

Extension of DIRAC to enable distributed computing using Windows resources

Wednesday 13 February 2008 12:00 (25 minutes)

The LHCb experiment, designed for high-precision studies of matter-antimatter asymmetries in the decays of b-hadrons, is one of the four main experiments at the CERN Large Hadron Collider (LHC). DIRAC has been developed to meet the experiment's need for processing petabytes of data per year, using globally distributed resources. It can be used either as a standalone Grid implementation, or as an optimisation layer on top of another system, such as the EGEE Grid, and has performed impressively in data challenges held since 2002. Although mostly written in Python, which is largely platform-independent, various features of the implementation have previously limited use of DIRAC to Linux machines. We have extended DIRAC to allow its use also on Windows platforms, making the core code more generic in a number of places, integrating Windows-specific solutions for certificate-based authentication and secure file transfers, and enabling interaction with Microsoft Windows Compute Clusters.

1. Short overview

We give details of the implementation, deployment and testing of a distributed computing system that provides transparent access to both Linux and Windows resources. The system presented is an extension of the DIRAC Workload and Data Management System, developed in the context of the LHCb experiment, and used successfully with Linux machines for several years. We have added the possibility to also use Windows resources, significantly increasing the experiment's data-processing capabilities.

3. Impact

An initial, small-scale deployment of the new system allows jobs submitted through DIRAC to be run on 100+ Windows CPUs, distributed between the Universities of Bristol, Cambridge and Oxford, and allows jobs to be submitted from Windows machines to run at the 120+ sites with Linux nodes made available through DIRAC. We have tested the different submission paths, and have successfully used the distributed Windows resources to optimise selection criteria for one of the b-hadron decay channels of interest in LHCb. Some sites are able to offer dedicated Windows clusters, not previously accessible through Grid systems, and others have large numbers of Windows machines that may be idle at certain periods, for example in teaching laboratories. The Windows-enabled version of DIRAC allows these resources to be added to existing Grid-based Linux resources, under a single workload management system, increasing data-processing capabilities by a significant factor.

4. Conclusions / Future plans

The DIRAC system continues to evolve, and we are helping ensure that newer releases are portable across platforms. We plan to deploy DIRAC at more sites with Windows machines available, and in particular aim to demonstrate the gains that are possible by using non-dedicated resources. Tests so far under Windows have involved running only a single application per job, and as a next step we will be running chained applications, covering simulation, digitisation and reconstruction.

Provide a set of generic keywords that define your contribution (e.g. Data Management, Workflows, High Energy Physics)

Workload and Data Management, Distributed Windows Resources, Cross-platform Job Submission

Authors: TSAREGORODTSEV, Andrei (Centre de Physique des Particules de Marseille); PARKER, Andy (Univer-

University of Cambridge); COLES, Jeremy (University of Cambridge); HARRISON, Karl (University of Cambridge); LYUT-SAREV, Vassily (Microsoft Research); LI, Ying Ying (University of Cambridge)

Presenter: COLES, Jeremy (University of Cambridge)

Session Classification: Workflow and Parallelism

Track Classification: Existing or Prospective Grid Services