

DIRAC Workload Management System

Thursday 10 May 2007 12:00 (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).

DIRAC (Distributed Infrastructure with Remote Agent Control) is the Workload and Data Management system (WMS) for the LHCb experiment. The DIRAC WMS offers a transparent way for LHCb users to submit jobs to the EGEE Grid as well as local clusters and individual PCs. This paper will describe workload management optimizations which ensure high job efficiency and minimized job start times.

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.

DIRAC submits Pilot Agents to the EGEE Grid via the gLite WMS as normal jobs. Pilot Agents then request jobs from the DIRAC Workload Management System after the local environment has been checked. Therefore DIRAC realizes the so-called PULL paradigm which ensures a high efficiency for LHCb Grid jobs.

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 possibility of using generic VO Pilot Agents is very exciting and DIRAC is ready to exploit tools such as glexec in order to optimize workloads. This would allow DIRAC to work in a 'filling' mode by which multiple jobs may be requested for execution by Agents deployed to Grid Worker Nodes in a secure way.

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 computing requirements of the LHCb experiment can only be fulfilled through the use of many distributed compute resources. DIRAC provides a

robust platform to run
data productions on all the resources available to LHCb
including the EGEE Grid.
More recently, user support was added to DIRAC that greatly
simplifies the procedure
of submitting, monitoring and retrieving output of Grid jobs for
the LHCb user
community.

Author: Dr PATERSON, Stuart (CERN)

Presenter: Dr PATERSON, Stuart (CERN)

Session Classification: Workflow

Track Classification: Workflow