Challenges in Deploying the World's Largest Scientific Grid

Simone Campana, James Casey, Patricia Mendez, Roberto Santinelli, Andrea Sciaba, Jamie Shiers
CERN, Geneva, Switzerland
firstname.lastname@cern.ch

Abstract

Detectors at the world's largest scientific instrument – the Large Hadron Collider (LHC) at CERN in Geneva – have already taken their first data. Although in response to the passage of cosmic rays and not collisions between accelerated particles, it marks a milestone in the preparations of the computing and data processing services that must run for well over a decade and possibly closer to two. The LHC Computing environment is being established using a world-wide distributed and federated Grid, consisting of more than 100 sites. In order to ramp up the required services to the necessary level of functionality, reliability and scale, a series of so-called "Service Challenges" are being undertaken. The first two such challenges focused on basic infrastructure, whilst the final two include increasing levels of complexity and functionality and are designed to demonstrate that all of the offline Use Cases of the LHC experiments can be handled at the appropriate scale. Service Challenge 3 focuses on the data processing typically performed by a small team of production managers, whilst the final challenge includes all remaining Use Cases – including the ability to handle end-user analysis from a user-community of several thousand physicists.

In this paper we summarize the progress to date together with the main challenges that need to be overcome in the not-too-distant future – the first version of the full production computing environment being scheduled for end-September 2006.

1 Introduction

The Large Hadron Collider is currently being constructed in the 27km circular tunnel (buried 100m below the surface of the earth) that previously housed the LEP accelerator. Four main experiments (ALICE, ATLAS, CMS and LHCb) with a total of some 6000 physicists coming from several hundred universities or laboratories from around the world will be involved in the analysis of the torrents of data that will be produced as from around September 2007. These four experiments will run for at least 15 years and it is expected that they will:

- Discover the Higgs particle – a key particle in the Standard Model that could explain why particles have mass;
- Discover super-symmetric particles. If they exist, this would be a serious push for super-symmetric theories or "String Theories" aiming at the unification of the fundamental forces in the Universe;
- Make a precise measurement of the parity violation parameter(s). This is important to understand why our Universe is made of matter instead of an equal – and annihilating – quantity of matter and antimatter;
- Understand the early times of the Universe (from 10^{-20} to 10^{-6} seconds after the Big Bang), in particular how the soup of quarks and gluons stabilized into nucleons and then nuclei and atoms.

The complete detector of an individual LHC experiment has some ten million electronic channels. Interactions between particles accelerated in the machine take place every 25 nanoseconds. These interactions generate terabits of information that must be filtered by 2 or 3 stages of trigger logic producing between 200 MB/s and 1.2 GB/s to be recorded on permanent storage for subsequent analysis. The total data volume per experiment is estimated to be a few PB/year at the beginning of the machine's operations, leading to a total yearly production of several hundred PB for all four experiments around 2012.

The mission of the LHC Computing Grid (LCG) Project [1] is to build and maintain a storage and analysis infrastructure for the entire high-energy physics community that will use this machine. It currently consists of 160 sites, with 15000 CPUs and 9 PB of storage.

The requirements of the experiments have been defined in Computing Model documents of each of the experiments and have been refined in individual Computing Technical Design Reports that were published in June 2005. The requirements for all experiments for the year 2008 correspond to a CPU capacity of 140 million SPECint2000, to about 60 PB of disk storage and 50 PB of mass storage.

The data from the LHC experiments will be distributed around the globe, according to a hierarchical model. The raw data emerging from the data-acquisition systems will be recorded on tape and initially processed at the Tier-0 centre of LCG located at CERN. The data will be distributed to a small number of Tier-1 centres – major computer centres with sufficient storage capacity for a significant fraction of the data and with round-the-clock operation. Analysis tasks requiring access to large subsets of the raw, processed, and simulated data will take place at the Tier-1 centres. A list of the current Tier-1 centres can be found in Table 3 below.

The Tier-1 centres will make data available to Tier-2 centres, each consisting of one or several collaborating computing facilities, which can store sufficient data and provide adequate computing power for end-user analysis tasks and Monte Carlo simulation. Individual scientists will also access these facilities through Tier-3 computing resources – these can consist of local clusters in a University Department or even individual PCs.

Excellent wide-area networking is essential to the operation of the LHC Computing Grid. The national and regional research networking organizations are collaborating closely with the Tier-1 regional centres and CERN in a working group to devise an appropriate architecture and deployment plan to match LHC requirements.

Data challenges and service challenges – the latter being the primary focus of the remainder of this paper – probe and evaluate current software and hardware solutions in increasingly demanding and realistic environments approaching the requirements of LHC data taking and analysis. The service challenges, together with a clear definition and implementation of the basic services required by the experiments, form the basis of the overall plan for establishing the full-scale global LHC Grid service for the start-up of LHC in 2007.

The LCG Project depends upon several other projects for the supply of much of the specialized software used to manage data distribution and access as well as job submission and user authentication and authorization, known as the *Grid middleware*. These projects include Globus, Condor, the Virtual Data Toolkit and the gLite toolkit.

The majority of the computing resources made available by the members of the collaboration are operated as part of the EGEE Grid, a consortium of national Grid infrastructures and computing centres from 34 countries. Other resources are operated as part of other grids, such as the Open Science Grid (OSG) in the United States and the Nordic Data Grid Facility (NDGF). Achieving interoperability between different grids without compromising on the functionality constitutes a major challenge.

2 LCG Service Challenges

So as to be ready to fully exploit the scientific potential of the LHC, significant resources need to be allocated to a series of Service Challenges. These challenges are an essential on-going and long-term commitment to achieving the goal of a production-quality worldwide Grid at a scale beyond what has previously been achieved.

Whilst many of the individual components that make up the overall system are understood or even deployed and tested, much work remains to be done to reach the required level of capacity, reliability, and ease-of-use. These problems are compounded not only by the inherently distributed nature of the Grid, but also by the need to get large numbers of institutes and individuals, all with existing, concurrent and sometimes conflicting commitments, to work together on an incredibly aggressive time scale.

The service challenges must be run in an environment that is as realistic as possible, which includes end-to-end testing of all key experiment use-cases over an extended period, demonstrating that the inevitable glitches and longer-term failures can be handled gracefully and recovered from automatically. In addition, as the service level is built up by subsequent challenges, they must be maintained as stable production services on which the experiments test their computing models.

The first two challenges [2] — December 2004 and March 2005 — focused on the basic infrastructure and involved neither the experiments nor Tier-2 sites. Nevertheless, the experience from these challenges proved extremely useful in building up the services and in understanding the issues involved in offering stable production services around the clock for extended periods.

During the remainder of 2005, the Service Challenges will expand to include all the main offline use cases of the experiments apart from analysis and will begin to include selected Tier-2 sites. Additional components over the basic infrastructure will be added step by step, including experiment-specific solutions. It is important to stress that each challenge includes a set-up period, during which residual problems are ironed out, followed by a period that involves the experiments but during which the focus is on the 'service', rather than any data that may be generated and/or transferred (that is, the data are not necessarily preserved and the storage media may be periodically recycled). Finally, there is an extended service phase designed to allow the experiments to exercise their computing models and software chains.

2.1 Setup and Throughput Tests

Service Challenge 3 (SC3) consists of a set-up phase that started on 1st July 2005, during which a number of throughput tests were performed (reported on below), followed by a Service Phase due to run from 1 September 2005 until the end of the year.

The primary responsibility of the participating sites at the infrastructure level was to provide a conformant SRM 1.1 [3] interface to their managed storage. The implementations currently in use are based on CASTOR [4] at CERN, CNAF and PIC, dCache [5] at all other Tier-1s except for NDGF, who currently use DPM [6], and either dCache or DPM at Tier-2 sites. A reliable file transfer service has been set up based on the gLite File Transfer Service (FTS) [8] at CERN and at many of those Tier-1s that will support Tier-2s during the throughput tests of the set-up phase. A service based on the LCG File Catalogue (LFC) [6] has been provided at CERN for ATLAS, CMS and LHCb. ALICE, ATLAS and CMS require a local file catalogue at all sites, LHCb would like read-only replicas for reasons of availability at two external sites.

2.2 Results of Service Challenge 3

In terms of file transfer services and data rates, the goals of Service Challenge 3, that started in July 2005, were to demonstrate reliable transfers at rates of 150 MB/s per Tier-1 managed disk to managed disk and 60 MB/s to managed tape, in all cases through an agreed SRM interface. The total aggregate data rate out of CERN was targeted at 1 GB/s. These target data rates are 50% higher than those successfully achieved in Service Challenge 2, with the data rate out of CERN scaled by the increase in participating sites. *All* Tier-1 sites participated in this challenge, although a small number did not have the necessary network bandwidth installed to have been able to have achieved the target data rate. However, testing the basic infrastructure and gaining experience with the corresponding services was viewed as an important step.

Although a period of continuous running with the Tier-1 sites was achieved between 13th and 19th July, this was only at the rates of 500-600MB/s previously seen in Service Challenge 2.

Site	**Daily average (MB/s)**
ASCC, Taiwan	10
BNL, US	107
FNAL, US	185
GridKA, Germany	42
CC-IN2P3, France	40
CNAF, Italy	50
NDGF, Nordic	129
PIC, Spain	54
RAL, UK	52
SARA/NIKHEF, NL	111
TRIUMF, Canada	34

Table 1 - Best daily throughputs sustained per site

The tape write target was also not met, with sites on average managing a rate of 30MB/s.

A period of debugging followed in August to try and understand exactly where bottlenecks were occurring and for what reasons. Diagnosing tape problems were excluded initially, and this led to three lines of attack

- Understand performance implications of long round trip time (RTT), i.e. US and Asia Pacific sites;
- Understand transfers with low number of streams (<10) substantially underperformed;

- Understand why transfers to European sites with ample bandwidth did not fill the bandwidth

These tests are being currently carried out with individual sites who exhibit the characteristics under investigation. In parallel, together with DESY – the developers of the dCache SRM implementation – research was done to provide an "optimal" set of configuration and usage parameters. This will be provided to Tier-1 sites using dCache in order for them to get close to a well-tuned system quickly.

A small number of Tier-2 sites were also foreseen, focusing on those with good local support, both at the level of the required infrastructure services and from the relevant experiment. The file transfer goals between Tier-2 and Tier-1 sites were to show sustained transfers using 1 GB files of ~3 files / hour Tier-2 to Tier-1 over several days. In a number of cases, Tier-1 sites have chosen also to test Tier-1 to Tier-2 transfers – not strictly speaking part of this challenge as this corresponds to analysis data flows – part of service challenge 4. Furthermore, all bar one Tier-1 site is actively involved in establishing services for and at Tier-2 sites and a large number of such sites are preparing for the service phase. This is particularly encouraging as the total number of foreseen Tier-2 sites well exceeds 100 and their role in the overall LCG production computing environment will be significant – particularly in the area of Monte Carlo production and processing and in many cases support for end user analysis. In terms of the number of participating Tier-2s, the reliability of the services and the sustained transfer rates, the goals for this part of the challenge have been largely exceeded.

2.3 Experiment Support Activities

As described above, a key differentiator of service challenge 3 over previous tests is the inclusion of additional services for the experiments, including their own offline software, services and use cases.

Obtaining end-user "buy-in" – and indeed full production use – of the Service Challenge 3 infrastructure is considered a key milestone on the road to fully functional production services – due in one year's time – and hence are reported on in detail below.

The Service Challenge Wiki includes information concerning the status of the general purposes services that have provided (FTS, LFC, BDII nodes deployed for SC3, etc), daily log files, that track the progress of the sites throughout the service challenges and the interventions provided by the experts, and information concerning the Tier-1 and Tier-2 sites. Documentation, contact mailing lists and other support information is also provided. A multi-level support structure is also in place, with first level support – box-level monitoring and intervention – being handled by the central operations team. 2nd level support requires basic knowledge of the applications but still involves follow strict procedures, backed up as necessary by expert 3rd level support.

2.3.1 Support for the FTS

A set of FTS pilot servers has been setup at CERN, consisting of one node per experiment inside the SC3 network and hence with access to the high-speed routes defined from Tier-0 (CERN) to Tier-1 sites and scheduled transfers on them. The nodes are called fts-<VO_name>-test.cern.ch. These nodes are intended to permit the experiments to gain familiarity with the service before the Service phase, adapting their codes as necessary and to perform preliminary transfer tests aimed primarily at determining site readiness. Access to these nodes has been restricted to a small group of persons per each experiment. In each server, a set of channels has been pre-defined following the experiment requirements. For example, for the Alice experiment, the following channels have been defined: CERN-CERN, CERN-GRIDKA, CERN-IN2P3, CERN-INFN, CERN-NDGF, CERN-SARA.
Those Tier-1 sites supporting Tier-2 sites via the FTS have also deployed a FTS server. The client installation of FTS has been distributed in the LCG2.5.0 release – targeted at Tier-2 sites participating in SC3. An FTS UI has also been installed at CERN and configured according to the experiments' requirements. Finally an SC3-specific BDII has been configured containing the information of the LFC catalogs and the Storage Elements available at the various sites.

Service nodes are also available, following the convention fts-<VO_name>.cern.ch.

2.3.2 Support for the LFC

Each experiment has a pilot LFC server deployed at CERN for the same reasons as above. These machines are named: lfc-<VO_name>-test.cern.ch and they are visible from the BDII setup for the SC3 throughput phase.

The client and the server codes have been deployed with the LCG2.5.0 version. Similarly, production servers for the service phase are available, following the convention: lfc-<VO_name>.cern.ch.

2.3.3 VO Boxes – "Agents & Daemons"

The LHC experiments require a mechanism to allow them to run long-lived agents at a site. This requirement was clearly identified by the Baseline Services Working Group, aimed at establishing the basic services required for SC3 and beyond. These agents will perform activities on behalf of the experiment and its applications, such as submitting jobs to a CE, monitoring those jobs, scheduling file transfers, or scheduling database updates. No such general service currently exists, although such actions are handling by batch jobs – a misuse of the batch system. The first prototype of these boxes has been deployed in the LCG2.6.0 version. It allows a direct login via *gsissh* for the software administrators of the corresponding experiment and the registration of a proxy for an automatic renewal. It is possible to submit a few jobs through these boxes. It is important to emphasize that this LCG2.6.0 version of the VO Boxes is a prototype, with a more complete version in the next LCG2.7.0 release.

2.3.4 stdout/err monitoring tools

During the production phases of the experiments within the LCG environment the unavailability of the standard output and error files until the end of the job has been raised as an issue. In those cases where these files give information of the status of the job, it becomes a fundamental issue to provide a tool able to retrieve the partial standard output and error files at running time. As well for those jobs, which perform important checks for the rest of the job (i.e. access to storage resources or other services) the possibility to retrieve the result of these checks before the job ends can avoid long jobs running with an incorrect performance.
This motivated the baseline services workgroup to find a solution based on the experience gained with ad hoc tools provided by the experiment support group of the deployment team for the productions of the Geant4 collaboration performed in the LCG resources.
A set of tools have been deployed in the LCG2.6.0 version to solve these problems.

2.4 Experiment Specific Activities

2.4.1 ALICE

The ALICE experiment will run its whole production within the SC3 environment. Only those sites involved in SC3 are candidates for Alice.

A VO Box is required at each site. The experiment plans to run its own agents and to install its software in these boxes.

As file catalog, ALICE will use the LFC as a site-local catalog and this therefore has to be installed at each site. Their own central catalog (AliEn catalog) will play the role of a metadata catalog.

The collaboration has already begun to test the FTS performing a few transfers in the channel predefined for them.

2.4.2 ATLAS

The ATLAS activities on SC3 can be separated in two phases: a "testing phase", ending in September 2005 and a "service phase", from September to December 2005.

The main goals of the testing phase are the evaluation of new Grid services, the development and validation of new ATLAS specific components, and the integration of such components with the Grid infrastructure. More in details, the ATLAS Data Management system has been re-engineered, passing from an architecture based on a central file and metadata catalog, to a system relying on a central location catalog and local replica catalogs at each site. For this reason, some of the main activities in July 2005 were the migration of ATLAS entries from EDG-RLS to LFC, testing of LFC performance and functionality, testing of the POOL File Catalog interface with LFC and cleaning of corrupted entries in the file catalogs (LFC and EDG-RLS) and the LCG storage elements. The reliable FTS has been tested transferring large blocks of files and integrated into the ATLAS Data Management service.

The new ATLAS Data Management system, together with the new Distributed Production system will be deployed in September 2003, concluding the testing phase. ATLAS pilot jobs, together with ATLAS specific Site Functional Tests have been prepared and will be employed to guarantee the correct configuration of each site.

Their involvement in the service phase will consist of two parallel activities: the Distributed production and a Tier-0 exercise.

The Distributed Production activity will start in October 2005 with a small scale exercise. The new infrastructure will be tested and deployed to produce roughly 1M events for a data volume of 5TB. From November, the large scale production (10M events) will start.

The Tier-0 activity is aimed to simulate the reconstruction of the data as they would come out of the detector and their distribution to Tier-1s and Tiet-2s, according to the ATLAS computing model. Such exercise, also starting at the beginning of November 2005, will stress several ATLAS and Grid services, like the ATLAS Data Management systems, the CPU farms, Storage Elements at every Tier, the file catalogs and several databases.

2.4.3 CMS

The SC3 activity of the CMS experiment is aimed to test the data transfer infrastructure. CMS will use their own PhEDEx tool to perform the transfers. However, the use of FTS underneath PhEDEx in the longer term is not ruled out.

CMS is using the POOL MySQL catalog. An interface POOL-LFC has been provided by the Grid Deployment team. CMS together with the EIS team is performing functionality and performance tests of this interface. It is planned as well the migration from XML and MySQL to the new catalog.

2.4.4 LHCb

LHCb is planning to evaluate the provided services during the throughput phase. The experiment requires the LFC as the central catalog. In this sense, a conversion of the format of their replicas is needed in order to let LFC be used through the DLI interface by the Workload Management System. The inclusion of the LFC in their own framework Dirac is foreseen, as is the FTS. Initial tests using the latter are currently being performed using the channel defined for this purpose.

3 Goals of Service Challenge 4

Service Challenge 4 needs to demonstrate that all of the offline data processing requirements expressed in the experiments' Computing Models, from raw data taking through to analysis, can be handled by the Grid at the full nominal data rate of the LHC. All Tier-1 sites need to be involved, together with the majority of the Tier-2s. The challenge needs to successfully complete at least 6 months prior to data taking. The service that results from this challenge becomes the production service for the LHC and is made available to the experiments for final testing, commissioning and processing of cosmic-ray data. In parallel, the various centres need to ramp up their capacity to twice the nominal data rates expected from the production phase of the LHC, to cater for backlogs, peaks and so forth. The analysis involved is assumed to be batch-style analysis, rather than interactive analysis, the latter is expected to be performed primarily 'off the Grid'. The data rates that must be supported per site are approximately 50% higher than the target for Service Challenge 3, but must be written to tape, rather than disk, at the Tier-1 sites. The total aggregate data rate out of CERN that needs to be supported in Service Challenge 4 is 1.6 GB/s to tape at the Tier-1s, whilst the final service must be capable of running at twice this data rate.

Preparations for the services required for this challenge and indeed the initial version of the production LHC Computing Environment, have already started – overlapping with debugging of the throughput tests in Service Challenge 3 and it own service phase. Benefiting from the experience of running production services, the key hardware requirements are currently being analysed. The focus is redundancy, high availability and scalability, achieved where possible in software – thus simplifying greatly the hardware aspects.

In addition, "site-audits" are foreseen, where a team of experts analyses in detail site plans for ramping up to the nominal data rates, for the provision of the needed services to the local community and for handling inter-Tier issues and support.

An important differentiator of this final challenge is the need to support not only a strictly limited number of production users but also the full communities of the LHC experiments. Not only does this imply a number of additional services but also requires a fully fledged distributed support model. The Tier-0, Tier-1s and to a lesser extent the Tier-2s have to offer a highly reliable and functional service 24 hours a day for extended periods. End-users should not see 'cracks' in the service or major differences in functionality or support level depending on where there jobs run – which should indeed be transparent.

The majority of the over-100 Tier-2 sites that have so far been identified have a 'natural' Tier-1 that should support them (in terms of providing the service end of the reliable File Transfer Service, archival storage for Monte Carlo products generated at the Tier-2 and delivery of analysis data to the Tier-2). Negotiations are on-going with the various Tier-1 sites and the experiments to ensure that all Tier-2 sites are affiliated to a Tier-1 – a small number of which are expected to obtain these services directly from CERN itself.

Table 3 below lists the nominal data rates that should be achieved in Service Challenge 4.

Note that the *target data rate* is the data rate that must be sustained continuously during the normal operation of the LHC machine. These targets must be demonstrated during SC4. The Grid and its component Tier-0 and Tier-1 centres must be capable of sustained operation at twice this rate to allow for catching up after service interruptions, and to be able to absorb locally generated load from the Tier-1 or Tier-2 centres.

The data rate to BNL is calculated on the assumption that the full ESD will be exported to this laboratory, rather than the fractional share that is sent to other Tier-1 sites. The resources provided by a given site and the size of the user community supported will also influence these figures.

Date	Description
01/09/05	**Service Challenge 3**: start of stable service phase, including at least 9 Tier-1 and 10 Tier-2 centres.
31/12/05	**Tier-0/1 high-performance network operational** at CERN and 8 Tier-1s.
31/12/05	**750 MB/s data recording demonstration at CERN:** Data generator→ disk → tape sustaining 750 MB/s for one week.
31/01/06	**All required software for baseline services deployed** and operational at all Tier-1s and at least 20 Tier-2 sites.
30/04/06	**Service Challenge 4 Set-up:** Performance and throughput tests complete: Performance goals (see Table 3 below) CERN-disk → network → Tier-1-tape. Throughput test goal is to maintain for three weeks an average throughput of 1.6 GB/s from disk at CERN to tape at the Tier-1 sites. All Tier-1 sites must participate. The service must be able to support the full computing model of each experiment, including simulation and end-user batch analysis at Tier-2 centres.
31/05/06	**Service Challenge 4:** Start of stable service phase, including all Tier-1s and 40 Tier-2 centres.
30/09/06	**1.6 GB/s data recording demonstration at CERN:** Data generator→ disk → tape sustaining 1.6 GB/s for one week using the CASTOR mass storage system.
30/09/06	**Initial LHC Service in operation:** Capable of handling the full target data rate between CERN and Tier-1s (see Table 3). The service will be used for extended testing of the computing systems of the four experiments, for simulation and for processing of cosmic-ray data. During the following six months each site will build up to the full throughput needed for LHC operation – twice the nominal data rate.
01/04/07	**LHC Service Commissioned**: A series of performance, throughput and reliability tests completed to show readiness to operate continuously at the target data rate and at twice this data rate for sustained periods.

Table 2: LCG Deployment milestones

Centre	ALICE	ATLAS	CMS	LHCb	Target Data Rate MB/s
ASCC		X	X		110
CNAF	X	X	X	X	220
PIC		X	X	X	200
CC-IN2P3	X	X	X	X	220
GridKA	X	X	X	X	220
RAL	X	X	X	X	220
BNL		X			152
FNAL			X		50
TRIUMF		X			65
SARA/ NIKHEF	X	X		X	175
NDGF	X	X	X		90
Target data rate at CERN					1,600

Table 3: Target data rates for CERN and Tier-1 centres in SC4

The rates are based on the data flows and rates defined in the experiments' computing models and technical design reports. There exist a number of plausible use cases – primarily related to analysis – that can result in significantly higher data rates, particularly between Tier-1 and Tier-2 sites. Examples of such use cases [8] include:

1. A small 1 TB dataset transported at "highest priority" to a Tier-1 or a Tier-2 or even a user group where CPU resources are available. Allocating 3 Gb/s to such activities with perhaps 2 proceeding in parallel (in the presence of other flows and allowing for some headroom) would allow the transfer to complete in around 45 minutes.
2. 10 TB need to moved from one Tier-1 to another or a large Tier-2. It takes 450 minutes, as above so only ~two per day can be supported per 10 Gb/s link.

It is therefore to be expected as the functionality of the LCG is proven that further demands on networks and basic services such as the reliable File Transfer Service will be made.

4 Conclusions

Much progress has been made in identifying and delivering the core services that will be required to process the physics data that will be generated by the LHC and to validate the computing models and offline software prior to the arrival of real data. Work still needs to be done in a number of critical areas, including ramping up reliable transfers to the needed data rates, user support, monitoring and services in general.

5 Acknowledgements

Numerous people have made this work possible, including the Grid projects mentioned above, support staff at the many participating sites (Tier-0, Tier-1s, Tier-2s) as well as the LHC experiments. Their contribution is gratefully acknowledged. Particular thanks are due to all those who contributed to the Service Challenges – through delivery of software components, preparation and running of the tests and the setup and running of the services themselves.

6 References

[1] LHC Computing Grid Technical Design Report, CERN-LHCC-2005-024, available at http://lcg.web.cern.ch/LCG/tdr/

[2] Ian Bird, Leslie Robertson, Jamie Shiers, *Proceedings of the 2nd International Conference on Mass Storage Systems*, http://globalstor.org/

[3] Storage Resource Manager (SRM) Working Group: http://sdm.lbl.gov/srm-wg/

[4] CERN Advanced STORage Manager, (CASTOR): http://castor.web.cern.ch/castor/

[5] dCache: http://www.dcache.org/

[6] Jean-Philippe Baud, James Casey, The Evolution of LCG-2 Data Management, Proceedings of CHEP04, Interlaken, Switzerland, September 2004.

[7] EGEE Middleware Architecture, EGEE-DJRA1.1-594698-v1.0, https://edms.cern.ch/document/594698/1.0/

[8] Harvey Newman, private communication.