LCG Grid Deployment Board Meeting

Minutes of the GDB meeting

CERN, 11 January 2006

Version 1.1

Amendments history:

Name	*Area*	*Date*
K.Bos	*overall*	*26 Jan 2006*
Ryszard Gokieli	*Particpants List*	*3 Feb 2006*

Agenda: http://agenda.cern.ch/fullAgenda.php?ida=a057701

Minutes: Alberto Aimar

Attendees: Please refer to list at the end of the minutes

Meeting Summary

- K.Bos started the GDB meeting with an introduction providing a summary of the main open issues. He reminded the experiments that they should send their VO Boxes questionnaires and the sequence diagrams of their use cases. Such information is needed by the VO Boxes Task Force that meets on the 24-25 January 2006. Other issues mentioned were the clarifications needed on how much memory per core is needed by each experiment and the exact calendar of the GDB meetings in 2006.
- J.Shiers presented a summary of the SC3 results and the plans for the SC3 Throughput Tests re-run in January. He also summarized the issues faced during the Christmas shutdown and how communication is being improved in order to operate the LCG as a production Grid Service.
- The computing coordinators of all four LHC experiments (ALICE/F.Carminati, ATLAS/D.Barberis, CMS/L.Bauerdick, and LHCb/N.Brook) presented a summary of their experience with SC3 and their SC4 goals. The achievements obtained, the experience gained and the improvements needed are described in detail in each presentation.
- A.Aimar presented the situation of the Phase 2 planning, the work in preparation of the Quarterly reports. He also showed the templates that should be used and what level of detail the plans should describe.
- P.Mato presented an overview of the Applications Area projects. He described the products delivered by each project and also how the Applications Area's services, platforms and configurations are now shared among all projects and with the LHC experiments.
- R.Trompert reported on the current set-up of the SARA-NIKHEF Tier-1 site and their choices in terms of mass storage system (dCache). He also presented their SC3 configuration, the experience gained and problems faced during the SC3 challenge. In particular he highlighted how needs of other VOs may need tighter security compared to the one provided by the systems developed for the LCG.
- J.Templon presented the options that are available in order to represent, in the current GLUE schema, the need of the experiments to specify which data must always stay

online on disks. This issue must to be investigated further and needs clarifications on what the experiments expect vs. how the sites intend to organize the data.

- G.Stewart presented some results of the tests done moving data among different implementations of SRMs. This operation showed major problems in moving data from dCache to DPM systems. This is an urgent issue for the GridPP infrastructure and the teams developing dCache and DPM are investigating on it.

1. Introduction (K.Bos)

See also the presentation.

The actions pending from previous meetings are mostly done.

The sequence diagrams from the experiments are urgently needed in order to be studied for the VO Boxes workshop, send them to J.Templon.

GDB will adopt the wiki style for meetings, minutes, planning, milestones and progress reports already used be other LCG groups (SC, MB, etc).

Ruth Pordes replaces Vicky White as US representative at the GDB.

VO Boxes

In November's GDB it was decided that, in the short-term, the VO boxes would be deployed to the LCG sites but that would be subject to:

- the completion of the security and operation questionnaires by the experiments
- the evaluation and acceptance by the sites

All information about the questionnaire is available at:
https://uimon.cern.ch/twiki/bin/view/LCG/VoBoxesInfo

In the longer term a task force should be formed to look for commonalities among use cases and services needed by the experiments in order to provide general services to cover those needs.

Actions:

- K.Bos – Define a procedure for VO box deployment
- Experiments - Provide sequence diagrams on the current usage of VO boxes.

The VO boxes Task Force will:

- meet in Amsterdam on the 24-25 January 2006
- the first report will be presented at the GDB of the 8 February (C.Loomis & J.Templon)

Memory requirements per core

The latest memory requirements per core are:

- ALICE, ATLAS: 2 GB
- CMS LHCb: 1 GB

Experiments should confirm that this is still the amount of memory needed, and sites should take these values into account in their procurement.

Issue for discussion:

Should a job be automatically terminated if it expands beyond a given maximum size?

New VO for GEANT4

In the previous meeting GEANT4 was proposed and accepted as a new VO. Since then the Physics application has been running on a few sites and with very positive results. The creation of the VO also highlighted that the procedures needed to agree on and create a new VO are long and rather complex.

Issue for discussion:

How can we speed up the process so that LHC-related VOs are accepted and available more quickly?

GDB Meetings in 2006

January 11	
February 8	just before Feb.10-12 SC4 Workshop 13-17 CHEP, Mumbai
March 8	
April 5-6	in Rome combined with HEPIX, Storage Workshop
Not in May	May 1- 6 ISGC, Taipei
June 7	June 12-16 T2 Workshop CERN
July 5	
Not in August	
September 6	at BNL
October 4	
November 8	
December 6	

Coming Meetings

- Jan.31 - T0/T1 network meeting in Amsterdam
  http://agenda.cern.ch/fullAgenda.php?ida=a068
- Feb.3 - 1st WLCG Collaboration Board Meeting
- Feb. 8 - GDB meeting at CERN
  http://agenda.cern.ch/fullAgenda.php?ida=a057702
- Feb. 10-12 SC4 workshop in Mumbai
  http://agenda.cern.ch/fullAgenda.php?ida=a056461
- Feb. 13-17 CHEP06 in Mumbai
  http://www.tifr.res.in/~chep06/

2. Summary of Service Challenge III (J.Shiers)

See also the presentation.

The Service Challenge 4 will become the initial WLCG Production Service and therefore SC4's goals are to test and validate all remaining offline Use Cases not considered by SC3. They details of these tests will be defined at the Mumbai workshop (Tier-1 sites) and in June 2006 (for Tier-2 sites).

Service Challenge 3

SC3 was more than testing throughput rates: several data management services were involved, SRM was required at all sites, as well as reliable gLite FTS, and the LFC file catalog deployed.

All Tier-1 and 20 Tier-2 sites took part to SC3, covering all the LCG regions. The regular workshops among sites and experiments were very useful and will continue.

The list of issue found in SC3 was presented, by N.Brook, in the previous GDB. By now many of those issues are resolved

The Throughput Tests did not steadily meet the expected targets and will be retried in January 2006. The disk tests were at about 50% of the target rates and without stability.

A lot of work is being done to fix the problems and the Throughput Test will be redone in order to reach the nominal rates (see slide 11 of the presentation).

SC3 underlined the complexity of providing reliable, high and sustained transfer rates. In addition SC3 was also useful to solve many issues and it highlighted the need of a "Grid Operations Room" in order to centralize the running of the Grid.

Another improvement is that after SC3 the LCG Grid is more operating as a permanent Service and the collaboration with the sites and experiments has been constant. There is a weekly conference call to address problems as they are raised. And this communication also helps for the preparation of SC4.

The weekly conference call is now permanent, at 16:00 Geneva time on Mondays,

Telephone +41227676000, access code 0164222

These weekly Monday's meetings always discuss:

- Summary of last week's operational issues
- Summary of each experiments' activities
- Outlook for the coming week (or two…)
- Other topical issues, regarding some SC3 setup / status etc.

Minutes go to "service-challenge-info@cern.ch" and to the MB mailing list.

3. Reports from the experiments

3.1 ATLAS (D.Barberis)

See details in the presentation.

ATLAS SC3 Summary

The original plan included complete tests of Tier-0 operations, including data distribution to Tier-1s and Tier-2s, following the Computing Model as described in the C-TDR (June'05).

In summer 2005 there was a re-scoping of the ATLAS SC3 in order to match the situation at the time: The distributed production part was removed from SC3 and the Tier-0 tests were separated in two parts: (1) Internal Tier-0 data flow and (2) Data transfer from Tier-0 to Tier-1 sites.

The results of the ATLAS Tier-0 tests (20 December 2005) are that the Tier-0 operation test reached:

- 30% of full ATLAS data rate (320 MB/s) from event filter to CASTOR
- 50% (of 340 MB/s) from CASTOR to Tier-0 CPU farm
- 50% (of 140 MB/s) from Tier-0 CPU farm to CASTOR

In December there were 16 CASTOR servers in operation and will be 48 for the rerun of the Tier-0 operation test in middle Jan. 2006.

The SC3 goals for the T0-T1 part of the service challenges are a test of the Grid middleware: FTS: reliable file transfer service, SRM: unified mass storage interface. LFC: file replica catalog.

SC3 had a 'throughput' phase, and then allowed the LHC experiments to run their own software in addition. ATLAS started officially on 1st Nov 2005. The aim was to ramp up week by week 2%, 5%, 10%, and 20% of the operational data flow values (files, speed etc) in the TDR.

The results are that out of the 10 Tier 1 sites - as of today - ATLAS managed to ship data to 9 of them. It took a lot of time and effort to set up the components necessary for each site. The reasons for the delays were usually because of human errors and inefficiencies (ATLAS and others) and of operational problems such as: storages are inaccessible, disks fill up, grid certificates expire, hardware/software upgrades, and more.

In summary the data transfer needed constant "babysitting": and we're a long way from fully functional systems able to sustain the data rates needed. Even if the Castor @ CERN problems have since then been fixed and the Grid middleware deployed by LCG was stable and gave good enough performance.

The plan earlier in 2005 was to keep ramping up and it achieved the maximum possible rate of 220 MB/s (reached several times but maximum for one hour) while the maximum sustained rate for one full day was only of 90 MB/s.

The communication with the sites was difficult and will be one of the issues to improve from SC4 onwards.

ATLAS SC4 Goals

In SC4 ATLAS will try to complete Tier-0 tests, including:

- Internal data transfer from "Event Filter" farm to Castor disk pool, Castor tape, CPU farm
- Calibration loop and handling of conditions data
- Transfer of RAW, ESD, AOD and TAG data to Tier-1s
- Transfer of AOD and TAG data to Tier-2s
- Data and dataset registration in DB (add meta-data information to meta-data DB)

ATLAS also plans to run Distributed Production with the full simulation chain running at Tier-2 (and Tier-1) sites and with data distribution to Tier-1, other Tier-2 and CAF sites.

The Distributed Analysis test will cover random job submission accessing data at Tier-1 (some) and Tier-2 (mostly) sites and will test the performance of job submission, distribution and output retrieval.

In order to achieve these goals ATLAS will need to have active FTS channels between all sites, and:

- FTS useable by anyone with a valid certificate.
- SRM "Baseline Services version" deployed everywhere
- Disk-only areas at all SEs without sudden migration of files to tape
- Agreed (and secure) way to deploy experiment's services
- Full implementation of VOMS groups and policies for job submission and data management
- Queues with different priorities for production and analysis jobs

In SC4 the services infrastructure and the middleware should really run as a production set-up, without any unplanned maintenance or sudden upgrades interruptions

3.2 ALICE (F.Carminati)

See the details in the presentation.

ALICE SC3 Goals

The SC3 main goals for ALICE were the verification of the distributed computing infrastructure, the validation of their software solutions and the possible production of useful physics.

Unlike the other experiments ALICE did not test the complete computing model as it is described in the TDR and its work was complementary to what other experiments did.

ALICE SC3 Results

ALICE tests in numbers were:

- Event sample (last two months running) 22,500 jobs completed
- Centres participation (22 in total). The Tier-1 sites participating were: CERN, CNAF, GridKa, and CCIN2P3.
- Jobs executed per site: CERN 19%, CNAF 17%, GridKa 31%, CCIN2P3 22%, total of all Tier-2 sites 11%

The general stress tests consisted of:

- Execution of 2500 jobs/24 hours
  average duration 8hours/job, for about 7500 jobs/day
- Using CASTOR2 with15K files/day (2 per job)
  each file is archive of 5 root files, 7.5 TB total

Some special tests for VO-box tests were provided by GridKa. Special thanks to K.Schwarz to the GridKa team for making 1200 CPUs available for the ALICE tests.

The global results of the SC3 ALICE stress tests were the following:

- Average duration: 12 hours, 50% of the target duration
- Jobs done: 2500, 33% of target numbers
- Storage: was 33% of the target

The *VO-box* behaved properly, without problems or interventions needed, the load profile looked proportional to the number of jobs running. The storage management software (*xroot* interface and *CASTOR2*) did not give problems. But ALICE objective was not to stress-test the mass storage system and the network

The usage of the *central AliEn services* was smooth and did not give any problem with number of jobs, no special load on proxy DB or any other service. The submission of 3000 jobs (6 master jobs) took 2 hours (0.8 jobs/sec).

ALICE VO-box specific and LCG software

A stable set of central services for production (user authentication, catalogue, single task queue, submission, retrieval of jobs) was available, with simple installation methods for the ALICE specific VO-Box software and good integration with the LCG VO-Box stack. It also demonstrated the scalability and robustness of the VO-Box model working during the mass job submission through the LCG WMS.

The issues faced were caused by the rapid updates of the MW software with inclusion of more computing centres. The centres managed to work as planned for the SC3 tests. The problems were due to the limited number of experts available. Not all services were thoroughly tested, this is valid in particular for LFC and for FTS.

ALICE operations: ALICE ran for four months of continuous MC generation and reconstruction, the contents of the events was the one requested by the ALICE Physics Working Groups.

Up to 2,400 simultaneous running jobs were repartitioned over the four Tier-1 sites (CCIN2P3, CNAF, GridKa, and CERN) and to over 10 T2's. A large number of events for user analysis were produced and the centres were responsive to the requests of changes, providing a stable operation set-up.

The use of resources was not steady, due to changes in software and the required tune-in at each centre and there was some communication problem with the centres on exactly what the VO needs to do (being fixed now by the ALICE Task Force). Lack of documentation prevented more site experts to participate in the VO operation and support

Storage management: The full migration of storage to CASTOR2, with extended tests of writing/reading and stability of operation, was very successful. The new CASTOR overcomes many old limitations and ALICE appreciated the fast response time of the CASTOR and IT experts to operational issues and quick resolution of problems.

ALICE also managed to execute extensive tests of xrootd as file transport protocol and tactical Storage Elements.

The open issues are due to the some functionality of CASTOR1 that is missing in CASTOR2. ALICE needs a uniform functionality of SRM across platforms and back-ends, allowing for general (not only FTS-type) SE tests, especially for user data analysis. For FTS ALICE needs higher level services easier to configure.

LCG Support: The interaction with LCG deployment team and site experts was excellent.

The focused Task Force meetings with participation of expert from the VO, LCG and computing centers have shown to be a very efficient discussion forum.

One open issue is that the LCG software needs a high level of expertise for deployment and support and is not clear how to combine that experience and the needs of the different Task Forces.

Application Software: All off-line ALICE basic software components (ROOT, AliRoot) went under extensive testing on several platforms (ia32, ia64, Opteron) allowing for very useful debugging and optimization. The performance was good, less than 2% of failure rate of jobs due to application software problems.

A relevant issue is that not all centers (especially Tier-2 sites) can cope with the high demands of the applications in terms of memory/CPU utilization. This is to be urgently addressed in the hardware guidelines.

In SC3 the sites have learned a lot about the problems that one has to solve in order to provide a reliable grid infrastructure. For the future ALICE is very worried about timing and stability of the whole system.

ALICE SC4 Goals

In SC4 ALICE will test the computing model as described in the TDR. F.Carminati stressed the fact that what is available at the end of SC4 is what will still be installed at LHC's start-up. An accurate planning needs to be done in a couple of weeks, with a precise time line, with clear prioritisation and costs of each feature (starting from F.Donno's list).

For ALICE is extremely important that all projects involved (EGEE, Deployment, experiments) work with the same target and the same objectives and on one product.

Action:

J.Templon asked ALICE, and the other experiments, to send to the Storage Group all inconsistencies that they notice among the different SRM implementations (CASTOR2, DCache, and DPM).

3.3 CMS (L.Bauerdick)

See all details in the presentation.

CMS SC3 Results

In the service challenges and in the future, the Grid Sites where CMS is going to run its software are:

- 7 Tier-1 sites: IN2P3, GridKa, CNAF, PIC, ASGC, RAL, FNAL
- until now 11 Tier-2s have been working actively in Service Challenge, out of 28 that have started their activities with CMS

CMS is directly working with the sites through local CMS members, the sites in the service challenge contribute providing hosting for CMS data and resources for CMS analysis and they report in the CMS weekly computing integration and operations meetings.

During SC3 the sites have reported several problems concerning the reliability, stability and complexity of the CMS and LCG applications. The fabrics are still often not sufficiently reliable and the applications are still very complex or too computing intensive.

SC3 was a major pillar of CMS integration work, with specific important goals such as: (1) testing realistic end-to-end use cases and scenarios for data transfer and data serving, (2) involving all Tier-1 and a majority of the Tier-2 sites. In slides 7-11 the main use cases are described and include the case of data transfers, analysis and data selection on Tier-1 sites, and analysis and simulation on Tier-2 sites.

The CMS central operations are based on:

- PhEDEx for the dataset placement and transfers, using the underlying grid tools and SRM storage managers.
- CRAB for analysis jobs, using the LCG RB and CondorG.
- MonALISA and the CMS dashboard for monitoring, using data from RGMA

The site-specific responsibilities are communicated through CMS people present at each site. The main goals are to ensure that the mass storage systems function, grid interfaces respond, data publishing steps and job data access succeed. Many monitoring tools are provided to the sites, but having the whole process working requires a lot of effort.

During the whole SC3 CMS managed to transfer 0.3 PB of data and 140 TB during the last two weeks of SC3 Phase 2, as much as in the whole preceding year. The average data rate end-to-end was of ~20MB/sec.

The job submission to EGEE/LCG and OSG sites was performed using CRAB, in two weeks of SC3 Phase2, about 32000 jobs were executed on 38M events, with a 2/3 completion ratio.

The daily aggregate rate peaked at 120 MB/s from Tier-0 to Tier-1 sites and at 45 MB/.s from Tier-1 to Tier-2 sites (but usually was below 30 MB/s).

SC3 was mostly devoted to debug component and systems, the results were positive, with the sites ready for real use and large storage performing well at many sites. Some of the infrastructure was not reliable enough and some planned data and work flows were de-scoped. Testing and integration with grid software is being improved now by the task forces. The CMS people at the sites are very instrumental to success.

The Distributed Conditions database was also tested with Online-2-Offline (O2O) transfers of conditions data executed with the equivalent of 6-months of data: HCAL, ECAL, and SiTracker. The tests of Frontier for Tier-0 to Tier-N (02N) are on the way; the Frontier POOL plug-in allows

squid-based database access and is in place in 10 sites. The first experience of Frontier and squid are promising but it requires more functional and stress testing.

CMS SC4 Goals

The details will be discussed at the SC4 workshop(s) but the main goal for CMS is to be able to use WLCG service for the CSA2006 use case. It should aim for mass storage validation at 400MB/s (by July 2006) at the Tier-1's and 100MB/s at Tier-2 centers.

SC4 should also demonstrate the T1-T1 and T2-any T1 connectivity. The CMS computing model has modest T1-T1 and limited T2-T1 transfers, but need to verify the ability to have generic T1-T2 transfers for each T2.

During SC4 CMS plans to scale up job submission: from currently 3000 jobs/day (for user analysis), to 200k jobs/day.

SC4 goals for CMS will be very similar to SC3 goals but needs by the end of challenge a submission rate that demonstrates the ability of a sustained use of WLCG service. The overall goal is to reach the 50% utilization of the resources available to CMS and 50% of the required T1-T2 permutations.

The important CMS use cases to test include data placement and job running, workflow at Tier-0, placement of data samples at T1 sites, data skimming, selection and transfer to T2s, calibration/alignment distributed infrastructure and also simulated analysis at Tier-2 sites.

CMS thanked the sites of the WLCG collaboration and infrastructure for the work being done and for the close collaboration.

3.4 LHCb (N.Brook)

See all details in the presentation.

LHCb SC3 Report

During SC3 Phase 1 (Oct-Nov 05) LHCb had planned to demonstrate that their Data Management meets the requirements of the LHCb Computing Model. In Phase 2 (Nov-Dec 05) they planned to demonstrate the full data processing sequence in real time, with integration of the Data and Workload Management subsystems.

The results of Phase 1 were partially successful:

- distributed 1TB of stripped data in one week from Tier-0 to Tier-1 sites, was almost successful by the criteria defined in advance;
- distributed 8 TB of data to the Tier-1 sites in two weeks, the result was acceptable but LHCb would like to repeat it as part of the SC3 re-run;
- removal of all replicas on Tier-1 sites, via LFN, within 24 hours failed. Mostly due to inconsistencies of SRM implementations;
- communication between Tier-1 sites failed because FTS did not support third-party transfers and because the configuration is very complicated (still been set up, not tested yet).

In Phase 2 data distribution was tested using the current tools to run data production in T2 centers only, with data transfers to the corresponding T1 centers. Stripping jobs were submitted, using automatic submission tools, at T1 centers as soon as the reconstructed data become available.

The SC3 lessons were extremely useful for starting to really work with the sites and with the experiment- and WLCG- services and software. Many new components were used and they showed the need for additional functionality and reliability in services (FTS, LFC, SRM, etc). CASTOR2 showed major improvements, but the coordination with site-specific systems and

organization is an issue (mass storage system, security, network, services, etc). FTS performed well for specific sites and channels after tuning. Other problems, such as time-out issues, and other limitations have been reported.

In the same period LHCb improved DIRAC for analysis, adding secure job submission, execution with the user credentials and with automatic pilot agents for submission to LCG.

More than 20 users in LHCb provided very useful feedback.

The performance tests were done using Ganga 4.0.2 and DIRAC as back-end. The tests were done with the submission of 200 jobs at RAL, CERN, PIC, NIKHEF and FZK. 197 jobs completed successfully, the remaining 3 succeeded after the first re-submission.

The job submission remains slow (~25 s per job) but will improve in the next version. The focus was on trying to cover the most complex use cases first.

LHCb SC4 and 2006 Goals

In 2006 LHCb envisages two data challenges. Their intention to start the first production challenge by end of Feb'06, with a finalised event model and producing about 200M events for which about 7.5 MSI2k.months (over 2 months) and 300 TB of storage at CERN for DC06 stage will be required

The LHCb DC'06-1 challenge will use the production services and test:

- the distribution of RAW data from CERN to Tier-1 sites
- reconstruction and stripping at Tier-1 sites including CERN
- DST distribution to CERN and to other Tier-1 sites.

The LHCb DC-06-2 challenge will be in October 2006 and will test the functioning of:

- reconstruction using Conditions data (COOL-based)
- all LHCb Tier-1 site running database service supporting COOL & 3D

The data will be accessed directly from the SE through protocols supported by ROOT and POOL and not by GridFTP/srmcp.

4. Status of Planning/Quarterly Reports (A.Aimar)

See also the presentation.

The information presented is available on the LCG Planning Wiki site https://uimon.cern.ch/twiki/bin/view/LCG/Planning. Since last GDB presentation all the Tier-1 sites have sent a new version much improved and that follows the templates distributed in November.

Once again it was stressed that the plans should cover more than just the procurement dates. The goal is to detect early if some important milestone will have a delay, and act accordingly. In particular the plan of each site should cover all 2006 and should include

- Clear capacity and performance availability at key dates Service challenges and future LCG services, such as acquisition of CPU, disks, tapes, network equipment, etc.
- Clear planning of installations and changes in the services provided, such as SRM 2.1, LFC, FTS, CE, RB, BDII, RGMA, etc.
- Several steps needed to set-up equipment and service. For example the plans should include milestones for: the selection of the product, the procurement milestones, times for installations, testing and delivery of the services

- Important infrastructure milestones that can be show stoppers, not only software and computers. For example cooling and ventilation, electrical works, 24x7 operations support, on call system, etc

The plans submitted are quite different in content and also in level of detail. A "planning check list" is being prepared and will be used before end of January to verify the plans. All Tier-1 sites should maintain updated their plans the Tier-2 sites can send their plans and report on a voluntary basis for now.

In LCG Phase 2 the reporting procedures are slightly changed. Each project (site, area projects, experiment) still provides a quarterly report. But now the QR will be lighter and more structured than in the past. Each project leader only has to:

- comment past milestones of the quarter, provided by the PO
- provide an outlook on the coming milestones
- describe other achievements, open issues and problems of the quarter

The Quarterly Reports should be filled and sent back by the 15 January 2006. A review will then be organized in order to provide a summary to the Overview Board.

In addition a more detailed "SC4 planning" activity has started, with the goal to exactly define timescale and contents of all services that will be available during SC4.

5. LCG Applications Area (P.Mato)

See also the presentation.

The focus of the LCG Applications Area is to deliver common physics applications software for the LHC experiments. It is organized in a way that it work and focus on the real experiment needs. Success is defined by the adoption and positive validation of each product by the LHC experiments.

The projects of the Applications Area are:

- SPI – Software Process Infrastructure (A. Pfeiffer)
  Software and development services: external libraries, savannah, software distribution, support for build, test, QA, etc.
- ROOT – Core Libraries and Services (R. Brun)
  Foundation class libraries, math libraries, framework services, dictionaries, scripting, GUI, graphics, SEAL libraries, etc.
- POOL – Persistency Framework (D. Duellmann)
  Storage manager, file catalogs, event collections, relational access layer, conditions database, etc.
- SIMU - Simulation project (G. Cosmo)
  Simulation framework, physics validation studies, MC event generators, Garfield, participation in Geant4, Fluka.

The *SPI (Software Process Infrastructure)* project provides general development services such as:

- Savannah service, providing bug tracking, Task management Download area, etc (>160 hosted projects, >1350 registered users).
- Software services executing installation and distribution of software (external and LCG AA projects). More than 90 external packages installed in the external service for many platforms and versions.
- Software development service, provides tools for development, testing, profiling, QA, as well as scripts and documentation adapted to LCG context

- Web and Documentation, maintains and improves existing web pages and automate content wherever possible (doxygen, LXR, wiki, etc)

ROOT provides the basic functionality typically needed by any Physics application. The current work packages are:

- BASE: Foundation and system classes, documentation and releases
- DICT: Reflexion system, meta classes, CINT and Python interpreters
- I/O: Basic I/O, Trees, queries
- PROOF: parallel ROOT facility, xrootd
- MATH: Mathematical libraries, histogramming, fitting
- GUI: Graphical User interfaces and Object editors
- GRAPHICS: 2-D and 3-D graphics
- GEOM: Geometry system
- SEAL: Maintenance of the existing SEAL packages

The main recent change in the libraries and services has been the merge of the SEAL and ROOT project. The merge consisted first in merging the development teams into a single team and then in plan an evolution of the software products into a single set of new core software libraries. For the time being the old SEAL functionality will be maintained, as long as the experiments will require it.

The *PROOF* (Parallel ROOt Facility) project aims to provide the necessary functionality that allows to run ROOT data analysis in parallel applications. A major upgrade of the PROOF system has been started in January 2005. The system is evolving from processing interactive short blocking queries to a system that also supports long running queries in a stateless client mode. Since September PROOF developers have access to a dedicated PROOF farm (32 dual processor nodes) good for developing the system but not for final users. A proposal for a PROOF Testbed is being prepared in collaboration with Alice and CMS.

The mandate of the *POOL* project is to provide data persistency for LHC physics applications. The project covers two main technology domains:

- Files - based on ROOT I/O for complex data structure: event data, analysis data
- Relational Databases – such as Oracle, MySQL, SQLite that are more suitable for conditions, calibration, alignment, detector description data

POOL itself is a framework for the persistency of arbitrary C++ objects and relationships, with file-based (ROOT) or RDBMS back-ends, it also provides file catalogues access and works with the users (ATLAS, CMS, LHCb).

In addition POOL also delivers:

- CORAL that is a general, technology-independent interface to Relational Database from C++. And is used by several other projects (COOL, POOL, ATLAS) as a technology isolation layer.
- COOL that is a framework for the handling of detector condition data associated with a time validity (used by ATLAS and LHCb)

The *SIMULATION* project includes a wide range of activities in the domain of event generators and detector simulation. It also coordinates the effort to validate the physics produced by the simulations and verifies that the results are adequate for the LHC needs.

It participates, following the needs of the LHC experiments, to the development of several simulation packages such as Geant4, Fluka, and Garfield and makes available several event generators packages to the HEP community.

All the Applications Area software (external and internal) is available under "/afs/cern.ch/sw/lcg" and is organized in a very standard manner: "<package>/<version>/<platform>". The current supported platforms are:

- slc3_ia32_gcc323: current production platform for all experiments
- win32_vc71: used mainly by LHCb as a second platform for development/testing
- slc3_ia32_gcc344: new compiler port. Almost ready.
- slc3_amd64_gcc344: requested by experiments to use resources available in other centers
- osx103_gcc33: requested by experiments as a second platform for development/testing. But has lower priority for the time being.

Everything involving the installation, configuration and the distribution of the software is also automated and in common.

The deployment of the Applications Area software (internal and external) is currently a responsibility of the experiments. For most of the packages the deployment is trivial because it just consists in a sharable library to be copied in a given directory.

But some packages require coordination between areas for consistent external software configurations and versions. POOL has suffered the most (e.g. conflicting with the file catalogs external libraries).

6. Site Report: SARA & NIKHEF (R. Trompert)

See also the presentation.

The SC3 infrastructure starting point was the existing DMF-based HSM, which has no SRM implementation and does not support some features that are standard in SRM (ex. File pinning).

dCache was chosen because it provides a SRM interface and flexibility with respect to different HSM backends, in case one will need to change backend in the future.

The SC3 configuration used different types of nodes depending on the function:

- Pool nodes
  - 4x dual Opteron, 4GB memory, 2x 1GE
  - 2TB disk cache, 12x 250GB SATA, 3ware RAID controller, disk I/O 200MB/s RAID0 (used during SC3) and 100MB/s RAID5, XFS
- Admin nodes
  - dual Xeon, 4GB memory, 2x 73GB internal disk, 2x 1GE
- MSS gateway nodes (disk servers)
  - 2x dual Xeon, 4GB memory, 2x 73GB internal disk, 2x 1GE, dual HBA FC, 1.6 TB CXFS file system (SAN shared file system)
  - runs CXFS client, read/write data directly to/from CXFS file system and RFIO daemon to put/get data to/from pool nodes
- MSS server (CXFS/DMF)
  - 4 CPU R16K MIPS, 4GB memory, 12x FC, 4x GE, 2x 36GB internal disk, 1.6 TB CXFS file system (SAN shared filesystem), 3x STK 9940B tape drives
  - CXFS MDS server, regulates access to CXFS filesystem

    - DMF (Data Migration Facility = HSM system), migrates data from disk to tape and back
- Network
    - dedicated 10GE network between CERN – Amsterdam
    - GE internal network between pool nodes and MSS gateway nodes

The expected SC3 throughput and results were as follow:

- Disk/disk: 100-110 MB/s. Problems with stability of the nodes: solved by limiting the number of I/O movers
- Disk/tape: 50 MB/s: Not enough bandwidth, and storage infrastructure was not dedicate to LCG

The percentage of computation resources used was of:

- LHCb: SARA: 28%, NIKHEF: 21%,
  for a total of GB in: 7638, GB out: 5, GB stored: 3335
- ATLAS: SARA 0%, NIKHEF 39%,
  for a total of GB in: 881, GB out: 0, GB stored: 900 (much less that the 20 TB planned).

Unfortunately the delays in the set-up of the infrastructure did not allow the participation of SARA to the ALICE data challenges. The main issues were the networking and routing problems because the 10G switch was not dedicated to LCG. Therefore it had to be configured to use either the CERN-dedicated line or the GEANT line.

Some problems of communication caused network changes (in the subnets) not to be reported or error situations not detected quickly (file transfer and network down during Christmas).

Other problems included failed transfers by attempting to overwrite files: this is not allowed by PNFS and under some situation also by dCache. Oracle databases sometimes hang and need to be restarted.

The dCache security (gsi)dcap can be circumvented. By using dccp it is possible to get anything in /pnfs/grid.sara.nl/data/<vo> by anyone. The Unix permissions on directories are not honoured and files in a directory with –“rwxr-x---“ are world readable. File permission are honoured but when data is copied in /pnfs it gets –“rw-r--r--.” The only difference using gsidcap is that you are authenticated but the behaviour above stays the same. Write permissions instead are properly protected. This whole behaviour is not adequate for several VOs that consider it too open.

On a general note SARA does not like to run rfio/dcap protocols for data access because are not authenticate.

The current SC4 are being updated and the main work will be to separate the Tier-1 tape storage for the LCG, from the general storage, and setup properly the DB nodes for FTS and LFC.

7. GlueSEType Definition (J.Templon)

See also the presentation.

The reason for this presentation is the problem of telling to a site that some files should remain on disk and never be stored onto tape. This is needed by all experiments, for data that need to be accessed always quickly.

J.Templon showed a few of the fields that could represent this information and how they are currently used, and often misused. See the presentation for the details.

What is the right way to recognize ‘disk’ SEs reliably and to pin files in online staging area of MSS-backed SEs? The situation will be even more complicated when VO subgroups show up. Several models exist (1 SE/1 group, 1 SA /1 group, etc) and will be necessary to find out what experiments need and how sites want to organize the storage.

Action:

Sites should read discuss and answer to the questions that are in the document (Modelling Storage Resources with GLUE information providers) attached to the agenda of this GDB meeting.

A discussion then started about the meaning and amount for "disk" capacity available at the sites. If it includes the disk caches that will be in front of the mass storage systems or includes only the disks that are used for processing directly at the experiment level. It was decided that the issue should be discussed further by the following MB.

J.Templon urged that it is time to start understanding how to model real SE as needed by the experiments with current GLUE schema. And to find out what experiments and users really want and how typical T1 and T2 centers plan to organize their storage. A small group is currently active and contributions on the subject are very welcome.

8. dCache to DPM performance problems (G.Stewart)

See also the presentation.

The goal of trying to move data among different storage system was to verify file transfers, learn to tune SRMs and uncover network problems, as part of SC4 preparations. With the following setup:

- Tier 2 to/from Tier 1: Target rate 300Mb/s, Transfers of 1TB
- Tier2 - Tier2: Target rate 100Mb/s

Soon was evident that that transfers from RAL dCache to Glasgow DPM was were (2Mb/s). After investigation with the FTS team was a problem with the configuration of the underlying gridftp transfer but this was not sufficient to improve considerably.

On Slide 4 one can easily notice the poor transfer time from the each of the dCACHE to the DPM sites. This is a *general* problem confirmed testing with other dCache sites (Edinburgh, DESY) and DPM sites (Edinburgh, Durham, pi.infn.it).

The issue was raised with dCache and DPM developers in December and all suggestions and shortcuts have been tried unsuccessfully. This issue is very urgent issue to resolve for GridPP (with 13 DPM sites installed/planned) and for the LCG as a whole.

The GridPP DB concluded that this is "the" most urgent issue to solve in order to prepare SC4 in time.

Actions

Item No.	Description	Owner	Status
0412-1	Contact Dave Kant at RAL re input of NorduGrid accounting data	A NorduGrid representative	Ongoing
0503-5	Ensure that all sites in country are publishing accounting data	All	Ongoing
0503-7	Begin common work list for OSG-EGEE to enable further discussion on scope and priorities for joint working	Vicky White	Done
0505-2	Write short note on what the issues of running Phedex and FTS together actually are and what the options are – schedules for each Tier-1 separately	Tier-1 managers	Open
0506-4	Provide feedback to Kors on proposed GDB dates and arrangements	All	Ongoing

LCG Grid Deployment Board Meeting

Item No.	Description	Owner	Status
0506-5	Encourage regional/site security contacts to review the new incident response procedures	Country representatives	Done
0509-1	Please review list of proposed GDB meeting dates for 2006 and provide feedback to Kors	All	Done
0510-1	Confirm SC infrastructure plans for January at November meeting.	Jamie Shiers	Done
0510-2	Develop the discussion on service monitoring to include input from Tier-1s etc.	Ian Bird/Jamie Shiers	Done
0510-3	Review and offer feedback in respect of Phase-II milestones (use link on agenda page or http://lcg.web.cern.ch/LCG/PEB/Planning/DRAFT%20-%20High%20level%20milestones%20-%2008oct05.pdf). The milestones will be put forward for approval at the November GDB.	All	Done
0510-4	Report by the next GDB on experiment plans to test the DAQ-Tier-0-Tier-1 system	Experiments	Open
0511-1	Review the December 20th LCG Service Coordination agenda and decide who should attend from your region: http://agenda.cern.ch/fullAgenda.php?ida=a056628	Country representatives	Done
0511-2	Circulate to the experiments/GCB list the operations questionnaire and (again) the security questionnaire	KB	Closed
0511-3	Define procedure by which the common services required by the VO Boxes can be investigated and an integration work plan prepared	KB	Done
0511-4	Provide sequence diagrams to demonstrate how and why VO Boxes are to be used	Experiments	Done
0511-5	Ensure named contacts in the Tier-1 plans are correct. Review/add milestones to the plan (see slide 16 of Alberto's talk). Send a new version to Alberto within 1 week	Country/Tier-1 representatives	Done
0511-6	Respond to the question posed by Nick Brook – what do you want to know more about from the experiments (information they can provide to enable better site planning)?	Country representatives	Done

List of Attendees -TBC

X means attended

V means attended via VRVS

Country	Member		Deputy	
Austria	Dietmar Kuhn	X		
Canada	Randy Sobie		Robert McPherson	
Czech Republic	Milos Lokajicek	X		
Denmark	John Renner Hansen		Anders Waananen	
Finland	Klaus Lindberg		Jukka Klem	X
France	Fabio Hernandez	X	Dominique Boutigny	
Germany	Klaus-Peter Mickel	X	Holger Marten	X
Hungary	Gyorgy Vesztergombi		Dezso Horvath	
India	P.S Dhekne			
Israel	Lorne Levinson	V		
Italy	Mirco Mazzucato		Luciano Gaido	

LCG Grid Deployment Board Meeting

Country	Member		Deputy	
Japan	Hiroshi Sakamoto		Kawamoto Tatsuo	
Netherlands	Jeff Templon	X	Ron Trompert	X
Norway	Peter Kongshaug		Farid Ould-Saada	
Pakistan	Hafeez Hoorani			
Poland	Michal Turala		Jan Krolikowski	
	Ryszard Gokieli	X	Marcin Wolter	
Portugal	Gaspar Barreira		Jorge Gomes	
Russia	Slava Ilyin		V. Korenkov	
Spain	Manuel Delfino		Andres Pacheco	X
Sweden	Niclas Andersson		Tord Ekelof	
Switzerland	Christoph Grab		Allan Clark	
			Marie-Christine Sawley	
Taiwan	Simon Lin		Di Qing	X
United Kingdom	John Gordon		Jeremy Coles	
United States	Ruth Pordes	V	Bruce Gibbard	X
CERN	Tony Cass			
ALICE	Alberto Masoni	X	Yves Schutz	
	Federico Carminati	X		
ATLAS	Gilbert Poulard	X	Laura Perini	
	Dario Barberis	X		
CMS	Lothar Bauerdick	V	Tony Wildish	
LHCb	Ricardo Graciani		Andrei Tsaregorodstev	
	Nick Brook	X		
Project Leader	Les Robertson	X		
GDB Chair	Kors Bos	X		
GDB Secretary	Jeremy Coles		Alberto Aimar	X
Grid Deployment Mgr	Ian Bird	X	Markus Schulz	X
Fabric Manager	Bernd Panzer	X		
Application Manager	Pete Mato Vila	X	Oxana Smirnova	
Security WG	David Kelsey			
Service Challenges	Jamie Shiers	X		
Quattor WG	Charles Loomis			
Networking WG	David Foster			

The following also attended:

Name	**Area**	**Name**	**Area**
Giorgio Maggi	Italy	Alberto Aimar	CERN
		Dirk Duellmann	CERN

Andrea Sciaba	CERN	Simone Campara	CERN
Christina Vistoli	CERN/Pisa		

Also attending remotely:

Name	**Area**	**Name**	**Area**
Mark van de Sanden	Amsterdam		