

# Table of Contents

<b>Summary of Open Issues reported by LHC experiments.....</b>	<b>1</b>
1. Security, authorization, authentication.....	1
2. Information System.....	1
3. Storage Management.....	1
4. Data Management.....	2
5. Workload Management.....	4
6. Monitoring Tools.....	5
7. Accounting.....	6
8. Applications.....	6
9. Deployment Issues.....	6
10. Operations.....	7
11. Castor standing open issues.....	7
12. Miscellaneous.....	7

# Summary of Open Issues reported by LHC experiments

---

## 1. Security, authorization, authentication

### 1. VOMS available and stable

Priority High

### 2. VOMS groups and roles used by all middleware

*Support for up to  $o(10)$  groups.*

Priority High

### 3. VOMS supporting user metadata (LHCb)

*Storing arbitrary user metadata should be possible in VOMS with an easy interface to access the user parameters, e.g. passing them in the VOMS proxy*

**Development:** This issue has been discussed already with the VOMS developers.

It is a feature already foreseen to come with some release of gLite.

A short term solution which does not require proxy format modifications has been provided to LHCb. A unique ID is stored together with the user DN and provided via a simple interface.

For instructions please check [here](#).

Priority Medium

### 4. Automatic handling of service proxy renewal

*The user should not need to know which server to use to register his proxy for a specific service.*

Priority High

### 5. Service needed for automatic renewal of Kerberos credentials via the Grid (ALICE)

Priority Medium

### 6. Recommendations on how to develop experiment specific secure services

*Best framework to write a secure service interacting with the Grid using delegated and automatically renewed user credentials;*

*API or "development guide" for security delegation standards and documentation;*

*GSI delegation vs. Myproxy, GT2 vs. GT4 vs. Web services, etc.*

Priority High

## 2. Information System

### 1. Stable access to static information

*Grid Information System (BDII or equivalent) should provide a stable access to the static information (services end-points and characteristics).*

Static and dynamic information should be splitted. Caching can be a solution.

Glue schema should be the same in gLite and LCG.

Priority Medium

## 3. Storage Management

### 1. SRM interface provided by all Storage Element Services

*SRM must be a fully supported specification as indicated in*

*Baseline Service group report*.

*In particular, the functionalities provided with SRM V2.1.1 are requested.*

*Mostly needed are: space reservation, file pinning, bulk operations.*

Priority High

## 2. Common and homogeneous functionality (same semantic) for all Storage Services

The APIs between SRM v1 and SRM v2 are different.

*Tests are needed to verify that the SRM implementation for a given SE type is compliant to the spec.*

*Smooth transition from SRM v1 to SRM v2.*

SRM v1 and v2 have to be maintained in parallel.

gfal or FTS should hide the differences between v1 and v2.

*SE interoperability issues must be solved.*

The functionality must be homogeneous.

*Applications must be able to access SRM functionalities at sites.*

SRM client libraries should be available to the applications.

**Priority** High

## 3. Support for disk quota management

*Support for disk quota management both at group and user level should be offered by all Storage Services (requested in particular by ATLAS, CMS and LHCb). For MSS space is considered to be unlimited.*

Developers of CASTOR, d-Cache and DPM cannot promise anything before the 3Q 2006.

**Priority** Low

## 4. Checking of the file integrity/validity after the new replica creation.

The copy operation should perform a checksum (on demand). The minimum is to check that the file size remains the same.

**LHCb/ATLAS** Remove and other operations have to be validated so that they have the correct effect on the fabric.

**Priority** Critical

## 5. Highly optimized SRM client tools

*SRM clients should be based on a highly optimized C/C++ library (gfal).*

*In particular, command line tools based on the C/C++ API (and not java based) should be available. Python binding is required.*

**LHCb:** no direct access to the information system should be required for any operation.

**Priority** Critical

## 4. Data Management

### 4.1 File Transfer Service

#### 1. Availability of File Transfer Service clients

*FTS Clients available on all SC3 sites on WNs and VOBOXes*

**Priority** High

#### 2. FTS "improvements" and feature requests as specified in the FTS workshop

*Please, check:*

FTS Workshop agenda and minutes [↗](#)

The relevant points are reported in what follows.

The [status plan](#) for FTS can be found here.

**Priority** Critical

#### 3. Reliability

*Keep retrying until told to stop. Allow for real-time monitoring of errors for transfer (parseable errors preferable) so that reshuffling of transfers, cancellation, etc. is possible.*

*Signal conditions such as source missing, destination down, etc.*

**Priority** High

#### 4. A service is needed for automatic file transfers between two sites on the Grid

*Start the transfers giving as input information the name of the SE (source and destination) and the file SURL (note: the file transfer service should not be linked to any specific catalogue; the SURL is the best specification for the file)*

Priority Critical

5. **Central entry point for all transfers**

*FTS should provide a single central entry point for all the required transfer channels including T0-T1, T1-T1 and T1-T2/T2-T1 transfers and for the T2 sites running analysis tasks.*

Priority Critical

6. **FTS should handle the automatic proxy renewal if necessary**

Priority Critical

7. **SRM interface fully integrated within FTS**

*Possibility to specify type of space, lifetime of a pinned file, etc.*

Priority Medium

8. **Support priorities, with possibility to do late reshuffling**

Priority Low

9. **Support for plug-ins to allow interactions with experiment's services**

Priority High

4.2 File Placement Service

1. **FPS plug-ins for VO specific agents**

*FPS should provide easy plug-in of the VO specific agents to implement retry policies in case of any kind of failure.*

Priority Low

2. **FPS should handle higher level operations**

*FPS should handle higher level operations such as data routing if necessary; replication operations (without specification for the file source); File Transfer Requests with multiple destination sites.*

Priority Medium

4.3 Grid File Catalogue Service

1. **LFC as global and local file catalogue**

*CMS is using LFC as global file catalogue for current MC production (phased out during 2006). Expected access rate: 100Hz peak, few Hz average as file lookup.*

Priority High

2. **LFC requested features** Support for replica attributes: tape, tape with cache, pinned cache, disk, archived tape, etc.

*Custodial flag: The concept of Master Copy that can't be deleted.*

*CMS: The availability of such attribute is mandatory for CMS.*

Priority High

3. **POOL interface to LFC**

*The functionality of accessing file specific metadata should not be provided by POOL but probably by an appropriate service such as the RSS.*

*This issue will be discussed in the TCG.*

Priority Critical

4. **Good performance**

*Performace that privileges read access, up to read-only unauthenticated instance if it helps.*

*The LFC should be highly optimized with respect to different kinds of queries, bulk operations for file and replica registration should be supported.*

Priority Critical

## 4.4 Grid Data Management Tools

1. **lcg-utils available in production**

Priority High

2. **POSIX file access based on the LFN**

*The C/C++ API (gfal library) should be able to provide POSIX file access based on the file LFN. This should include an efficient strategy for the "best replica" choice in the context of a running job. The strategy should take into account site location, prioritization of the different storage classes, the current state of the networking, etc.*

Priority Medium

3. **File access API (gfal library) using multiple instances of LFC**

*The basic file access API (gfal library) should be able to talk to several instances of the LFC catalog to ensure redundancy for high availability as well as load balancing for efficiency.*

Priority High

4. **Reliable registration service**

*Supporting ACL propagation between storages and catalogs and bulk operations.*

Priority Medium

5. **Reliable (bulk) file replica deletion service**

*Use Case: delete all SC3 data (specify a set of files) sitting on a storage element - a simple way to control that the deletion actual occurs, with automatic handling of failures.*

*ATLAS: Need to be able to delete N files in M hours.*

Priority Critical

6. **Staging service needed**

*A higher-level service to deal with staging of collection of files (datasets).*

*Such service should also operate locally at the level of a T1.*

Priority Medium

## 5. Workload Management

1. **Stable and redundant service**

*ALICE: Need a site specific configuration which contains a set of primary RB's to be used by each VO (it can be one RB or more depending on the VO requirements) and a second set of RB's which will be used in the case the first set is down. The 2 sets can be different from region to region.*

*LHCb: A list of RB's available for the VO should be defined and an easy or transparent switching mechanism from one RB to another should be provided.*

*Ideally, a single RB end-point should be provided with an automatic load balancing between the RB services behind. No loss of jobs or loss of the job results due to temporary unavailability of a RB service should happen. =The resulting RB service should provide for load balancing, resilience to failures, and scalability.*

Priority Medium

2. **Capability of handling 10\*\*6 short (>= 30') jobs in 1 day with RB service**

*ATLAS/CMS: Feature needed for SC4. The final short job number is evaluated to be 10\*\*6; thus the capability has to scale to 10\*\*6 by summer 2007.*

*LHCb: ~1Hz submission rate.*

Priority High

3. **Efficient use of information system in the match making**

*Capability of sending the jobs to the sites where the input files are present and having enough free CPU slots.*

Priority High

4. **Efficient input sandbox management (Caching of input sandboxes at sites ?)**

Priority Low

5. **Latency for job execution and job status reporting should be proportional to the expected job duration.**

Priority Medium

6. **Support for different priorities based on VOMS groups/roles**

Support requested at the global level.

*ATLAS: This should be possible without relying on a unique centralized DB (gPbox)*

Priority High

7. **The RB should reschedule the jobs in its internal task queue, using a prioritization system**

*This RB requirement does not require rearrangement of the site queues triggered by anything outside the site (RB or other services), but only of the RB internal queue. Then the jobs submitted to the different sites should be normally handled by the batch systems, in fair scheduling mode.*

*This feature is already available in gLite RB.*

Priority High

8. **Fair share across users in the same group**

Priority Medium

9. **Interactive access to running job**

*For debugging and monitoring purposes*

*CMS: top, ls, and peek at individual file level needed.*

Priority Medium

10. **Computing Element service directly accessible by services/clients other than RB**

*Get the status of the computing resource and, in particular, the number of waiting/running tasks for the given VO.*

*Submit, monitor and manipulate jobs through the CE service interface.*

Priority High

11. **Allow running special jobs (Agents) on a worker node to steal other jobs (LHCB)**

*Agents can steal execution of the jobs belonging to other users on the same worker node.*

*The Agents will run for as long as there is CPU time available on a given queue.*

Priority High

12. **Allow for changing identity of a job running on the worker node (LHCB/ATLAS)**

*This is the same as the trusted identity change service.*

*LHCB: Interrogate the site policy service for permission to run a job of a particular user.*

*In case of the positive answer, the new user proxy will be acquired from the VO service for subsequent job operations.*

*The Agent job continues even after the user job execution finished.*

*ATLAS: Using WMS to submit jobs doing data transfer on behalf of multiple users.*

Priority Medium

## 6. Monitoring Tools

1. **Tools needed to monitor transfer traffic**

Priority Medium

2. **SE monitoring**

*Needed statistics for file opening and I/O by file/dataset from SE's. Abstract load figures.*

Priority Medium

3. **A scalable tool to collect VO specific information for global operations**

*Job status/failure/progress information Monalisa or R-GMA do it.*

Priority Critical

4. **Publish/Subscribe to logging and bookkeeping and local batch system events for all jobs in the VO.**

*R-GMA can do it.*

Priority	Critical
----------	----------

## 7. Accounting

### 1. Support for accounting, with site, user and group granularity (DGAS or equivalent)

*VOMS group information should be obtained from Proxy.*

Priority	High
----------	------

### 2. Possibility to aggregate by VO (user) specified tag

*Application type (MC, Reconstruction, etc.), executable, dataset*

Priority	Low
----------	-----

### 3. Storage Element accounting aggregated by datasets (e.g. PFN directory)

Priority	Low
----------	-----

## 8. Applications

### 1. Address library conflicts with Middleware

*Castor, LSF, POOL, DPM, etc*

Priority	Critical
----------	----------

### 2. Improvements/new features for the POOL File Catalog interface

*ATLAS Being discussed with POOL and LFC teams.*

Priority	Critical
----------	----------

## 9. Deployment Issues

### 1. LFC global file catalogue available at CERN

*Request coming from CMS and LHCb.*

Priority	Critical
----------	----------

### 2. Read-only mirrors of the central LFC service

*Read-only mirrors should be available at a subset or all the T1 sites.*

*The mirror update frequency is of the order of 30-60 minutes.*

Priority	High
----------	------

### 3. Each site should provide a Storage Element with an SRM interface

Priority	High
----------	------

### 4. Different classes of SEs

*Tier1 sites as well as analysis Tier2 sites should provide different classes of storages with distinct SRM end-points:*

*MSS storage (if available ) for non-frequently accessed data (archives);*

*Disk storage with write access for production managers;*

*Disk storage with write access for all the VO users.*

*A mechanism for choosing the SE at a given site with the above mentioned characteristics should be provided.*

Priority	High
----------	------

### 5. XROOTD deployed at all sites

Priority	Medium
----------	--------

### 6. VOBOX deployment at sites

*ALICE: Needed at all sites*

*ATLAS: Needed at all sites*

*CMS: Needed at all sites*

*LHCb: Needed at all T1 centers and selected T2*

Priority	High
----------	------

### 7. VOBOX should be considered basic provided Grid services

*VOBOX are provided as basic services with specific functionality. As such, it is the responsibility of site administrators to keep them up-to-date for what concerns the middleware services they provide. It is instead responsibility of ALICE to keep the experiment software installed on these machines*

*up-to-date and to take care of possible problems that can occur when running the experiment specific agents.*

**Priority** Medium

8. **Each site should provide a Computing Element service accessible directly (LHCB)**  
*Same interface but information access on the nodes needed.*

CREAM and CMon seem to satisfy this requirement.

**Priority** High

9. **Support for short jobs**

*Every site should have dedicated queue for short (less than 30 min e.g. jobs) so that those are executed with priority. Job latencies should be proportional to job duration.*

**Priority** Medium

10. **Standards for CPU time limits**

**Priority** High

11. **Support for queues with at least 2 different priority levels**

**Priority** High

12. **Support for a system at the local queue level able to rearrange job priorities (ATLAS)**

*ATLAS: Requirement for a priority system including local queues at the sites, able to rearrange the priority of jobs already queued at each single site in order to take care of new high priority jobs being submitted. Such system requires some deployment effort, but essentially no development since such a feature is already provided by most of the batch systems, and is a local implementation, not a Grid one.*

**Priority** Medium

13. **Tools to allow for setting up of site dependent part of the VO environment (CMS)**

*Besides global VO software manager role, a mean is required to allow each site to handle the site dependent part of the VO environment setup and to fix problems with software installation.*

**Priority** High

## 10. Operations

1. **Extend Site Functional Test to a heartbeat test for all major functionalities**

*Job execution, file transfers, storage access, etc.*

**Priority** Medium

## 11. Castor standing open issues

1. **Problem using Castor2 and SRM 'isCached'**

*Castor2 has different diskpools at the backend, but the SRM only sees one of the diskpool. So a file is put onto a diskpool but is seen as 'not being cached' by the SRM because it's checking the wrong diskpools. Diskpools should either be transparent: provided that the copy between pools is fast - or not transparent, but then visible/mapped somehow to the "grid" part.*

**Priority** High

2. **A User DN is mapped to one Castor pool only**

**Priority** High

## 12. Miscellaneous

1. **xrootd interfaced with SRM**

*xrootd is about to provide SRM interface. xrootd should be provided in production. This discussion will be taken in the TCG.*

A set of workshops should be organized to discuss in details issues like this.

A first list of issues to discuss in workshops will be compiled in the TCG/BSWG.

**Priority** Low**2. CMS does not require Posix-like open of non-local SE's****3. Hosting long-lived processes**

*Work on a standard set of secure containers? e.g Apache+mod\_gridsite as a site component? How to run agents using those services? As normal jobs at the site?*

*Is it worth looking into the model of FTS with it's VO-specific agents framework? Can the same principles be applied elsewhere? Is it possible to have more documentation on this?*

**Priority** High**4. Publishing experiment specific info**

*Where should experiment specific info be published? BDII, R-GMA, ...?*

**Priority** Medium**Legenda :**

Priority	Delivery Date
Critical	January-February 2006
High	February-April 2006
Medium	Mid SC4
Low	After SC4

**Major updates:**

- Main.flavia - 29 Nov 2005 - Initial compilation starting from experiments input
- Main.flavia - 06 Dec 2005 - More input from experiments
- Main.flavia - 07 Dec 2005 - Including comments coming from discussion at BSWG
- Main.flavia - 09 Dec 2005 - Including comments from Federico Carminati
- Main.flavia - 12 Dec 2005 - Added VOMS instructions for getting User ID Metadata
- Main.flavia - 13 Dec 2005 - Added reports on the development plans of middleware (FTS, VOMS)
- Main.flavia - 11 Jan 2006 - Added experiments priority

This topic: LCG > SummaryOpenIssuesTF

Topic revision: r12 - 2007-02-19 - AlbertoAimar



Copyright &© 2008-2020 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

Ideas, requests, problems regarding TWiki? Send feedback