Contribution ID: **146**   Type: **oral presentation**

DIRAC Data Management System

Friday 11 May 2007 09:40 (20 minutes)

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).

The LHCb experiment being built to utilize CERN's flagship Large Hadron Collider will generate data to be analysed by a community of over 600 physicists worldwide. DIRAC, LHCb's Workload and Data Management System, facilitates the use of underlying EGEE Grid resources to generate, process and analyse this data in the distributed environment. The Data Management System, presented here, provides real time, data driven distribution in accordance with LHCb's Computing Model.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.

During WLCG's Service Challenge 3 (2005) the LHC experiments tested their bulk data transfer frameworks for the distribution of files from CERN to their associated Tier1 sites. LHCb's Data Management System was interfaced to gLite's File Transfer Service (FTS) and showed, during optimal performance, it could meet the requirements of the Computing Model for data distribution from Tier0. During LHCb's Data Challenge 2006 the Data Management System components for data driven distribution was used to transfer over half a million files, using FTS, and their registration in grid catalogue (LFC). Future developments of the Data Management System will embrace the new functionality provided by the Storage Resource Manager (SRM) v2.2 implementations allowing LHCb to fully manage space reservation, pinning and storage classes of its data.

With a forward look to future evolution, discuss the issues you have encountered (or that you expect) in using the EGEE infrastructure. Wherever possible, point out the experience limitations (both in terms of existing services or missing functionality)

The DIRAC Data Management System has recently been reviewed to allow transfer requests present on site VOBOXes to be processed using FTS to increase the efficiency and redundancy of data distribution. This development as well as the systems mentioned above will be presented here.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications

The data volumes produced by the LHC experiments are unprecedented, rendering individual institutes and even countries, unable to provide the computing and storage resources required to make full use of the produced data. EGEE Grid resources allow the processing of LHCb data possible in a distributed fashion and LHCb's Computing Model is based on this approach. Data Management in this environment requires reliable and high-throughput transfer of data, homogeneous access to storage

resources and the cataloguing of data replicas, all of which are provided by EGEE infrastructure and utilized by LHCb.

Primary author: SMITH, Andrew Cameron (CERN - University of Edinburgh)

Presenter: SMITH, Andrew Cameron (CERN - University of Edinburgh)

Session Classification: Data Management

Track Classification: Data Management