Grid2003 Project Lessons

Version	Date	Author	Status
1	11/28/03	RP, PFC,AR	Initial draft, long list of metrics, problems, recommendations.
2	12/3/03	RWG, MM	Added to the above, with specific iVDGL and monitoring material.
3	12/4/04	XZ	Added ATLAS info.
4	12/8/03	RWG	Added Ruth's PPDG statement, lessons from Scott Gose, CMS SRM demonstrator, LCG-Grid3 interoperability
5	12/12/03	RWG	Added installation observations from F. Luehring, Ed May
6	12/16/03	RWG	Added comments by Ian Fisk (CMS), Jason Lee (Netlogger). The whole document needs a stylistic edit.
7	12/29/03	RWG	Added comments by P. McGuigan.
8	1/1/04	RP	Prepare for U.S. CMS SCOP review, summarize input from steering meeting

Contents

1 Overview

1.1 Purpose of this Document

This document presents results from the Grid2003 Project which resulted in deployment of the "Grid3" grid environment, execution of application demonstrators, and collection and archival of performance metrics. It includes an analysis of the achievements, lessons learned and recommendations for future work. The intended audience of this document is: Stakeholders: Steering committees of GriPhyN, iVDGL and PPDG, the U.S. ATLAS and U.S. CMS software and computing management; Facility management – the line management of the site administrators for Grid3 sites; Application management – the line management of the application administrators for those applications running on Grid3.

1.2 Goals of the Project

The Grid2003 project was defined and planned by Stakeholder representatives in an effort to align iVDGL project goals with the Software and Computing projects of the LHC experiments. The planning process converged during the iVDGL Steering Committee[i] at Argonne Laboratory, June 4-5, 2003, with a set of agreed to principles, that the Project must:

- 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.

Planning details were iteratively defined, and are available in the iVDGL document server, c.f. Plan V21. The goals of the project included meeting a set of performance targets using metrics listed in the planning document. The central project milestone can be summarized as delivery of a shared, multi-VO, multi-application grid laboratory in which performance targets were pursued through deployment and execution of application demonstrations during the period before, during, and following the SC2003 conference in Phoenix (November 16-19). The Project was organized as a broad, evolving team including the application groups, site administrators, middleware and core service providers, and operations. The Project was able to call on additional effort through the Stakeholder organizations.

The design and configuration of the Grid were driven by the requirements of the applications. The Project included those responsible for installing and running the applications, the system managers responsible for each of the processing and storage sites (including U.S. LHC Tier1 Centers, the iVDGL funded prototype Tier2 Centers, resources from physics departments and leveraged facilities from large scale computing centers) as well as the groups responsible for delivery and support of the grid system services and operations. The overall approach of the project was "end-to-end" in terms of giving equal attention to the application, organization, and site infrastructure and system services needed to achieve science applications running on a shared grid.

The applications running on Grid3 included official releases corresponding to production environments that the experiments will use in run in production and analysis over the next year. Applications from the computer science research groups (GridFTP, Exerciser, Netlogger) were used to explore the performance of different aspects of the Grid.

The project plan included basic principles of the project, which contributed to making life simpler and more flexible. In particular the decisions to: have dynamic installation of applications; not presume the installation and configuration of the "worker" processing nodes; use existing facilities and batch systems without reinstallation of the software; all contributed to the success of the project.

The active period of the project, where the people involved were expected to contribute and be responsive to the needs of the project, was a period of 5 months from July through November 2003. Subsequent to this Grid3 remains, with many applications running, but there are reduced expectations as to response time to problems and the attention of the members of the team. The collaborative organization of the project allowed us to address problems as they arose and focus our efforts in response to unanticipated issues. The team took decisions and was flexible enough to accept additional sites and applications into the project as it evolved. Identifying people with coordination roles has helped the project to scale in size and complexity. These roles were filled by responsibles from their respective projects, including: Sites (iVDGL operations team), Applications (iVDGL applications coordinator, with liaisons from each VO's Software and Computing project), Monitoring (Grid Telemetry), Operations (iVDGL operations) and Troubleshooting (VDT). The iVDGL and PPDG project coordinators, who also represent U.S. ATLAs and U.S CMS respectively, coordinated the Project.

2 Achievements

2.1 The "Next Phase" of the iVDGL Laboratory

As stated, the Grid2003 Project was organized to meet several strategic project goals, including building the "next phase" of the iVDGL Laboratory, according to the stated goals and mandate of NSF funded iVDGL ITR Project. The iVDGL Project previously had two deployments (both in 2002): a small testbed consisting of VDT deployed on a small number of U.S. sites (iVDGL-1), followed by an a second, joint deployment with the EU DataTAG project (iVDGL-2) which coincided with the IST 2002 and SC2002 conferences (the "WorldGrid" Project). Grid3 (originally proposed iVDGL-3) is the third phase of the iVDGL Laboratory.

The iVDGL mandate includes, as stated in proposal:

"We propose to establish and utilize an international Virtual-Data Grid Laboratory (iVDGL) of unprecedented scale and scope, comprising heterogeneous computing and storage resources in the U.S., Europe—and ultimately other regions—linked by high-speed networks, and operated as a single system for the purposes of interdisciplinary experimentation in Grid-enabled[ii,iii] *data-intensive scientific computing*[iv,v]*… The laboratory itself will be created by deploying a carefully crafted data grid technology base across an international set of sites, each of which provides substantial computing and storage capability accessible via iVDGL software. The 20+ sites, of varying sizes, will include U.S. sites put in place specifically for the laboratory; sites contributed by EU, Japanese, Australian, and potentially other international collaborators; existing facilities that are owned and managed by the scientific collaborations; and facilities placed at outreach institutions… An international Grid Operations Center (iGOC) will provide the essential management and coordination elements required to ensure overall functionality and to reduce operational overhead on resource centers. The system represents an order-of-magnitude increase in size and sophistication relative to previous infrastructures of this kind."*

The Grid2003 Project deployed, integrated and operated Grid3 with ~25 operational processing sites comprising at peak ~2800 CPUs for more than 3 weeks up to, during and after the SC2003 conference on November 16, 2003. Other iVDGL proposal themes in which progress was made:

- Multiple VO grid: six different virtual organizations participated with 10 applications deployed and successfully run. All applications were able to run on sites that were not owned by the organization whose application it was. The applications were all able to run on non-dedicated resources.

- Multi-disciplinary grid: during the project two new applications, one from biology and the other from chemical informatics, were run across Grid3. The fact that these could be installed and run on a Grid infrastructure designed and installed for Particle and Astrophysics Experiments gives us added confidence that this infrastructure is general and can be adapted to other applications as needed.
- Use of shared resources: many of the resources brought into the Grid3 environment were leveraged facilities in use by other VO's. Examples include successful incorporation sites
- Grid Operations and establishment of the iGOC. Resources from the Indiana University-based Abilene NOC were leveraged to provide a number of operations services, including: VOMS services for iVDGL VO participants (CS, Biology, Chemistry), the MonALISA Grid3 database which served double duty for online resource display and archival storage for the Metrics Data Viewer (MDViewer) used for analysis of Grid3 metrics, the top level GIIS information service, development and support of the iVDGL:Grid3 Pacman cache, coordination, development and hosting of site status scripts and displays, creation/support of Ganglia Pacman caches and hosting of top-level Ganglia collector and web server.
- Dynamic resource allocation: the University of Buffalo CCR was able to configure their local schedulers bring resources into and out of Grid3 nightly according to local policies, satisfying local requirements and (external) Grid3 users.
- International connectivity: though one site was located abroad (Kyunpook National University, Korea), international operations were not a primary focus in contrast to last year's WorldGrid VDT-EDG interoperability demonstration project which focused on TransAtlantic Grids.
- VDT installation and configuration. Improvements included enhanced support for post-install configurations at sites with native Condor installations. Pacman3 development was concurrent to most of the project and was not used for deployment. However, initial tests of Pacman3 with iVDGL:Grid3 have demonstrated backwards compatibility of the new tool.
- VDT testing and robustification. The Troubleshooting team, lead by the VDT group, oversaw a number of VDT improvements and patches in response to bugs uncovered by site administrators and application users and developers. These included, most importantly, patches required for job managers and provisions for the MDS 2.4 upgrade.

2.2 LHC Production and Analysis at Scale in a Shared Grid Environment

As stated, the Grid2003 Project was organized to meet several strategic project goals, including "Provide the infrastructure and services needed to demonstrate LHC production and analysis applications running at scale in a common grid environment."

Figure 1 shows the integrated and differential Grid3 usage (in CPU-days, by VO) over the course of the main demonstration period. Both U.S. ATLAS and U.S. CMS were able to run production systems at scale during this period. Note that the experiments continue to exercise production on Grid3 (right plot) with an average of 700 CPUs in daily use.

The Grid3 environment has provided an opportunity for U.S. ATLAS and U.S. CMS to work together and deploy a joint infrastructure for their production processing needs. As a result of the success of the project, both experiments plan to continue to operate production facilities and applications in this multi-organization environment for the 2004 Data Challenges.

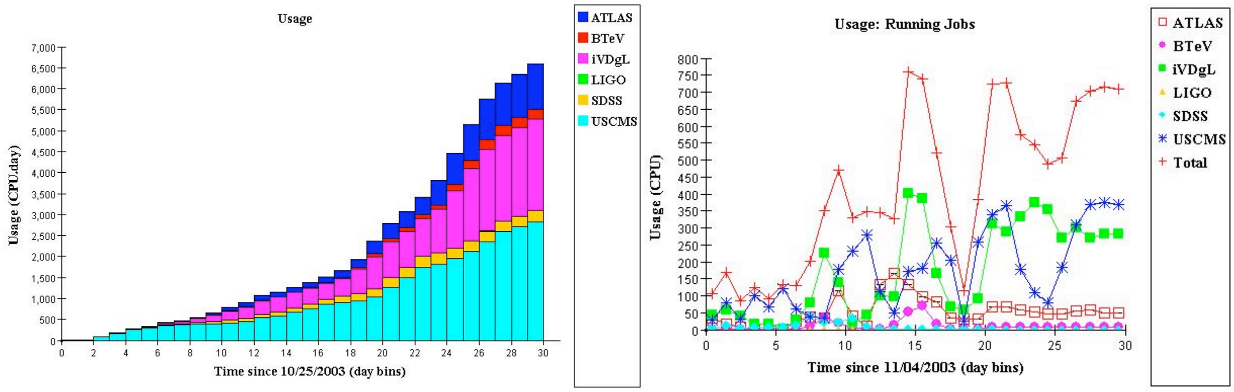

Figure 1 Integrated and differential CPU usage during Grid2003, by Virtual Organization.

2.2.1 ATLAS PreDC2 Demonstrator

For ATLAS, we installed the ATLAS grid-enabled application package GCE-Server, which is based on GriPhyN Virtual Data System (VDS), on 22 Grid3 sites. Automatic (user-level) installation tools based on Pacman were developed, and effectively used the Grid3 GLUE schema extensions for application installation attributes. Client hosts (GCE-Client) were installed outside Grid3 for job submission. More than 5000 jobs (Geant3 based simulation followed by reconstruction) were processed on about 18 large computing sites, with total data I/O of about 1.1 TB (the results are summarized in Figure 2). The data processed were Higgs and Top samples, all registered in RLS. Using the DIAL distributed analysis tool, clear four electron signals were reconstructed from the produced Higgs datasets. The produced datasets were stored at BNL as part of the grid jobs, and continue to be analyzed by DIAL developers and the SUSY physics working group. A parallel activity was to submit jobs to the LCG-1 production testbed using GCE (Chimera). This work is still on-going, and several concrete steps have been taken to incorporate differences between the LCG and typical Grid3 site configurations.

However, during the demo runs, we also observed a high failure rate, about 30%, of the submitted jobs, especially for the long jobs that run more than 4 or 5 hours. Failures were defined as errors in any step of the job sequence which prevent perfect completion (pre-stage, job execution producing the output file, post-stage to the final storage element at BNL, and registration to RLS) so that most were partial failures. Approximately 90% of the failures were caused by problems at the computing site: disk filling errors, gatekeeper overloading or network interruptions. Our software did not handle the nightly roll over the worker nodes at UB gracefully, and so jobs still running were crashed and had to be re-processed. Another reason for the job failures came the middleware itself (a separate document detailing the specific errors is in preparation); many glitches were detected when we ran large numbers of production jobs over sites with differing batch queue systems.

In the future, to reduce the failure rate more stable computing elements need to be provided and maintained by site administrators, along with the iGOC, which provides monitoring and troubleshooting services for the Grid. Site resource configuration is another issue to study; one example is to separate the GridFTP server from the gatekeeper, so that the gatekeeper doesn't get overloaded. On the other hand, our job submission mechanism also needs to be improved. Site status and site policy information has to be taken into account, so that jobs only go to sites that are functioning and match the job requirements. After jobs are finished cleanup is important, so that we will not quickly fill up the disk space of the computing element. Another issue is to understand the different batch queue systems better. Sometimes we mis-understood the returned message from the remote site, which prevented us from interpreting the job status correctly. As grid middleware evolves continuously, we probably can expect more and more job scheduler particularities be hidden behind the unique grid interface, so that application users don't need to worry about all the details of the different batch queue systems.

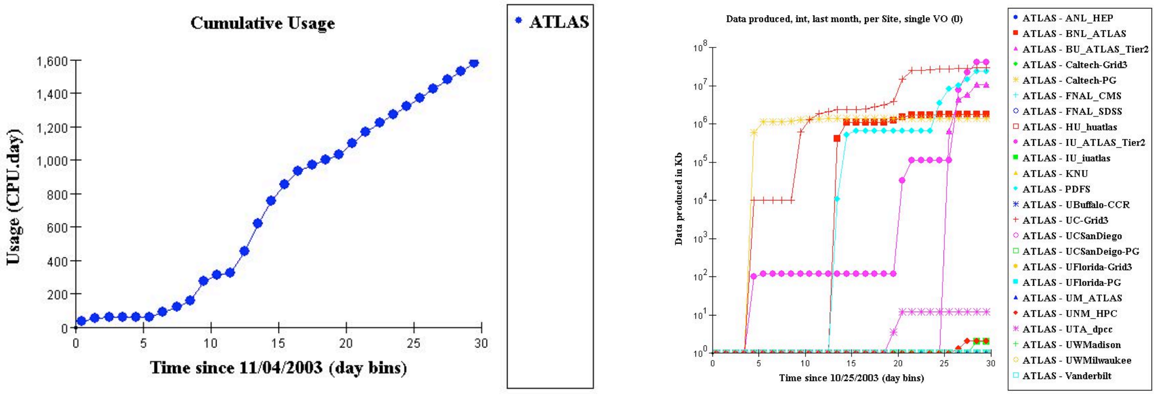

Figure 2 ATLAS production on Grid3 (11/04/03) to 12/04/03: CPU days and data produced, by site.

2.2.2 USCMS MOP Production

US-CMS used Grid2003 resources to produce simulated events for the upcoming CMS data challenge. US-CMS ran a GEANT3 based, statically linked FORTRAN application called CMSIM and a GEANT4 based, dynamically linked, C++ application called OSCAR. The timing of the Grid2003 assembly was not ideal for US-CMS because it almost completely overlapped with the ramp up of the data challenge production. This placed a significant load of the participants as well as limiting the resources we could devote to Grid2003. In the beginning only the US-CMS test resources were devoted to Grid2003 deployment, while the much larger production clusters continued to run a less general but more proven distributed system deployed for US-CMS. The production resources were first allocated for Grid2003 during the first week of October. Though not planned, this roughly corresponded to the end of the official CMSIM production. The OSCAR production was not ready to proceed until the end of SC2003. Subsequently, the simulation US-CMS performed on Grid2003 resources until SC2003 was primarily test samples designed to demonstrate the scaling and reliability of the system, but not saved for physics analysis. Since SC2003 US-CMS has successfully used Grid2003 resources on X (input from Craig) sites to simulate more than 1 million GEANT4 full detector simulation events for the CMS physics groups. Computing usage since the middle of November is shown in Figure 3. In general, the efficiency to run on Grid2003 resources is roughly as high as the original US-CMS production grid, once the sites have been fully validated. The official OSCAR production jobs are extremely long and not all sites have been able to accommodate running them.

US-CMS successfully dynamically loaded the application code to the areas specified in the GLUE extended schema published in MDS for most sites. The problems encountered were most frequently local configuration issues, often permission problems for the areas specified. It is clear that the sites administrators need better suites of tools to validate site configurations. The current method in which a site is validated by an external system and the results are published on the web is important, but not sufficient. The local site administrator needs to be able to check the site status, determine configuration problems, and iterate. The site services and configuration need also to be checked at a deeper and finer grained level.

US-CMS found the MDS installation was sufficiently unreliable until the upgrade to MDS 2.4, shortly before SC2003, that local site information could never be routinely dynamically queried from MDS. Instead information was taken from MDS or the Grid2003 web page and entered into static site configuration files. The primary functionality is the same, but this robs the local site administrator of more dynamic control of the facility. The US-CMS goal had been to allow the site administrator the ability to change the local site configuration to suit his/her needs and have the submitting jobs recognizes the changes immediately from MDS. This would allow administrators to respond to failed and over subscribed resources more easily by redirecting the applications through information provider changes. We still think this is desirable and hope the upgrade to MDS allows this functionality.

The rate of CMSIM to successfully complete is approximately 70%, which is consistent with the US-ATLAS estimates. We have yet to do a calculation of the percentage of OSCAR jobs which complete. One aspect we have noticed is that jobs often failed with site configuration problems, or failed in groups from site service failures. We saw fewer random job losses and more frequently a disk would fill up or a service would fail and all the jobs submitted to a site would die. The service level monitoring needs to be improved and some of the services need probably to be replaced. For example, the resource reservation available through the SRM based storage element service would have prevented a variety of storage related service failures.

US-CMS also ran into some non-technical issues in implementing and using Grid2003. Primarily this is an issue of US-CMS communication and priorities. In general we must make more realistic effort estimates at the beginning of the project and during the planning stages.

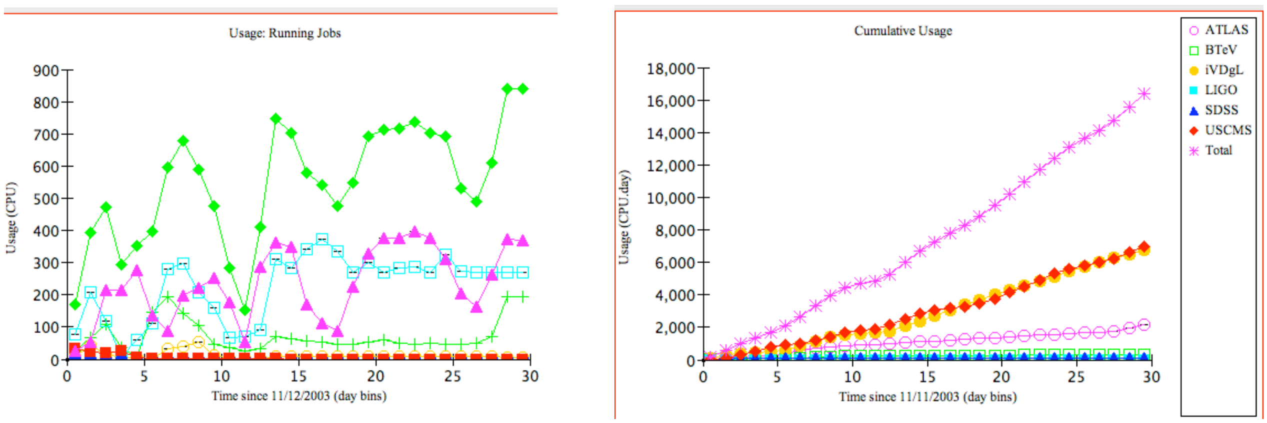

Figure 3: CMS daily and cumulative use of Grid2003 since the beginning of SC2003.

2.3 LCG-Grid3 Interoperability

Interoperability of Grid3 with the LCG is an ongoing commitment and goal. Attention was paid throughout the project to technical and programmatic issues that would potentially impact this goal. It was accepted that there was insufficient effort in Grid3 to actually address all these issues. The Grid3 User Registration policy was taken directly from the LCG registration policy. Specific interoperability demonstrators were discussed and agreed to at a meeting between the LCG Grid Project Managers (Les Robertson, Ian Bird), the U.S. ATLAS and U.S CMS computing management and the Grid2003 project leaders at a meeting in October 2003. These demonstrations were shown at the RSIS forum at CERN in December 2003 and further work is planned.

2.3.1 Information Schema Development

The GLUE schema attribute set for variables was extended to address needs of the applications in Grid2003. Specifically for site specific information (e.g. path names for application, data and temporary disk areas) and Virtual Organization and application information related to specific application demonstrators (e.g. for Chimera Virtual Data planning). During a meeting of the Joint Technical Board it was agreed that a new version of the Glue Schema would be released to include a merge of the Grid3 and LCG extensions and provide for interoperability between the environments. This work is underway.

2.3.2 ATLAS GCE Production on LCG-1 sites

A parallel activity to the ATLAS production done on Grid3 was to submit jobs to the LCG-1 production testbed using GCE (based on the GriPhyN VDS system). The submit sites were chosen to be Grid3 sites. The LCG-1 sites were identified to be one site at CERN and one site at Brookhaven. A separate report summarizes our observations based on initial testing efforts. This work is still on-going, and several concrete steps have been taken to incorporate differences between the LCG and typical Grid3 site configurations. As of 04-dec-2003, twenty-three (23) LCG-1 sites could be identified as supporting the ATLAS VO. Of these twenty-three, four (4) sites were defined as being part of LCGWEST with the remaining nineteen (19) sites being part of LCGEAST.

2.3.3 SRM Storage Service Demonstrator by U.S. CMS

This Storage Element (SE) has several components: a Disk Management system, dCache, developed as a joint project with DESY and FNAL and Storage Resource Manager (SRM) as the common interface. dCache allows many storage pools to be grouped together into a storage system. The pools can be everything from high performance RAID devices to individual disks on worker nodes. dCache handles load balancing by replicating files between pools if they are frequently requested. It also transparently handles the loss of a pool, provided the data is available on another pool or a MSS. dCache was designed as a front end to tape systems, which is the implementation at FNAL. The SRM interface is becoming the standard storage interface protocol. The implementation deployed in the U.S. handles advanced functionality of proxy driven transfers, resource reservation, and transfer queuing. The functionality offered by this storage element solves many problems, which we've seen with the CMS application. Applications can reserve space for a period of time, ensuring that space will be available as the application runs. Worker nodes inside a NAT or with no external network connection at all can trigger data be replicated to the storage element, and then access the data through the LAN. Moving the storage element functionality off the compute element also reduces the load and improves the reliability of the node.

This is an active development project and we do not see a wide distribution through Pacman in the Grid2003 cache in the next several months, but sites interested in the functionality should contact Michael Ernst (ernst@fnal.gov) and Timur Perelmutov (timur@fnal.gov). CMS has demonstrated interoperability between CERN and FNAL, through the SRM interface data was transferred from tape at CERN to tape at FNAL. The storage element functionality is expected to be deployed in the LCG-2 release. We expect at next week's CERN demonstrations to show data moving between storage elements.

2.4 Computer Science Demonstrators

We summarize results from a number of Computer Science demonstrators deployed during the project.

2.4.1 GridFTP Data Transfer Demonstrator

The goal of demonstrating 2 TB data transfer across Grid3 per day was realistic and was achieved (see the section on metrics). The Grid and the sites proved robust to long running data transfer demonstrations. Issues of account privileges, ports and firewalls caused the main problems in deployment and configuration. For the next GridFTP goal, it would be good to add a planning component, which takes account of the site and network capabilities. For example, some sites have faster connections than others, can those sites contribute more to the 2 TB/day goal or does each site need to contribute the same amount? What about sites that are down? Etc One important lesson learned was that you couldn't depend on all the sites in Grid3 being operational at any given time. That is only 80% of the sites may be working when a demo is run. This needs to be incorporated into the planning of any demos that are theorizing on how to go about reaching a metric. The client GUI user interface developed for this demonstration proved its worth in providing an easy to use interface for configuring and tracking the data transfers.

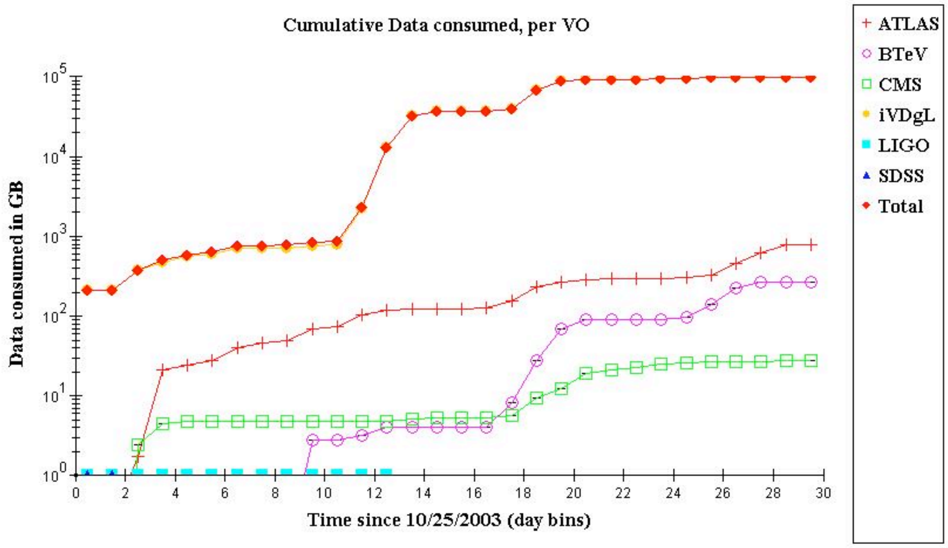

Figure 3 **Data consumed by Grid3 sites, by Virtual Organization, showing nearly 100 TB transferred during the main project period. The GridFTP demo accounted for the vast majority of data transfers.**

2.4.2 NetLogger Instrumented GridFTP

The Netlogger instrumented GridFTP was used in specific demonstrations and the information stored and published at LBL (http://netlogger.lbl.gov:8080/grid3.rpy). The demonstrations worked well, and the information is available for analysis. The fact that Grid3 had solidified on using one version of RedHat linux (and stuck by its decision) was a huge contributing factor in the fact that this project went so well. Once a program was working on one set of nodes, you could be fairly confident that it would work at all the other sites. Using Pacman for the software distribution of all the software (and using versions that tied all the various components together) was a big win for this project. Simply providing a 'standard' environment in this way helped tremendously.

Installation was problematic. It would have been useful to have a login shell at the different sites to debug problems (like the version of python, the internet connectivity of the nodes, the hostname of the node as returned by uname, etc.). In the end a job was submitted that installed the application. This worked fine, but getting to a point where there was a working job took quite a while.

2.5 A Common Grid Environment for LIGO and SDSS Applications

The stated goal for the Grid2003 project was to provide a common grid environment for LIGO and SDSS applications. Grid3 is a useful grid environment for LIGO applications, in particular for conducting a blind search for continuous gravitational waves from rotating neutron stars. Continuous gravitational waves, nearly sinusoidal signals in the rest frame of he source, are expected from rotating neutron stars. Due to selection effects of our observations we know of only a small fraction of the existing population of these sources. The chances are that the source nearest to us is actually known--hence, the interest in performing blind searches, scanning the regions of the sky where we have no a priori information of the presence of a source. The search codes to perform wide-area, wide-frequency blind searches were designed so that the computation load can be distributed among different and very loosely coupled computers -- the results of the calculations performed on one machine are completely independent of what is happening on any other machine. The second science run of the LIGO detectors took place over a time of 8 weeks during March-

April 2003. The data from the three LIGO detectors is being analyzed to search for continuous gravitational wave signals coming from the region in the direction of the Galactic Center and the galactic core.

Some of the searches already conducted have been successfully performed using Grid3 resources. We did, however, realize that like some of the other Grid3 applications, we were limited by the load due to the Globus jobmanagers on the number of jobs that could be scheduled into a Grid3 site. We have determined that the Condor GridMonitor substantially helps relieve the stress on the machine running the Globus gatekeeper and jobmanagers, and we would like to see it properly installed and configured at all Grid3 sites. This is especially important since performing any searches/jobs at a site requires moving at a minimum 20 gigabytes of processed data files onto the site. In order to recover the cost of moving these files we need to run as many jobs as quickly and efficiently as possible against the data we stage in. At this time we are re-factoring the scientific analysis pipeline and making adjustments to its representation as a Condor DAGman DAG in order to improve I/O performance. As soon as we have validated our new codes we plan to immediately begin using Grid3+ resources to continue the blind search.

2.6 Partnership with PPDG

The Grid2003 goals and methods are aligned with the PPDG focus on end-to-end applications, system hardening and deployment. In Grid3 PPDG participates in the application administration and deployment, troubleshooting, and system deployment. The participation of the DOE laboratories that are part of the Pilot Collaboratory is enabling the lessons learned from Grid2003 to be applied across to other particle physics experiments who are already in data taking mode and whose needs for production and legacy systems mean that they can only move more slowly towards a common system.

3 Milestones and Metrics

The Grid2003 project met the final milestones as documented in the plan in terms of demonstrating capabilities and applications for SC2003.

The Grid2003 project additionally met the metrics from the planning document as listed below. The "status" numbers fluctuate over the course of several weeks around SC. A webpage devoted to collection and analysis of Grid3 metrics has been setup (ref: http://grid.uchicago.edu/metrics).

1. **Number of CPUS: Target: 400, Status: 2163:** More than 60 % of available CPU resources are non-dedicated facilities. The Grid3 environment effectively shares resources not directly owned by the participating experiments (ref: list of sites).
2. **Number of Users: Target: 10, Status: 102:** About 10% of the users are application administrators who do the majority of the job submissions. However, more than 102 users are authorized to use the resources through their respective VOMS services.
3. **Number of Applications: Target: >4, Status: 10:** Seven scientific applications including at least one from each of the five GriPhyN-iVDGL-PPDG participating experiments were and continue to run on Grid3. In addition, three computer science demonstrators (instrumented gridftp, a multi-site I/O generator, and health monitors the grid) are run periodically. (ref: applications page)
4. **Number of sites running Concurrent Applications: Target: >10 Status: 17 :** This number is related to the number of Computational Service (CS,CSE) sites defined on the catalog page and varies with application.
5. **Data Transfers Per Day: Target: 2-3 TB, Status: 4 TB:** Metric met with the aid of the GridFTP-demo which runs concurently with scientific applications (ref:GridFTP statics page) http://www-mcs.ivdgl.org/mail_archive/grid3-core/msg01579.html

6. **Percentage of Resources Used: Target: 90%, Status: 40-70%:** The maximum number of CPUs on Grid3 exceeded 2500. On November 20, 2003 there were sustained periods when over 1100 jobs ran simultaneously (the metrics plots are averages over specific time bins, which can report less that the peak depending on chosen bin size). Each time we upgraded a component of the grid there was a significant length of time before stable operation was regained. In the latter part of the project most of the upgrades were for the monitoring systems, which did not prevent applications from running.
7. **Efficiency of Job Completion: Target: up to 75 %; Status: varies:** This varies depending on the application, and definition of failure. Generally speaking, for well-run Grid3 sites and stable applications this figure exceeds 90%. We have not had the time to explore why individual jobs fail.
8. **Peak Number of Concurrent Jobs: Target: up to 1000; Status: 1100:** Achieved on 11/20/03.
9. **Rate of Faults/Crashes: Target: <1/hour, Status: varies.** We have not started to measure this in quantitavely, but have begun to collect summaries from the applications groups.
10. **Operations Support Load: Target: <2FTEs, Status: Typically ~10 part time**: We were continuously adding applications and sites through the SC conference. We believe that once initial stability of a site was achieved they remained stable except for hardware problems, and that at that point the operations load was much reduced. In the case of hardware several sites replaced disks and/or nodes without perturbation to the operation of the overall system.

4 Positive Aspects

There were several positive aspects of the Grid3 environment and Grid2003 Project execution. We list some of them here:

1. **Multiple Application use of the Site Resources.** The job execution policy was implemented through the site batch queue priorities based on VO group accounts. This seems to have worked well in allowing applications from multiple VOs to run on each site. The policy that jobs of the VO owning the site have higher priority than those from other VOs seems to be working, though we have not done any in depth analysis of whether this is working in detail. There was no mechanism deployed for a site to publish its policy (due to time limitations).
2. **Software updates.** Several installations and updates of software were successfully achieved across Grid3 during the project: We upgraded from VDT 1.1.9 to VDT 1.1.11; from MDS 2.2 to MDS 2.4, from basic Glue Schema to the additional Grid3 schema and information providers; from use of Gridmapfiles to the VOMS system; as well as several versions of Mona-Lisa. The upgrades that had published update, schedule and testing plans were more successful.

   The use of Pacman for software packaging and distribution continued to be successful and provide the flexibility needed for updates as needed. Providing a common Grid3 Pacman cache enabled new packages to be easily made available to all sites and applications. This procedure was similar to last year's WorldGrid package configuration, but with a better scheme for handling application installations.
3. **Testing before Deployment**: The principal that new components and applications should be demonstrated on a single-VO grid was useful at the beginning. By the end of the project the issue of single or multiple VO was not regarded as significant. More significant is the software environment in which the new application or component is running – this needs to be demonstrated on a test or development grid whose software and configuration are as close to Grid3 as possible.
4. **Deployment of Multiple Monitoring Services:** MDS, MonaLisa, Ganglia, MDViewer and ACDC (the University of Buffalo monitoring database and job manager monitoring system) were all variously used for monitoring. Each of these monitoring applications has a different scope. All were useful.

5. **Robustness and Stability:** The Grid3 infrastructure was found to be robust to site failure, as long as core services were not affected. Applications were not robust to site failure. Our experience is that the Grid does not significantly "degrade" in availability over time, except for basic hardware failures. We took care to minimize the number of system and core service critical points that could cause failure to applications.
6. **Operations, Contacts and Support:** In general the distributed model for operations, contacts and support worked for this project. Identifying coordinators for various services helped in organization and enabled delegation of responsibilities to help in the 'scaling' problem as the size of the project grew, but clearly more work is necessary here to understand operations support at larger scales. The site administrators were responsive, and responsive to phone calls if needed when the actual deadlines were approaching. The use of email lists ensured maximum communication, but the volume of mail was overwhelming. Several bug/issues went untracked. Towards the project end a bug reporting system was established using the capabilities of the project Savannah portal.
7. **Delegated Management, Local Autonomy:** New tools and mechanisms to enhance delegated management and local autonomy, including that of specific *services*, would be welcome. Simple steps towards establishing a site validation protocol were made, but these must be improved to lessen the load on 'experts' incurred when a new (site, application, service or VO) joins the environment.
8. **Check Lists and Automated Scripts:** The Site Catalog kept track of the sites that had been validated and accepted as part of Grid3. Automated "heart beat" scripts were developed to test the basic services of each site on a regular basis and publish the results automatically into the site catalog and map. The Application checklist collected information about the requirements of each application demonstrator. The Validation scripts were developed (but not fully deployed) to help site administrators know when they had completed installation and configuration of their site. These should be improved and customized for Grid3. We need better MonALISA, and other monitoring sensor tests, which should be automated. These automated tests should be prepared early, since executing them manual consumes much time. They should be able to run in parallel over sites.
9. **Virtual Organization Management System:** Grid3 was the first U.S. deployment of the EDG/DataTAG Virtual Organization Management System (VOMS) as tested and recommended by the VOX joint project. The VOMS databases did reduce the effort to maintain plain gridmapfiles. The implementation of Group Accounts was useful – but not completely successful (see below).
10. **Application Installation:** Many/All applications were configured to be dynamically installed without the need for root privilege (i.e. without the need for site administrators to be involved). The need to manage multiple releases at a site was discussed by some VO administrators; advanced tools in this area would be welcomed.
11. **Grid3 Schema and Information Publishing:** The definition of additional attributes to define the directories to be used by the applications ($APP, $DATA etc) and extension of the Glue Schema to include these worked well. These extensions will be added with other LCG attributes into a new version of the common Glue Schema.
12. **Roles and Responsibilities:** Identifying the responsibilities of Application Administrators resulted in a small set of people being responsible for running applications across all sites on Grid3. These administrators were proactive about running many jobs on request, and following up with problems and issues they encountered. Identifying the responsibilities of Site Administrators resulted in a clear understanding of the expectations for the management of each site, cluster and headnode/gatekeeper system. Identifying Virtual Organization (VO) Administrators helped in clarifying the needed procedures and responsibilities needed for each organization participating in the multi-VO grid. We identified other Administrators e.g. GIIS/Information Services, as needed.

5 Problems

Many problems encountered during the project were understood and solved. In general, we list below only those of significance for which future work is needed and/or recommended.

1. **Application/Site Debugging:** While we were successful in running all applications fairly stably across many of the Grid3 sites in most cases each application had to be debugged on each site to achieve this. While this is more or less possible for 25 sites and 10 applications, this is clearly not a scalable solution. In practice, the larger facilities were the most successful in running the applications. The smaller test sites were useful, but in general were not as effective in running the applications over a long period.
2. **Configuring Sites and Services:** We spent inordinate amounts of time on configuration issues and fixing the same or similar problems. This was an especially significant problem for the site administgators and resulted in frustration and overload. Any time we upgraded or changed a core service it took much effort and many emails and phone calls to get back to a stably running system. (e.g. VOMS, MDS, MonaLisa).
3. **Condor configuration**: Site administrators had some difficulties configuring Condor on specific sites. The situation is improving over time.
4. **Disk Space Allocation and Cleanup:** Grid3 had no disk management services. Especially when many jobs were submitted during SC2003 week the shared and temporary disk areas were filled without warning. The space used by temporary, data staging and parameter files can fluctuate wildly depending on specifics of the application. Site administrators have the responsibility to notice and cleanup full disk areas.
5. **DOE SG Certificate Revocation:** During the course of Grid3 some certificates were revoked without the warning from the DOE Science Grid being sufficiently understood by the project. While this was probably just a mistake, the impact and confusion caused quite some effort.
6. **Effort Planning:** In several cases the effort required from the site administrators was significantly larger than that in the Grid2003 plan (.25FTE). The expected effort profile, especially the full time load during multiple week long periods of software update and configuration, needed to be better presented and understood at the start of the project.
7. **Exerciser:** The exerciser application goal is to be scheduled at each site at very low priority and publish information about those processor cycles on Grid3 that could potentially have been used by applications but were not, and to as a heartbeat for execution of jobs. The reasons why this has not been reliable are being looked into.
8. **gass-cache in AFS space:** Sites that have the gass-cache in AFS space have problems with access control on the disk areas (a kerberos ticket is needed). Solutions will in future be documented in VDT release notes. .
9. **Job Manager problem (Globus BugZilla 950) and Job Manager Patches:** Some applications were unable to run successfully on some sites – the output is not staged back correctly. http://bugzilla.globus.org/bugzilla/show_bug.cgi?id=950 . During the course of Grid3 patches were made available for some of the job managers in an attempt to alleviate the problem. There is anecdotal evidence that it helped in some sites, but not others. (see below) (incidentally we think Globus bugzilla #931, and #950 are related and should be merged.)
10. **Job Resource Requirements:** Some sites have constraints on the length of time jobs are allowed to run. Applications needed to specify the amount of CPU they would require, and in some cases could not be successfully submitted to some sites. This required use of the RSL walltime parameter.
11. **Load on the Head or Gatekeeper Nodes:** There were many instances of very heavy CPU load on site head/gatekeeper nodes. This was in some cases attributable to the number of jobs being submitted to a

site. In other cases it appears to be due to the monitoring services running on the gatekeeper node. In yet other cases it looks like some daemons may have "run amok". At present our diagnostic tools are lacking for being alerted to and being able to understand the causes. In the last month or so of the project it was suggested applications use the "Grid Monitor" to alleviate the load due to the job manager. This has helped at some sites, but not others. Work is still in progress to understand these issues.

12. **Linux Kernel versions:** We found that mixing nodes at lower versions of Redhat than 7.3 caused problems. LIGO wants to transition to Redhat 9.0. We do not understand the requirements and implications of running a mixed-linux kernel version Grid. We recommend some discussion of these issues over the next few months. In particular, it is not a "version of Redhat" problem necessarily. Other aspects that can affect compatibility include: version of Glibc; Packaging and distribution; versions of shared libraries; Location of configuration files (inetd,); Linux kernel version; gcc run time libraries; use of NIS. Pacman may address some of these problems.
13. **MDS issues:** The unreliability of MDS in providing information to the Grid3 GIIS led to a plan and decision to upgrade from a patched version of MDS 2.2 in VDT to a standalone packaged version of MDS 2.4. While the installation was successful, the problems were not solved without much effort being spent understanding configuration issues at each site. More post-mortem information is given in an appendix below.
14. **Site Host Certificates for each VOMS server:** Each Virtual Organization Management Server (VOMS) requires all Site Hosts to provide their X509 service certificate to enable the Site to create the local gridmapfile. This is not scalable and is error prone.
15. **Site Shared File System Directory Permissions**: Application use of Grid3 sites requires write permission to the shared file system and specific directories. Configuring these appropriately and robustly was error prone. Some sites left these directories world writeable, which should not be required.
16. **ssh overloaded by VDT gsi-ssh:** This has caused some problems and confusion. It would be good for VDT not to overload the "ssh" command.
17. **Virtual Data System (VDS) generated very large number of "small" DAGs:** Applications using Chimera/Pegasus that generated a very large number of DAGS and subsequent Condor-G jobs (of the order of 10,0000) had some problems. Workarounds were developed, generating a smaller number of larger DAGs.

6 Recommendations

We list here some recommendations for changes and/or future work as a result of the experiences of Grid3 to date. These will be input to the "next steps" document and together with input from the project stakeholders on requirements, priorities and available effort, will be used for future planning.

1. **Automation of Software Updates and Configuration:** Automated configuration, testing, and tuning, scripts give immediate payback in terms of effort expended to develop them compared to the effort spent doing this manually on a grid of the scale of Grid3. The project needs to be staffed and encouraged to spend the necessary effort to automate software installation and configuration as much as possible. MonaLisa is the first component, which does an "automatic update" in situ. The implications of this need to be discussed and understood. (Note: the automatic update allowed one to add new features, and have a preliminary working version that allowed more features to be added, transparently for both site administrators and users. A problem at the beginning that not enough attention and pre-release testing was made which resulted in faulty installations that had to be fixed manually, and on another occasion a buggy version damaging itself and other pieces of the Grid3 installation.) It means that versions in a "VDT distribution" can be changed once it is installed and running.

We need to to achieve push button installation for an entire Grid3+ site; maintain software over time entirely with a systematic update mechanism which requires the discipline of not fixing things site by site by hand; establish a procedure for testing new releases on a testbed before updating the production Grid3+ sites. Pacman 3 snapshots and bundles provide elements that make such a system easy to design Additionally, previous discussions in iVDGL described a site manager application, perhaps as an extension to Pacman3.

2. **API for Troubleshooting and Accounting information.** Specifically GRAM and GridFTP, the job submission and file transfer systems, respectively, should provide *direct* information without having to parse logfiles.
3. **Contact and Support Model**: It is recommended that the contact, operations and support model be revisited if Grid3 is to be sustained. Factorization of responsibilities, perhaps at the service level, should help and be explored. Automatic tools supporting delegation of management and local site autonomy should be identified and developed
4. **Document Port and Firewall Issues:** A document should be written detailing all port usages and including pointers to all information about issues related to sites being behind a firewall This document should include information on how to configure tcpwrappers for a site on Grid3. A matrix showing the port number, the local program requiring that port is open (e.g. Ganglia), and to whom the port should be accessible (the remote client such as the remote Ganglia collector at the iGOC). The document should include information of how one runs globus when not all ports can be opened. (Argonne's Jazz cluster was the only site, which required non-standard ports to be used. This didn't present much of a problem but it did require continually informing users to use these non-standard ports.)
5. **Efficiency Metrics:** The efficiency targets for Grid2003 were not met. We recommend an analysis of this, an increaseed focus on efficiency, and some more specific recommendations on how the situation can or should be improved (e.g. understanding the percentage of free cycles used).
6. **Head Node Issues:** More work needs to be done and tools provided to understand and if necessary alleviate the specific loads on Head Nodes due to monitoring (MonaLisa, MDS), and the Job Manager. Gatekeeper and job manager logs get very large when the Grid is running many jobs over many days. A Grid3 logfile rotator was written as part of Grid3. This should be included in the relevant Globus/VDT release. Additionally concrete recommendations are needed to define which services should run on which machines. (e.g. Separate GridFTP from Gatekeeper from batch system manager from Ganglia from?). Minimum hardware requirements should be published for head nodes.
7. **Job Execution Policies**: Tools should be deployed and analyses done to check that the current Grid3 job policies are actually working (hopefully the existing metrics are sufficient). A survey should be done of the applications groups to determine when each group will need more than one priority for their submitted jobs (e.g. as US CMS does already). A revised mechanism for job execution priorities may be needed at that time.
8. **Job Resource Requirements:** Experience with the variety of sites on Grid3 encourages us to recommend that sites publish more information about their policies for job execution and resource usage e.g. maximum cpu time allowed. We recommend some attention to understanding what information is needed, how it might most effectively be used, and a proposed initial implementation using an extension to the Grid3 schema.
9. **LCG-2 & Grid3 Interoperability:** Specific demonstrations of U.S. LHC application demonstrator job execution and data management across LCG-2 and Grid3 should be ongoing and be given sufficient priority in the project. Effort should be applied to ensure that the Information and Authentication Services will support this level of interoperability.

10. **Metrics and Monitoring Analysis**: The metrics information for Grid3 has only just been deployed. We recommend meetings or a workshop to review and analyse the metrics and monitoring information. The Netlogger demonstrator information should be included as part of this review. The information collected from the instrumented GridFTP should also be reviewed. We should also discuss how to collect and measure the metrics from the Grid3 plan that we so far have not measured e.g. % of jobs that succeed. We should review the level of effort applied to monitoring and determine if priorities are appropriately reflected.
11. **Registration and Use Policies, Certificate CA support:** We should develop Acceptable Use Policy, which users should sign and should be filed by the VO administrators. Negotiations should continue with each of the sites where the current Grid3 registration policy is unacceptable. If necessary a revision of the registration policy should be developed. The currently agreed upon Grid3 registration policy is not acceptable to several Grid2003 sites, notably ANL, LBNL and BU. Several sites have temporary policies in place only through the duration of the project. It is recommended that further work be done in this area in collaboration with the DOE Science Grid, laboratory facility management and LCG policy groups. Additionally, quarterly phone meetings are being held between representatives of the stakeholder organizations and the DOE Science Grid Certificate authority management. It is recommended that the project include a Liaison from the DOESG.
12. **Service Level Agreements and Operations:** Since Grid3 was running "application demonstrators" and is not a production grid, a distributed, "volunteer" model of operations is appropriate. More formal agreements and procedures will need to be developed before a shared Production Grid can be deployed and sustained. We started to develop Service Level Agreements for the VDT distribution site and the Operations Center. The stakeholders need to give guidance on their needs.
13. **Shared File System issues:** For Grid3 we took a pragmatic approach and required a shared file system across all nodes in a site. This configuration is not supported by the LCG, and in general is not an expected or acceptable requirement for an arbitrary site on a grid. We recommend understanding, for the applications currently running on Grid3, how to remove this requirement.
14. **Site Definition and Classification:** Defining classes of sites – Compute, Storage and Test – helped in the organization and management of the overall system. We recommend this be documented and extended for future work, and that service classifications (publication of supplied services by resource providers) be addressed. The precise definition of a Site and Service Provider, and the parameters and concepts associated with it also needs to be revisited to aid in deployment, integration and support. We should revisit these concepts in discussion with the LCG.
15. **Site Verification Script:** The site verification script should be completed, and new sites required to use it. It should be included in the VDT test harness or Grid3 Pacman cache.
16. **Software Licenses:** We recommend the stakeholders document the software license requirements and constraints they require for a shared grid infrastructure.
17. **Storage Services and Data Management:** Grid3 currently has little or no infrastructure support for these services. Depending on the priorities of the applications and stakeholder groups, these issues need to be discussed and a first deployment prototyped and tested.
18. **Temporary Disk Space - Management and Clean Up:** We recommend an effort to understand the requirements and principles for temporary disk space management and cleanup. If needed we recommend a small development effort to improve the tools and infrastructure for support of this management and cleanup. This will involve, and may affect, Applications, Core Services and Sites. We have no guidance at present to help site administrators estimate the disk space needed or how it will be used (e.g. size of log files per job) or how and when clean-up can be done. We recommend establishing a procedure or channel whereby application administrators can communicate this directly to target site administrators. The WorldGrid "project" concept addressed this and should be revisited.

19. **Troubleshooting:** Additional tools are necessary for troubleshooting, specifically tools for analyzing and querying log files, the ability to link a job ID on the execution side with a job ID at the submit (VO) side.

20. **Upgrade of the Grid infrastructure:** At the moment Grid3 is stable as long as the core services are not updated. We recommend that a project to upgrade the Grid3 infrastructure be executed in the course of the next three months, with the necessary procedures being developed. Possible upgrades are: Globus 2.4 – either packaged in GT 2.4 or GT 3.2; new versions of Glue Schema and Information providers, additional diagnostic and testing scripts. *Quote from a Site administrator*: "I think that after SC2003 there should be some time spent on improving updates. It seems like every time a package(s) is updated the instructions are inadequate and things break. I know this is not an easy task, but one that is a necessity for the Grid. I know that most of you, like me, have other responsibilities and more and more time is being spent on "fixing" things with Grid3"

21. **VOMS support:** We recommend development effort be allocated to a) provide a Pacman package with all host certificates for downloading by the VOMs administrators. b) Extend the site validation scripts email the host certificate of the head node to the Grid Site Coordinator for including in the package. c) Make the host scripts that generate the gridmapfiles more robust against the VOMS servers not being available. e.g. when a VOMs server is not available the latest available gridmapfile is used.

22. **Virtual Data Toolkit (VDT)**: Care must be taken not to allow the larger VDT responsibilites to be preempted by the Grid2003 and short term demonstration and milestone needs. It is time for the VDT stability release series to be delivered.The requirements on the VDT deliverables and team must be balanced across several different constituencies. Need to establish a process to ensore a proper balance of requirements amongst many VOs. E.g. Does porting to new chips/OS outweigh maintaining stability?

23. **Virtual Data Toolkit extensions** (to not forget): if GLOBUS_TCP_PORT_RANGE is configured previously, e.g. it is visible during VDT install, VDT should put it into all the necessary places (VDT setup script, Condor config {HIGH|LOW}PORT variables, globus-job-manager.conf for gatekeeper, (x)inetd conf environment for gsiftp servers. A document listing/describing all of the commonly modified configuration files and where the useful logfiles are should be provided by VDT.

7 Individual Reports:

These reports are included to provide more detailed information about specifics experiences from the project. They reflect input from volunteers and there is not an attempt to be complete.

7.1 Globus Bugzilla 950

This explains the Globus team's view on Bugzilla, Bug 950. More information can be obtained from Greg Nawrocki (nawrocki@mcs.anl.gov) It is described as "Globus 2.2.4 does not print output for various job managers, LSF and PBS". In this note, when this anomaly occurs in the context of the Globus Toolkit it will be referred to as the print output anomaly to distinguish it from what we believe to be the root of the problem, the lack of NFS close-to-open consistency. It is important to state up front that the Globus team does not consider this to be an error in the Globus Toolkit. Rather, we believe that this anomaly occurs as the result of an inconsistency within certain implementations of NFS. Nevertheless, the Globus team is committed to finding a solution by which this inconsistency can be circumvented, resulting in the print output anomaly no longer occurring. This work is considered to be a feature addition as opposed to a bug fix.

THE PROBLEM: The Globus Toolkit (GT) GRAM job manager assumes that the head node (HN) and compute nodes (CNs) in a parallel computer system share a file system that provides close-to-open consistency. This assumption means that Globus can handle the stage out of application output files as follows: a) The application running on one or more CNs closes all output files (and terminates). b) The

GRAM job manager running on the HN is notified that all output files have been closed. c) The GRAM job manager opens the output files and transfers them to the remote user. Unfortunately, it turns out that the Linux implementation of NFS v3 used on the TeraGrid (and at some other sites) does not provide close-to-open consistency. Thus, we encounter problems.

CLOSE-TO-OPEN CONSISTENCY IN LINUX NFS As noted at http://nfs.sourceforge.net: "Close-to-open cache consistency was introduced to the Linux NFS client in 2.4.20." If a Grid site NFS implementation is not providing close-to-open inconsistency, there are two (not mutually exclusive) possibilities: a) The Linux NFS implementations in use are old and/or buggy. b) The Grid site has configured NFS in a way that disables close-to-open consistency. E.g., as noted at http://nfs.sourceforge.net: "you can disable close-to-open support by using the 'nocto' mount option." In addition, it seems that there are other configuration options that may also have the effect of disabling close-to-open consistency including attribute-caching options. Presumably it should be fairly easy to determine if the NFS no close-to-open consistency problem occurs at the various Grid sites, independent of Globus Toolkit, and which of the above conditions may be causing this. When a Grid site exhibits this problem while running the Globus Toolkit, one should first determine if either of the two above conditions apply.

ADDRESSING THE PROBLEM:There are two approaches to overcoming the problem: we can modify NFS deployments where problems arise to provide close-to-open consistency, or we can modify Globus (feature addition) to work in the absence of close-to-open consistency.

7.2 ISI MDS team

From Ben Clifford and Larry Flon: Because Grid3 is based on VDT, and VDT is based on NMI, there is a lag between release of updates to the Globus Toolkit and their appearance on Grid3 sites. We should plan better for component updates. Pacman seems a viable mechanism to deploy component updates. Even so, the Globus Toolkit is not packaged to easily allow the simultaneous deployment of different component releases, which caused problems with GLOBUS_LOCATION having the proper setting when commands are executed.

+ MDS has been distributed with default configuration settings set to use the least amount of system resources. This was fine when the Toolkit is used in small testbeds, but not in large-scale deployments. We are now considering erring in the other direction, and distributing with default settings that will work in large grids.

+ Software "features": we didn't realize until this event that you cannot change GRIS timings dynamically. It is necessary to either do it in multiple steps (setting the ttl to 0, then to the desired value) or simply to restart the upstream GIIS. This is a mis-feature. It is also a mis-feature that a GRIS runs its information providers in sequence, rather than in parallel, which was the root of timing problems within Grid3.

+ Manageability: we had no easy and reliable way to determine what version of MDS was running at each site, and how it was performing. This is something we will address in the future.

+ Access: it would have made it easier to track down problems if we could have logged into the MDS nodes. If gsissh had been configured, it would allow anyone with job-execution authorization to login with a command shell. This is not something you would want in a production grid, but during the construction of the grid it would be most helpful.

7.3 Observations from US ATLAS

7.3.1 LCG-Chimera Interoperability Work

The following email was sent out on November 19 after some preliminary success was obtained in using Chimera to run an ATLAS simulation job on an LCG-1 site.

"Yesterday I was able to submit a Chimera/Pegasus job to process 4 events using atlsim in the ATLAS 6.5.0 release at the LCG-1 CERN site (adc0018.cern.ch). The job ran to completion, generating the expected ".his", ".log", and ".zebra" files. The .log and .zebra files were then successfully transferred to an output pool site at the University of Chicago (grid02.uchicago.edu). The ".his" file failed to transfer. This failure is under investigation. Both the ".log" and ".zebra" files were successfully registered in the RLS at the University of Chicago. A lot has been learned about how LCG-1 sites operate and what type of interactions can occur with non-LCG-1 sites. As expected, several minor script changes were needed to get the software to execute correctly in the lcg1 environment. A number of Chimera/Pegasus enhancements will, undoubtedly, be needed to support a uniform job submission to either an lcg1 site or a non-lcg1 site. The Chimera/Pegasus folks are working directly with Rob and Flavia to achieve this aim. The next step will be to make a few simplifications based on what I have learned already and submit longer running jobs (100 events -> 10 hours), transferring the output files to Brookhaven and registering them in the RLS at Brookhaven." - Jerry Gieraltowski – ANL

This result, however, proved to be a "lucky" fluke in that the grid gatekeeper for the stage-out of the output files chose the same worker node. More work is currently in progress by Jens Voeckler (University of Chicago) and Gaurang Mehta (Pegasus team at ISI) to make Chimera/Pegasus perform uniformly in either an LCG or Grid2003 environment.

Visibility of Working Directories and Executables. Assumptions made by Chimera/Pegasus which were valid in the Grid3 environment proved not to be true in the LCG environment. Specifically, the assumptions that a "working directory" and the actual executable were visible (i.e., shared) to all nodes are not true in an LCG environment. The fact that the grid gatekeeper (CE) can not "see" the storage element (SE) and that the use of pooled accounts results in varying logins on the worker nodes between successive jobs caused a number of problems; some of which have yet to be surmounted.

In-Progress: These items are currently being addressed by the Chimera/Pegasus team.

Missing LCG Certificate in VDT Release - The CERN LCG site uses a "host certificate" that is not included with any VDT release through VDT-1.1.11. Without the correct CA files associated with this certificate, jobs can not be submitted to the CERN LCG site.

Resolution: VDT support ticket #214 documents this problem. Alain Roy, of the VDT Team, incorporated the necessary certificate files for the missing certificate (fa3af1d7) is the upcoming release – VDT 1.1.12. Private patches of this certificate have been installed on servers at Chicago and ANL to support additional testing with this site.

Missing Port Assignments for Worker Node Out-Going Connectivity - The distributed list of ports to be assigned and opened on the various LCG nodes does not mention "outbound connectivity" ports on the WNs. The Brookhaven LCG-1 site could not support gsiftp data transfers to the WNs but other LCG-1 sites could. The actual range of ports to be open on the WNs is suggested to be 20000,25000 but the actual port range is left up to the local system admin. Some LCG-1 sites have defined a different range.

Resolution: The Brookhaven system admin (Jason Smith) has arranged to have a specific port range opened on their worker nodes. These connections support outbound connectivity only.

Correct Environment Setup Needed When Executing on Worker Nodes - Several Globus specific environment variables such as GLOBUS_LOCATION, LD_LIBRARY_PATH, and GLOBUS_TCP_PORT_RANGE are pre-defined for the user prior to execution on a worker node. However, it is not always true that the user can expect the variable PATH to be correctly defined to include Globus binaries.

Resolution: The solution to this problem is left to the user to employ. If your executable is expecting to use Globus commands, you must execute:

```
#source $GLOBUS_LOCATION/etc/globus-user.SHELL
```

where SHELL is either sh or csh before executing any other commands in your executable.

Intentional Host Blockage for GSIFTP Service - The domain name of a worker node is not always published outside of the host site. If this condition is true and a site being connected to from such a worker node is intentionally blocking hosts recognized by a specific domain, the connection will be blocked.
Example: Worker Node atlasgrid24.usatlas.bnl.gov has a numeric DNS designation of 139.66.44.2. The DNS of 139.66.44.2 is recognized as "atlasgrid24.usatlas.bnl.gov"only with the bnl site. If this worker node were to make a request to an external server which is allowing only connections from the domain ".bnl.gov", the connection would be blocked since the external server cannot translate 139.66.44.2 into any recognizable published domain.

Resolution: Servers which expect to interface to worker nodes at LCG sites must not block any domains. The only servers that would fall in this category, from the ATLAS perspective, would be database servers and gridftp servers. One server at Chicago and two servers at ANL have been setup to allow all hosts to connect to them.

File Permission Problems Due to Pooled VO Account Associations - Login accounts associated with a specific certificate and VO are dynamically assigned from an associated pool of accounts. Thus an ATLAS user may be associated with the account "atlas04" for one job and "atlas06" for the next subsequent job. The user must ensure that file permissions are correctly set for all created files which are intended to remain on the site for subsequent jobs.

Resolution: The user must ensure that created files are group writable. It is suggested that you execute "umask 002" as part of your executable if you are expecting to create any output files which must be available for subsequent jobs.

Missing PBS Patches from Globus - PBS patches, from Globus, to correctly stage stdout and stderr do not seem to be present in the current LCG-1 versions of the "pbs" and "lcgpbs" schedulers. Additional patches to fix known problems with "shared file systems" also do not seem to be included. These fixes have been identified by Jens Voeckler to be distributed with patches to VDT-1.1.11 and are identified by Globus Bug # 950.

In-Progress: Fixes for Globus Bug #950 need to be incorporated into the current LCG release of VDT. *Note 1*: LCG is currently deploying VDT-1.1.8 with additional patches specific to EDG/LCG. *Note 2*: The LCG job scheduler "lcgpbs" is not well understood by the authors so it is difficult to determine what has been modified and/or enhanced in this scheduler. Additional documentation from the authors of this scheduler is needed to understand the differences between this scheduler and the standard "pbs" scheduler.

7.3.1.1.1 Proposed Enhancements to Chimera/Pegasus

- Waiting on stuff to go in here from Jens and Gaurang.
- One possibility to enhance interoperability would be to provide a gatekeeper service on each UI with the only supported job scheduler being "fork". This would then provide a contact point between the LCG environment at the site and any other Globus-based environment external to the site. Some minor script development would be needed to translate any external job request into an LCG-specific job submission request.

7.3.2 Comments on Lessons Learned

From Ed May: The user accounts and accounting scheme appeared to be designed more for the purpose of accounting the grid usage and monitoring rather than providing a fixable and easy to use model for applications users. Assigning all users to a single unix ID resulted in confusion of who was actually running specific jobs at specific sites. This makes debugging applications more difficult. File ownership is confused and cleanup of crashed applications is difficult.Group accounts may be desirable and useful for stable production running of applications, however, more than one is desired per application. For applications development and testing, individual accounts are perferred.Mechanisms for accomodating sites which can not provide group accounts need to be explored more thoroughly.

Very little use of the "pacman" update mechanism was made, which caused extra work for administration.Too many new sites were added during the preparation for SC2003. A more structured approach to installing, testing, and verifying new sites before adding to the production list would be very desirable.

The iGOC needs to play a more proactive role in debugging problems with the grid infrastructure. Changes and additions to the Grid middleware infrastructure came too fast and were not fully tested on a separate testbed from the production testbed before deployment. More tools for sanity checking of the grid middleware configuration as deployed at a site and needed; and should be routinely run. Problem resolution was "ad hoc" and it was often unclear who was responsible for resolving specific problems and deploying and checking the resolution. A formal bug reporting and tracking system is really required for future work.

The separation of services on to different hosts at a large site is highly desirable. eg GridFTP and GridGateKeeper. Multiple gatekeepers at large site is probable desirable for capacity and reliability.

From Fred Luehring: The documentation needs work. Detailed comments are recorded elsewhere. Please contact luehring@indiana.edu for more details. Perhaps the most important problem with the installation procedure is that it essentially does not deal with the fairly common situation of a headnode with two ethernet interfaces: one for the outside and one for an internal network of the worker node. VDT does not have the option for this case and one has manually alter 10-20 configuration files in both the Globus and MDS24 directories when installing in such a situation. It would be helpful to have a document listing/describing all of the commonly modified configuration files and where the useful logfiles are. Note the problem with part of the globus-gatekeeper log file being written to /var still exists.

From P. McGuin: Some thoughts about my Grid3 experiences in no particular order:

1) Pacman / Installation issues - I think that Pacman is useful for the purposes of installing software but the grid3 installation is only as good as the package writers and cache maintainers. At one point, a pacman -updateall invocation managed to cause me a major headache as MDS 2.4 was reinstalled in a way that required me to perform surgery on the Pacman database. Package writers in general are not prepared to have their packages removed from a node. Hence each installation of the VDT always seems to require a "fresh" install. If better packages are written I think it would be easier to deal with problematic installation issues at various sites.

Other installation problems include the partial installation of a Condor jobmanager. I realize that a Condor installation will remain as part of the VDT and for some installations this makes a great deal of sense. However, for sites running other resource managers the partial configuration of the condor jobmanager is an annoyance. My simplistic method for dealing with situation is to remove the jobmanager-condor file from $GLOBUS_LOCATION/etc/grid-services. Another minor annoyance is that the PBS jobmanager setup package makes invalid assumptions about how we are setting up PBS. I always have to go back and rerun the globus setup scripts with different options after an installation or update of this Pacman package.

As mentioned in other places, the documentation for installing a grid3 site tends to be distributed and stale. More emphasis should be made so that this documentation is centralized and kept current before messages are sent to the mailing lists. My last complaint (on installation issues) has been made elsewhere, but Globus does not install well on on multi-homed (more than 1 NIC) machines. An option should be made, or instructions given, that will let me set the proper host name on the machine before installation begins so that I don't have to track down and manually correct the hostname in a myriad of scripts.

2) Support and Mailing list issues The mailing lists were/are a double edged sword in the effort to communicate in a timely manner with users and administrators. First, I would recommend that a new mailing list be created that is used for announcements only and will only accept submissions from core iVDGL management types. Second, cross-posting to multiple lists should be heavily discouraged. There were times when threads split between the lists and it was difficult to follow what was being discussed.

3) Problems with GridFTP: There is a bug in the current version of the gridftp server that causes it to fail when more than 32 mount points exist on the executing host and logging is enabled. A VDT support ticket has been opened and hopefully this will get fixed. Otherwise, we can not enable logging at our site.

Another issue that I see raised in the Grid 2003 Project Lessons document is the desire to increase the amount of traffic moved across sites to demonstrate some capacity capability. One method mentioned was the sending more data to sites with bigger pipes. Is this realistic? I would think that data movement will track execution locations. Unless bandwith is the primary discrimination for applications this may not be valid. (my $.02 worth)

4) Gatekeeper Load This has been a large problem for our site and I am still investigating reason why this is so. So far I see three or four items that warrant further investigation. First, each job accepted through the jobmanager (PBS in our case) leaves a monitoring process on the gatekeeper that checks the status of the running job every 10 seconds. This period is way too short! Ideally the monitoring timing parameter should be an RSL attribute. However, if RSL can not be changed at this point the monitoring frequency should be controllable by the site administrator. Typically I am accepting jobs that run 12 hours or longer. If I accept 50 such jobs, I can expect 216,000 qstat calls (in total) from the monitoring processes. I believe that another contributing factor to the load issues is the way that the VDT was installed at our site (mea culpa). The need for the Globus client tools to be visible on each worker and the ease of installation at the time led me to install the VDT on a network drive that is also supporting the applications and the other users. The contention for the network disk is surely a contributing factor to the load issues. The next install will be local to the gatekeeper and exported to the worker nodes. However this affects the ordering of the proper shutdown and startup of our cluster. A final contributing factor is the potential that the gatekeeper's local harddrive is not optimized for data transfers. I suspect that this is part of the problem but the evidence is anecdotal. Our system was originally configured such that the gatekeeper was running the Ganglia gmetad service which caused horrible load problems until the RRD databases used by gmetad was moved to a RAM based file system.

5) Monitoring There is way too much monitoring going on. Both MDS and MonaLisa fill different niches but each has drawbacks. MDS has certainly had problems and time will tell what MDS 2.4 can't do. MonaLisa is interesting but seems resource heavy and I never had time to investigate what configuration options could be used to reduce problems associated with it.

6) Accounts and such. As an administrator I am interested in easing my burden while allowing the VO's to do work. One issue that arises is tying batch jobs back to the submitter. When a batch jobmanager queues a job on behalf of a user the jobmanager should send an entry to the gatekeeper log(s) that indicates the batch queue ID returned along with the submitter's credentials. Ideally this entry would be tied to the data created by the xinetd started gatekeeper process to ease parsing of the logs.

Our site is still formulating usage priorities but my personal inclination is to create a situation where I can prioritize one VO's users over all other VO's. Within the prioritized VO there should be a method that allows the VO to specify a priority be that by account or centralized submissions. Outside of the prioritized VO, I would rather map all other grid users to a single account. This last account would have the absolute lowest priority in our cluster. To our local users resources consumed by grid users are the same as idle resources.

7.4 Email archives

- http://www-mcs.ivdgl.org/mail_archive/grid3-admins/threads.html Discussion list for Grid3 system administrators who are involved in and responsible for setting up and administering Grid3 sites. ~1200.
- http://www-mcs.ivdgl.org/mail_archive/grid3-core/threads.html Discussion list for Grid3 core. Those who are interested in issues concerning Grid3. ~1650

- http://www-mcs.ivdgl.org/mail_archive/grid3-apps/threads.html Discussion list for Grid3 Application Administrators. These are the people who are responsible for grid3 applications. ~191
- http://www-mcs.ivdgl.org/mail_archive/grid3-taskforce/threads.html Discussion list for Grid3 taskforce. Disscussion of issues related to grid3 policies and direction. ~1200

---

[i] iVDGL Steering Committee Meeting, Argonne Laboratory, http://www.ivdgl.org/planning/2003-06-steering/.

[ii] Foster, I. and Kesselman, C. (eds.). *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, 1999.

[iii] Foster, I. Internet Computing and the Emerging Grid. *Nature Web Matters,* 2000. http://www.nature.com/nature/webmatters/grid/grid.html.

[iv] Moore, R., Baru, C., Marciano, R., Rajasekar, A. and Wan, M. Data-Intensive Computing. In Foster, I. and Kesselman, C. eds. *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann, 1999, 105-129.

[v] Chervenak, A., Foster, I., Kesselman, C., Salisbury, C. and Tuecke, S. The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Data Sets. *J. Network and Computer Applications,* 2001.