data set. Work is already underway within the GriPhyN-LIGO working group to develop the necessary infrastructure using Pegasus and Chimera to generate the workflows.

The ComputeFStatistic is the code that computes the optimal detection statistic for a periodic signal emitted by an isolated pulsar for which we know position, spin-down rate and frequency (the gw frequency is twice the rotation frequency). If we want to perform blind searches we run the code over and over again with different target parameters. This is a computationally expensive task.

Every search is characterized by:

- source location
- frequency or frequency band
- spindown parameters
- source of interferometer data (L1, H1, H2, GEO)
- start and stop time of the observation

Each search requires that a conventional binary SFT (short Fourier transform) data file be accessible and contain the frequency band that the target signal spans during the observation time. Additional data files that contain the ephemeris data for the year must be present. We expect to stage the necessary SFT and ephemeris files for a collection of jobs to each site in Grid-3 from a LIGO/GEO Tier-1/2 center using GridFTP. Since the data needs to be available to each job that runs at a site it must be placed on a file system available to the job running on a node. If the file system(s) available to jobs on a node is/are not accessible via the internet (perhaps protected by a firewall or on a private network) then the stage must happen in two steps. We expect to publish the location of the staged data in a Globus RLS server so that its location is available to the job.

The minimum size of the data to be analyzed depends on the size of the frequency band B to search. We can estimate that (B+0.5)*2*Tobs*4 bytes is the size of the data set. For a band of 10 Hz and an observation time of 1E7 s, the data volume is 4GB.

The first job in the workflow extracts the necessary frequency band from the SFT data file and saves it to a file. The file path is an input to the ComputeFStatistic code, the second job in the workflow, which outputs the F statistic for a single parameter set. The last job in the workflow stages the output results back to the LIGO/GEO Tier-1/2 centers and creates entries in a database.

Each workflow instance will run for several hours on an average process and will produce a small output. The ComputeFStatistic code can be re-run with the same SFT data set and different parameters, provided that the SFT data set contains the necessary band.

4.5 CS Challenge Problems

4.5.1 VDT Scalability and Performance Experiments

IAN: I propose that we define an aggressive and comprehensive set of experiments designed to evaluate the performance and scalability of individual VDT components, and the robustness of the overall system. Some of these experiments can be combined with application experiments, but others should be distinct. A goal should be to produce a technical document that contains this information. We need someone to take responsibility for defining clear requirements and steps for this work.

4.5.2 Scheduling Algorithm Experiments

IAN: I'd like to see Kavitha Ranganathan use Grid-3 for some comprehensive evaluations of her various scheduling algorithms. We need someone to take responsibility for defining clear requirements and steps for this work.

4.5.3 NetLogger and Grid3

NetLogger can be used to reliably transfer monitoring data to a central monitoring data repository. LBNL will be responsible for the following Grid3 tasks:

- Deploy a "NetLogger Archive Service" on the DOE Science Grid. This will be a "production-level" service that will accept NetLogger data from all Grid3 systems, and store the data in a MySQL database. This data will also be archived on the NERSC HPSS system. The archive will initially have a simple SQL query interface, and will later support XPATH queries via OGSI. Initially supported queries will be based on combinations of host name, program name, timestamp range, and Grid Job ID. The Grid Job ID is an arbitrary user-specified string that is passed to NetLogger to help identify events that are part of the same "Grid Job". The NetLogger Archiver buffers all incoming data to disk before sending to the MySQL server, allowing us to control the incoming load on the SQL server.
- Document and fully support the "NetLogger message forwarding service". This service would typically be deployed one per site, or one per PC cluster. The forwarding service collects all NetLogger data from a local set of hosts, and then forwards this data to the archive. If the archive is unavailable, data is buffered to local disk, so that no data is lost. This service includes a "max disk usage" option to ensure that it does not use too much disk space.
- Work with Bill Allcock, ANL, to complete NetLogger instrumentation of the GridFTP server and the Globus_URL_copy client program. There will be two levels of instrumentation. The default level will generate NetLogger events at program start, end, and on errors. The detailed instrumentation level will generate NetLogger events for all significant I/O.
- Continue to work with Craig Tull (ATLAS/LBNL) to finish adding NetLogger instrumentation to the Atlas Athena code.
- Document and support the NetLogger client library for any other interested Grid3 applications or middleware. A small addition to NetLogger will be a "max disk usage" option to ensure that it does not use too much disk space in the case that the network is down and NetLogger data is going to a backup disk location.

- Add NetLogger to pyGlobus, providing NetLogger instrumentation to pyGlobus-based applications.

All of this should be ready by September 1, 2003.

- **Netlogger-Grid3 demo and services described at**

http://www-didc.lbl.gov/NetLogger/netlogger-grid3.html

4.6 BTeV Challenge Problems

The BTeV Grid-3 challenge problem will demonstrate an integration of BTeV Monte Carlo with GriPhyN virtual data tools. The clarity of the Chimera virtual data toolkit as a BTeV physics interface and the scalability of these tools for large Monte Carlo generation will be tested with mock data challenges run at the scale of the Grid-3 testbed.

The BTeV analysis workflow consists of Pythia event generation, Geant detector simulation, and BTeV reconstruction and analysis code. These steps will be configured in Chimera as separate processes with Chimera/Pegasus/DAGMAN/Condor-g completely defining and controlling the execution of the workflow. The physics sample will contain a generic b-physics signal (bottom-quark particles generated with all known decays) combined with minimum-bias background. The workflow processing time is about 15 seconds per event on a 2GHz machine, translating into a typical request for 2.5 million events generated with 1000 10-hour jobs across the Grid-3 testbed.

The BTeV Grid-3 demonstration code consists of a small set of BTeV executables packaged in Pacman without modification directly from the BTeV code version management (UVM) repository. Perl scripts are used to configure remote application and working directories using information from MDS with Globus. Pacman is invoked remotely to pre-install the BTeV application. A complete BTeV workflow is run remotely in a single (unique) workflow directory. The final stages of the workflow copy data to a Grid-3 storage site, register the files with the BTeV Replica Locator Service, and remove the workflow directory from the remote site.

The BTeV Chimera customization environment consists of a set of scripts that use the Grid-3 MDS schema to build the Chimera resource databases (condor pool configuration and the transformation catalogue). The BTeV Chimera implementation defines the complete provenance for a BTeV Monte Carlo analysis within the Chimera Virtual Data Catalogue (VDC). All physics analysis parameters are defined as VDC parameters so Chimera can generate the required input parameter files from the VDC. All transformations in the transformation catalogue are mapped directly to BTeV executables rather than launching wrapper scripts. This requires that the necessary run-time environment be configured in the transformation catalogue from information obtained from MDS. This also allows a clean separation of BTeV analysis code (versioned with UVM) from the Chimera management of that code so the complete provenance of the workflow is defined in the Virtual Data and Transformation Catalogues with a guarantee that the workflow provenance within Chimera (plus UVM) is complete. With the complete analysis definition found in the VDC, a physicist can use a VDC browser to obtain all the information necessary to reproduce that analysis.

5 Metrics and Throughput Parameters

We will establish metrics and performance targets and we will run demonstrators to measure and characterize the computing activity and gauge the resulting performance of Grid3. These are being

defined at the beginning of the project, but will need better understanding and refining throughout its course. We expect to define sets of metrics matched to the application demonstrators being run. The Grid3 infrastructure will include tools that collect, aggregate, and report measurements. Some examples of metrics which we would ultimately like to achieve are provided in the table below.

Metric (to be defined)	Target (to be defined)	Comments
Number of processors	up to 500	Possibly collected through MDS, archived and time-stamped
Data transferred per day	>2-3 TB	Data "flux" will need to be defined carefully.
Peak number of concurrent jobs	Up to 1000	Collect the total number of jobs running on Grid3, sorted by VO, archived so that time-dependent plots can be made.
Percentage of resources used	up to 90%	
Efficiency of job completion	up to 75%	Success to be defined. Other efficiency metrics could be identified.
Number of users	>10	Collect and sort by VO origin.
Number of different applications	>4	By an application "registry", such as proposed by the WorldGrid "project" mechanism, this could be simplified.
Number of sites running multiple applications	>10	Collect with time stamps, intervals to be determined.
Rate of Faults/Crashes	<1/hour	Measure, and perhaps categorize by degree of severity.
Operational Support Load of full demonstrator	<2 FTEs	How many people (and in which roles) were operating Grid3?

6 Sites and Resources

A list of sites, capacities (CPU/Storage), and use-policy will need to be established at the beginning of the project. This information has been collected in an excel spreadsheet:

http://www.ivdgl.org/grid3/grid3sites.xls

7 Schedule

Our goal to have demonstrators running at SC2003 which begins November 15. This does not mean that all demonstrators or all metrics need be met for this demonstration, but it does imply that a focused effort will be applied to showing significant functionality for this milestone. Applications and demonstrators will be migrated to run on Grid3: the multi-VO/application shared resource Grid – once they have been run stably on an individual application testbed (single VO, dedicated resources). The Grid3 project will take on the effort of integrating the demonstrators to run on shared sites and of doing the small incremental development needed to make this a success. Grid3 will provide tools to measure and display the performance parameters. During the first phase

there will be work to provide a working Grid infrastructure for sharing of resources across VOs while the application groups work on providing end-to-end applications that run for extended periods. Packaging, installation scripts, testing, and documentation are part of all software deliverables.

Some important dates:

- July 1 – "project start"
- July 31 – a Core Grid of sites deployed
- August 18 – Grid3 Integration week
- October 16-17 – GriPhyN/iVDGL all-hands (2nd face to face meeting October 13,14)
- November 15 – SC03

		Status: July 23rd:Later status will be updated in other documents
6/27	Steering Committee approval of Grid3 plan. Formation of project. Identification of people.	*Delayed to 7/27* *Grid3-core mail list defined*
7/1	First weekly meeting of the project. Decide on ~2 initial demonstrator applications (the simplest!) to be run across shared resources on Grid3 sites. Decide on process/testing for adding packages to Grid3 caches.	*Meetings started.* *Athena-reconstruction & CMSsim.* *Started document on Grid3 pacman cache design.*
7/11	Start development work for infrastructure needs for multi-VO application grid.	*Grid3 cache started.* *Waiting for VOX project taking stock*
7/18	Agree on GIIS structure, common schema and core information providers. Identify reliable hosts and support for: top level GIIS server, grid-wide , monitoring displays,	*Grid3 GIIS port and ip decided.* *Document started*
7/25	Review documents of GIIS structure, schema and information providers.	**Document finished.**
7/31	A simple Grid3 deployed on Core sites, with two application demonstrators running. Define monitoring requirements.	**Done during Grid3 week, August 18.**
8/1	Decide on VO management infrastructure and software to be used. Start packaging and testing of VO management software on Grid3 sites.	**Deployment of VOMS before the VDT 1.1.11**

	Discuss and make preliminary decisionon policy support. Decide on and document RLS/RLI infrastructure.	**upgrade of October 3.**
8/8	Define Storage Elements, software to be installed, support for SE in VDT, testing needs.	
8/15		
8/18-22	Working meeting for integration of first phase of Grid3 sites and first ~2 demonstrators including installation and configuration.	**Done.**
8/22	Complete development work for infrastructure needs for multi-VO application grid. Start development work for monitoring and display programs and portals.	**Grid3 extensions and IP's for this purpose; complete on Oct 3.**
9/5	Delivery of netlogger CS demonstrator	**Week of October 3.**
9/12	Demonstration of grid wide monitoring and performance display.	**Ganglia and MonaLisa displays**
9/19	Start "continuous" operation of Grid3 with ~2 applications. Initiate dissemination material.	**USCMS, USATLAS, SDSS submitted applications during the development period.**
9/26	Integration of additional application demonstrator. Dry run of operational procedures in response to problems. Decide attendance at SC2003 and where the demonstrations will be run.	**Applications: 6 applications from 4 VO's. 4 CS applications.**
10/3	Add sites to the grid. Review of documentation available and needed.	**Still adding sites at 10/15 (24 sites)**
10/10	Integration of additional application demonstrator. Review of interoperability with LCG-1 and addition of LCG-1 sites as resource providers.	**Interoperability reviewed on 10/10**
10/15-10/17	Working meeting for integration and operations. Cut off date for new application demonstrators. Transition to concentrating on operations, throughput parameters and metrics.	**Done, on 10/13-10/14.**

	Review of dissemination material available and needed.	
10/24	Test of integration of Grid3 and LCG-1 sites.	**VDT Client and LCG-1 gatekeeper okay**
10/31	Final configuration of sites participating in Grid3.	**Delayed till 11/7**
11/8	Final freeze of all software.	
11/15-11/19	SC2003 demonstrations	
11/25	Addition of demonstrators available since 10/17.	
12/16 (Δ12/12)	Face to face meeting with steering and management to take stock and understand status of performance metrics. Decide whether/how long to sustain the project and identify additional demonstrators and sites.	

8 Effort

An estimate needs to be made of the required effort for the project. Specific personnel need to be identified. In each case the effort must be accounted for in the corresponding project management plans of the collaborating organizations. We indicate where we think the effort might be needed.

Role	Effort	Total	Source	Names
Operational support at each site, including security, installation/config, site certification (testing) Site preparation (cluster management tools-rocks, …) (>10 sites)	.25 FTE/site	2.5	IVDGL Tier 2 system administrator (5 sites) Tier 1 user facility support (Fermilab, Brookhaven)	ATLAS: Xin Zhao/UC **Matt Allen/IU** Dantong Yu/BNL Ed May/ANL Iwona S./LBNL **P. McGuigan/UTA)** **S. Youssef (BU)** **S. McKee (UM)** CMS: Jorge Rodriguez/UF Terrance Martin/UCSD Shaowen Wang (UIowa)? Suresh Singh/CIT **Ian Fisk (Fermilab)** SDSS: **George Fekete /JHU** Vijay /Fnal

Role	Effort	Total	Source	Names
				LIGO: Brian Moe/UWM Bruce Allen (facility manager) Mike Foster (technical)/PSU & Sam Finn (facility manager) **BTEV** **Bobby Brown/VU** DOESSG: D. Olson (LBNL) **U Wisconsin:** **Dan Bradley/UWM-Phys** **Erik Paulson** ANL-Jazz: Scott Gose
Middleware support team • VDT • MDS • RLS • Packaging/Config	1.5 FTEs	1.5	IVDGL VDT .5 FTE (Wisc) Pacman support (iVDGL, BU) .5 FTE Globus (Chicago, ANL) IVDGL/PPDG .5 FTE	Carey Kireyev/UW (VDT support) Scott Gose/UC (Globus, iVDGL) JensVockler Chimera support) Gaurang (Pegasus)
System diagnostics and troubleshooting • End-to-end systems (applications) • Instrumentation • MDS performance • Grid3 Site Verification	2 FTEs	2	PPDG Condor team 1 FTE/Wisc Globus ISI iVDGL 1 FTE/ISI	Peter Couvares/UW (Condor support) Alan Desmet/UW (Condor support) Nosa Olmo/ISI (RLS and MDS support) **Craig Prescott (UF)**
Coordination, operations, and user model	~1 FTE	1	usatlas, uscms, ivdgl, ppdg	All project participants.
Deploy, use, feedback: Grid-wide, Multi-Application monitoring + operations Applications and operations specific monitoring Eg. Monalisa, Netlogger	2 FTEs	2	1 iVDGL operations center (IU/iGOC) .5 US ATLAS .5 US CMS ? DOE SG?	Leigh Grundhoeffer(iVDGL)/IU John Hicks (IU) John Weigand (USCMS) Marco Mambelli/UC (0.5)

Role	Effort	Total	Source	Names
				Monalisa contact: Yang Xia , Iosif Legrand (technical contact no specified FTE estimate) NetLogger contact: Brian Tierney
User interface (web portal), reporting, and dissimenation mechanisms	1 FTE	1	UC 0.5 FTE ? 0.5 FTE	Marco Mambelli, David Bury/UC **Andrew Zahn (UF)**
Per application group: preparation, deployment and diagnosis of application software: ATLAS GCE, CMS-PCP software (packaging, client and server side)	1 FTE	2.5	US CMS PPDG 1 FTE iVDGL/UF 0. FTE US ATLAS PPDG/ANL, BNL 1 FTE SDSS LIGO BTEV CS demonstrators	Nickolai Kouropatkine (CMS/Fermilab)/FNAL Bockjoo Kim (UF) Xin Zhao, Yuri Smirnov/UC Jerry Gieraltowski /ANL **Dan Engh/UC (BteV)** Mike Wilde (UC), Brian Tierney (DOESG, LBNL) **Transdemo (M. Enrst) GridFTP:Scott Gose** CS-LIGO (Gaurang Mehta)
Project coordination, international collaboration and liaison to LCG, Teragrid	1 FTE	1	Usatlas, uscms, ivdgl, ppdg	Taskforce
Outreach and dissemination Poster, demo-booth, brochures, documentation/tutorial	1.5 FTE	1.5	Fermilab CD, Indiana	.5 Anne Heavey (FNAL) poster/documentation **Eileen Berman/FNAL** **Rich Baker/BNL** **Fred Luehring (Rob Quick)/IU**
Total		15 FTEs		

9 Dissemination

The demonstrator will have a central web portal where information is aggregated and logged, and where people can visit and view the activities, etc. There should be sufficient manpower and endorsement to send monthly news updates to the funding agencies of throughput achieved in terms of various activities. For example:

- Number of jobs/workflows/tasks (various representations)

- Simulated or analyzed events delivered to the experiment
- CPU used
- Faults per hour, faults characterized
- Data transferred
- Operational and middleware support effort
    - Application support effort
    - Middleware extensions
- Value for:
    - LHC deliverables
    - CS milestones achieved
- User feedback, numbers, comments

This work corresponds to the 'common reporting tool, event logging framework' description listed in the monitoring section above. The project registry proposed by the WorldGrid document would be relevant here.

10 Issues

There are a few significant technical issues that we have identified:

- Particular attention must be paid to packaging, distribution and support of multiple versions of the Grid3 middleware and environment.
- The CERN and Globus versions of RLS must be reconciled on the timescale of Grid3.

11 Next Steps

What happens next after the construction and demonstrator phase?

Possible text: During the last phase of the project we will assess and document the achievements in terms of the project goals and deliverables as part of the regular reporting of the collaborating projects. The project management team will be responsible for working with the steering and management groups to define and establish any follow-on projects and/or extension of the current project. It will be part of the milestone to deliver the next phase, well before the end of the Grid3 project.

12 Appendix

12.1 iVDGL Architecture

This picture will need to be revised for the project, but the basic services and interfaces can be identified with today's components. To be discussed.

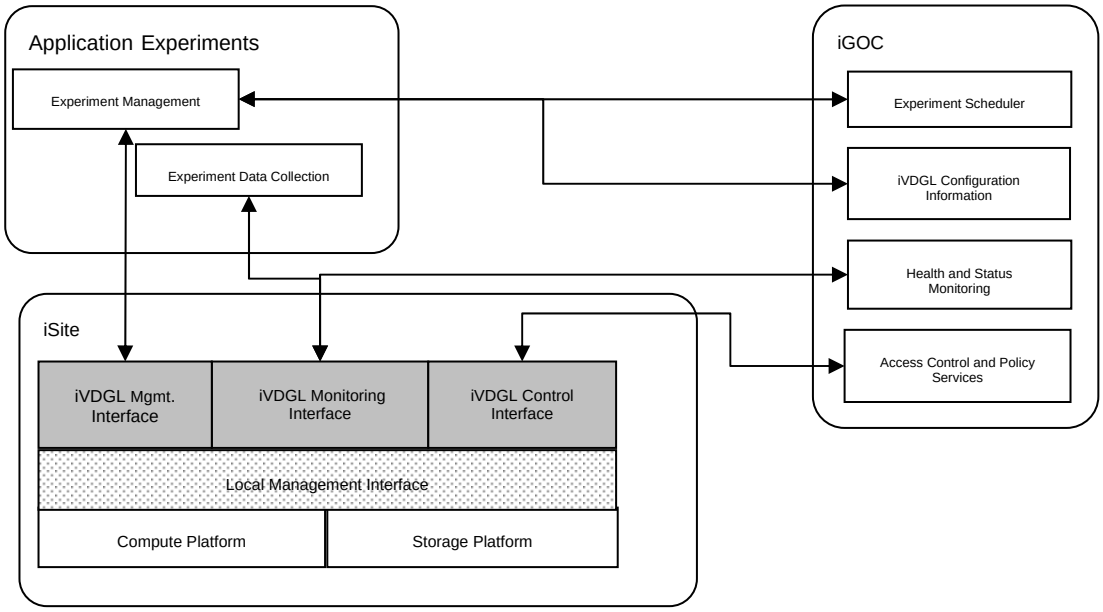

Figure 1. iVDGL architecture. Local sites support standard iVDGL services that enable remote resource management, monitoring, and control. The iGOC monitors the state of the iVDGL, providing a point of access for information about the Lab, as well as allowing for configuration for particular experimental activities. The application experiments interact with iVDGL first by arranging to conduct an experiment with the iGOC, and then by managing the needed resources directly via iVDGL services. (This operating model will need revisiting.)

12.2 iVDGL Year 2 proposal: Demonstrate Value in Application Experiments

Deployment: Expand iVDGL to 18 sites, 13 in the U.S. and 5 in Europe, all running second-generation software including new URCs. Establish high-speed network connectivity within U.S. and across the Atlantic using the SURFNET and DATATAG research links (OC-48). Establish iGOC trouble ticket system. Develop iVDGL monitoring systems which handle collection and presentation of grid telemetry data. Conduct regression tests of all components. Start logging of iVDGL trouble tickets for human factors studies. Evaluate reliability and scalability of iVDGL monitoring. Deploy a set of infrastructure support services (bandwidth management services) being developed by other organizations such as Internet2 e.g. QoS, DiffServ, and Multicast.

Experiments/Applications: First large-scale application runs across 15 sites and Gb/s networks. *ATLAS*: Continued integration of grid services within the Athena framework. Data Challenge 2: 10% complexity scale involving 5-10 iVDGL sites. Performance and functionality tests will be used in ATLAS computing technical design report. Validate the LHC computing model64. *CMS*: 5% complexity data challenge using 10 iVDGL sites and approximately 50 users. Use DTF to explore use of NRC class facilities. Completion of a CMS production cycle where half the efforts are completed using Grid tools, including: first pre-production set tools for task monitoring, optimal task assignment to sites, in addition to the tools used in the previous year. *LIGO*: Port LIGO scientific algorithm code to the iVDGL for pulsar searches over large portions of the sky; port code to perform lower-mass in-spiraling binary coalescence searches; work with EU partners to replicate this capability on the EU grid; *SDSS/NVO*: Tests of code migration between iVDGL sites. Grid-enable gravitational lensing application code. Integrate first SDSS data release. Run

prototype power spectrum calculations in production mode, on medium scale. *Other Apps*: Work with PACIs and others to define experiments. *Middleware/Tools*: Deploy instrumentation archives. Start collecting and analysis usage patterns on iVDGL services, scalability studies.

12.3 Task Force

From the Steering Committee meeting of June 4-5, 2003:

- Mandate: Propose the project work plan (including schedule) for Grid3, negotiate the effort, and propose a project structure. This should include inventory and procurement plans for resources which would be used under the assumption of a successful Grid3 environment.
- Time frame to deliver this plan to (June 27, 2003):
  - iVDGL, GriPhyN and PPDG Steering Committees
  - USATLAS, USCMS, LSC, SDSS S&C project
- Membership:
  - iVDGL Coordinator – Rob (co-chair)
  - PPDG Coordinator – Ruth (co-chair)
  - VDT Team representative – Peter C
  - GriPhyN Coordinator and CS demonstrator representative– Mike Wilde
  - US-CMS representative – Ian Fisk
  - US-ATLAS representative – Rich Baker
  - LSC representative – Scott Koranda
  - SDSS representative – Jim Annis
  - iVDGL Operations – Leigh G
- Request for comment period the week of June 23

13 References

Collected references:

- References for VO management:

  http://zuni.cs.vt.edu/grid-security/globus-authz-mod.html

  http://www.dutchgrid.nl/DataGrid/wp4/lcas/edg-lcas-1.1/node1.html

  http://www.uscms.org/s&c/VO/doc/doc.html

  http://www.uscms.org/sandc/reviews/doe-nsf/2003-01/docs/DC04_054.pdf

  http://hep-project-grid-scg.web.cern.ch/hep-project-grid-scg/DataGrid-07-D7.6-0112-2-0-SecurityDesign.pdf
- iVDGL proposal: http://www.phys.ufl.edu/~avery/ivdgl/itr2001/proposal_all.pdf
- The WorldGrid design document, April 2003:

  http://physics.bu.edu/~youssef/wg/Worldgrid-Design-1.0.doc
- Memo on WorldGrid and iVDGL 3 to iVDGL steering committee:

  http://www.ivdgl.org/mail_archive/steering/msg00083.html
- ATLAS data challenge plans:

- February 2002: http://atlas.web.cern.ch/Atlas/GROUPS/SOFTWARE/DC/DC1/DC1-Report-V1.0-211002.pdf
- May 2003-05-29: http://agenda.cern.ch/fullAgenda.php?ida=a031389

- CMS DC04 plans, November 2002: http://www.uscms.org/sandc/reviews/doe-nsf/2003-01/docs/DC04_054.pdf
- Open Science Consortium, Proposal Draft, May 17, 2003: http://www.uscms.org/oscc/
- Open Science Grid Presentation to Fermilab Computing Division, May 14, 2003 http://cdinternal.fnal.gov/Org2003/BriefingMtg/2003Briefings/2003-05-CDbriefing.pdf