Geneva, 27 October, 2005

MEMORANDUM

A/To: LCG Service Team at CERN

De/From: LFC Support Team

Concerne/Subject: LFC Production Requirements Analysis

This memo describes the requirements for the production deployment of the LCG File Catalog (LFC) at CERN. It also shows how these requirements map into WLCG service levels and describes the logical layout of the service to meet the requirements. We finally describe the rollout plan to move from the current pre-production setup to a production setup.
Note that we do not consider

- What constitutes a “reliable server unit”, which may be one or more machines with connecting infrastructure such as shared disks
- Database backup and failover, assuming this is provided transparently by the underlying database service.

Requirements

ALICE

ALICE will use the LFC as a local catalog component at all sites. They will use the LFC for this component at CERN. They will use the Alien File Catalog as a central metadata catalog, but this is not part of the service provided by CERN IT.

ATLAS

ATLAS have migrated all the data they require from the EDG RLS to the pilot LFC. They use this as a global catalog currently, and would like to keep this global catalog at the same service level until at least the end of the year. They will move to using a distributed catalog model in their new Distributed Data Management (DDM) software, but this will not be ready until the end of the year. When this is in place, CERN will be required to provide a local catalog like all other grid sites.

CMS

CMS will migrate all their data from the EDG RLS to the pilot LFC. They will use this as a global catalog component for approximately 6 months. They are also in the process of developing a new data management framework. For this, they will require a local catalog component at all sites, but the technology choice for this at CERN has not yet been specified. The LFC is one of several possible options.

LHCb

LHCb wish to have a “global” catalog deployed at CERN, which will contain information on all grid replicas of LHCb data. Their main concerns are scalability of the single catalog and its response time under load.
They do not consider it necessary to have locality of catalog information and the actual data. They are happy with a single catalog endpoint for all write operations, but would like some “read-only” catalog endpoints to address the scalability and reliability concerns. These could be deployed either at the Tier-0 (Prevessin) or at a suitable Tier-1 (e.g. Lyon). Other requirements were stated:

- # Clients: O (10k) jobs running at a time. Query rate in 10-100Hz range.
- Synchronisation of local catalogs could happen as infrequently as 1 time/day. Ideally they would like 1 update/job cycle – this is approximately 20 to 30 minutes.

Other VOs

We currently provide a central catalog component for other CERN VOs – geant4, UNOSAT as well as for the dteam testing and integration activity. These would still require a central catalog, which would get intermittent and low-level usage.

WLCG Service Classes

The deployment of the LFC will be done within the context of deploying services for the WLCG service at CERN. The LCG Memorandum of Understanding (MOU) defines different levels of service availability. These have been translated into WLCG Service Classes as below (https://uimon.cern.ch/twiki/bin/view/LCG/ScFourServiceDefinition#Services)

Class	Description	Downtime	Reduced	Degraded	Available
C	Critical	1 hour	1 hour	4 hours	99%
H	High	4 hours	6 hours	6 hours	99%
M	Medium	6 hours	6 hours	12 hours	99%
L	Low	12 hours	24 hours	48 hours	98%
U	Unmanaged	None	None	None	None

For the LFC service for LHC VOs we regard the global catalog (i.e. LHCb) to be "Critical" service class, and the local catalog component to be "High" Service class. While the LHC MOU does not cover non-LHC services, we would consider the global catalog component for non-LHC VOs to be of a quality similar to the "Low" service class.

Work is ongoing within FIO to define appropriate hardware configurations that map to these server levels, based on mid-range server configurations with shared disks.

Requirements Analysis

From the requirements above, it is clear that there are three distinct categories of service required:

- A highly-scalable and reliable global catalog (LHCb).
- A local catalog component, which will hold entries for files at the Tier-0. This should have the same scalability and reliability constraints as the storage system at CERN.
- A "catch-all" global catalog for other small CERN VOs and test VOs.

In order to share infrastructure, and make failover and redundancy as easy as possible, we propose to pool the service for VOs within a category of service, e.g. all the LHC VOs which require a local catalog will share the same infrastructure.

It is noted, however, that in order to give as much isolation as possible, by default different VOs will be mapped to different hardware via DNS aliases, and will only share resources if the hardware or software on their node fails. For example, the DNS aliases lfc-alice and lfc-atlas would point to different machines, but the DNS alias would be set up in such a way that if lfc-atlas went offline, the system would automatically fail over without human intervention to the machine hosting the lfc-alice alias. Upon successful intervention, the lfc-atlas alias would automatically switch back. It is hoped that FIO/CS will provide the system to do this based on values of particular metrics within LEMON.

Although CMS have not yet decided on their technology choice for local catalog at CERN, we shall provide them with a local catalog in the pooled local service for them to continue their evaluation.

For service reasons, including recovery, we believe the best way to produce a read-only replica for LHCb is to do it at the Tier-0. Initially this would be a read-only LFC daemon pointing at the same database backend. We would like to move this to point at a backup copy of the database, preferably hosted in Prevessin for reliability reasons.

Figure 1 shows the logical schematic of the components to be deployed, showing the three distinct categories of service.

LFC Production Deployment Layout

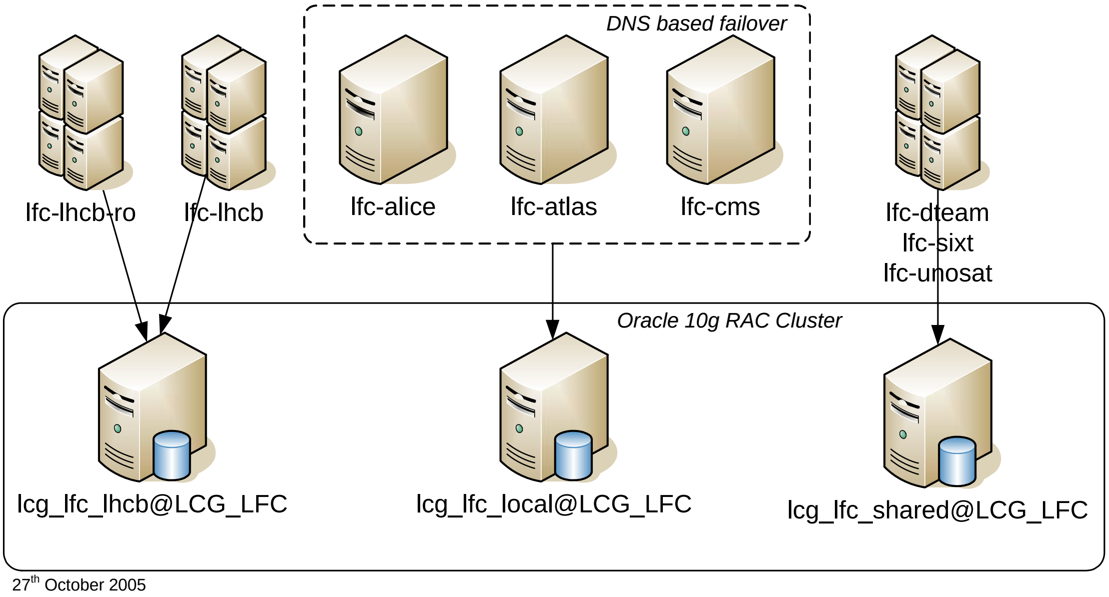

Figure 1: LFC Production Deployment Layout

Table 1 below shows the various aliases that would be used initially, and their purpose, along with the initial mapping onto physical hardware.

Alias Name	Host	Usage
lfc-alice	lfc007	Local ALICE catalog
lfc-atlas	lfc008	Local ATLAS catalog
lfc-cms	lfc009	Local CMS catalog
lfc-lhcb	lfc010	Global LHCb catalog
lfc-lhcb-ro	lfc011	Global LHCb R/O catalog
lfc-alice-test	lfc007	Local ALICE catalog (1)
lfc-atlas-test	lfc002	Temp ATLAS global catalog
lfc-cms-test	lfc003	Temp CMS global catalog
lfc-lhcb-test	lfc010	Global LHCb catalog (2)
lfc-dteam-test	lfc005	Global DTEAM catalog (Certification)
lfc-dteam	lfc004	Global DTEAM catalog (Production)
lfc-shared	lfc004	Global shared catalog – for UNOSAT, SIXT, GEANT4

Table 1 Mapping of aliases onto physical hosts

Notes:

1. This is mapped to the same host as lfc-alice, since they have already been using the lfc-alice-test endpoint as a local catalog and populated it
2. This is mapped to the same host as lfc-lhcb, since they have already been using the lfc-lhcb-test endpoint as a global catalog and populated it

Rollout Schedule

We plan to use the intervention between phase 1 and phase 2 of the SC3 service phase (Oct 31st/Nov 1st) to re-configure the current pre-production service to use the database accounts on RAC as described above. This will avoid more costly migrations later on when the number of experiment entries in the catalog increases.

The work being done with regards to high-availability hardware configurations for the LFC server nodes is not complete, so this will only be carried out after the conclusion of SC3 service phase. We will use the current (non high availability) hardware configuration for the LFC servers, with the aim of updating this before the start of SC4.