LHCb Analysis Centres

History

- First draft started: 26 May 2009
- First draft completed: 29 May 2009
- Last update: 3-Feb-10

Introduction

The LHCb Collaboration relies heavily on Grid resources for implementing its Computing Model that has been described in the "LHCb Computing TDR" published in 2005.

The present document describes further on which sites LHCb analysis can be performed and in particular how non-Grid analysis is supported by the LHCb Computing project, which was not part of the Computing TDR.

LHCb Grid Analysis Centres

As defined in the LHCb Computing TDR, all Tier1s are LHCb Analysis Centres. This means that all real data (μ)DST files are replicated to all these sites[1]. MC data are distributed to CERN and two Tier1s.

Note that CERN shared batch and interactive facility (lxplus, lxbatch), being available to the whole Collaboration are considered by WLCG as part of the Computing resources made available to LHCb and accounted as such in the LHCb Computing Resources pledges[2].

Other Grid sites may possibly be used as LAC under the following conditions:

- Provide sufficient storage resources in the form of a Grid-SE (i.e. SRM-enabled) for handling sizeable datasets on disk (typically more than 100 TB). The LHCb Applications and DIRAC should support data access at that flavour of SE[3].
- Provide CPU resources that otherwise would not be pledged for LHCb, i.e. additional resources to those pledged for simulation.
- Provide enough local LHCb manpower (i.e. in an LHCb institute related to that Grid site) for managing the datasets, in coordination with the LHCb Data Management team.
- Be open to the entire LHCb Collaboration, i.e. conform to the definition of LHCb Grid site as specified above.

---

[1] The only limitation to this policy is the provision by sites of enough disk storage resources. In this case older datasets may be removed from storage in order to allow more recently produced datasets to be replicated. Sites may also receive only real data and no simulated data.

[2] To the exclusion of any possible LHCb Tier3 located at CERN.

[3] The Computing Project commits to implement this support if not yet available.

Non-Grid Analysis

In addition to analysis on the Grid at LACs as described above, some LHCb analysis will be performed on various types of Computing resources:

- Local clusters, a.k.a. Tier3 centres. They can possibly be co-located with a Grid site (i.e. share the same SE and/or the same WNs), or have their own non-Grid local storage. They may even have access to an SE attached to a LAC, in which case no replication of data would be necessary.
- Desktop/laptop with sufficient local storage, a.k.a. Tier4.

None of these Computing Resources is pledged nor accounted through the WLCG and Computing RRB and its usage is entirely under the responsibility of the local LHCb team (Tier3) or user (Tier4).

Nevertheless, the LHCb Computing project supports these sites in providing tools, depending on the site storage configuration.

File replication to a Grid-SE

For Grid-SE accessible to a Tier3/4, DIRAC replication tools should be used for transferring the required dataset to the site. To this effect, the SE should be declared in the DIRAC Configuration System (CS) and replicas will be registered in the LHCb File Catalog (LFC).

The advantages of replication are:

- Central knowledge of replicates of some datasets, i.e. no need to keep a local bookkeeping of which files have been copied.
- Compatibility with ganga for local job submission using LFNs. Jobs can then be submitted to the Grid if needed without changes.

When needed, files should mandatorily be removed using the adequate DIRAC tools, in order to keep the LFC in synch with the storage.

Whenever a Tier3/4 has direct access to a LAC storage, no replication is required and data access should take place using the associated file access protocols (e.g. at CERN, provided the Castor client library is available on the resource).

File copy to non-Grid storage

For local storage attached to a Tier3/4 that does not provide Grid access (i.e. no SRM), files should simply be copied using the DIRAC copy tool. Usually one would expect this storage to be a file system (possibly shared by a cluster). The local copies of the files would *not* be registered on the LFC, nor will the SE be registered in the DIRAC CS.

Such storage has to be fully handled by the group or user that is using it. File access protocol should be one of those supported by the LHCb Core Software.

Software distribution

The LHCb Computing Project provides and supports installation tools for the supported applications on the supported platforms. This includes also distribution of ganga and DIRAC and their installation either on shared or private resources.

The LHCb applications are also available for the CERN Virtual Machine (CernVM) that should be used in case the platform used is not supported by LHCb Core Software.

Glossary

- CS: DIRAC Configuration System.
- DIRAC: LHCb Grid interface for Workload and Data management
- Ganga: user interface for job submission, developed jointly by ATLAS and LHCb.
- LAC: LHCb Analysis Centre.
- LFC: LHCb File Catalog.
- LFN: Logical File Name.
- SE: Storage Element on the Grid. It must provide an SRM interface.
- SRM: Storage Resource Manager. A service abstracting storage resources on the Grid.
- WLCG: Worldwide LHC Computing Grid.