Journal of Physics: Conference Series **664** (2015) 062009 doi:10.1088/1742-6596/664/6/062009

The GridPP DIRAC project: Implementation of a multi-VO DIRAC service

Bauer D, Colling D, Currie R, Fayer S, Huffman A, Martyniak J, Rand D and Richards A

Blackett Laboratory, Imperial College London, Prince Consort Road, London SW7 2BW UK

E-mail: lcg-site-admin@imperial.ac.uk

Abstract. The GridPP consortium provides computing support to many high energy physics projects in the UK. As part of this GridPP offers access to a large amount of highly distributed resources across the UK for multiple collaborations. The userbase supported by GridPP includes hundreds of users spanning multiple virtual organisations with many different computing requirements. In order to provide a common interface to these distributed a centralised DIRAC instance has been setup at Imperial College London. This paper describes the experiences learnt from deploying this DIRAC instance and the modifications that have made to support the GridPP use case.

1. Introduction

The GridPP consortium [1, 2] is a collaboration in the UK which provides computing support for multiple High Energy Physics (HEP) experiments within the UK. GridPP provides resources to hundreds of users spanning across multiple Virtual Organisations (VOs). Many of these VOs have little or no internal computing support and are reliant on resources provided by GridPP. The resources offered by GridPP are highly distributed and so in order to provide more centralised support a grid middleware solution has been sought. After considering possible alternative technologies, the DIRAC Interware [3] project was chosen to provide a fixed interface to the different resources offered across the UK. Providing a DIRAC solution to each VO is a potentially expensive task in terms of both dedicated hardware and manpower. In order to significantly reduce this overhead a single instance is preferred, as such a centralised DIRAC service is hosted for GridPP at Imperial College London.

2. DIRAC

The DIRAC Interware project was started in 2003 to provide a grid job submission and managment framework. It was originally designed for the LHCb collaboration and is still used extensively throughout the LHCb computing framework. Due to its origins the DIRAC project was originally designed around the use-case of a single-VO, with more recent developments focussing on improved support for multiple VO instances. DIRAC is currently used within many HEP collaborations including Belle2, LHCb, IHEP, ILC and many others. This use of DIRAC has contributed to the robustness and reliability of the of core DIRAC framework. DIRAC offers a highly modular platform to develop upon which allows for new modules to be written in a way which takes advantage of the core proven technologies.

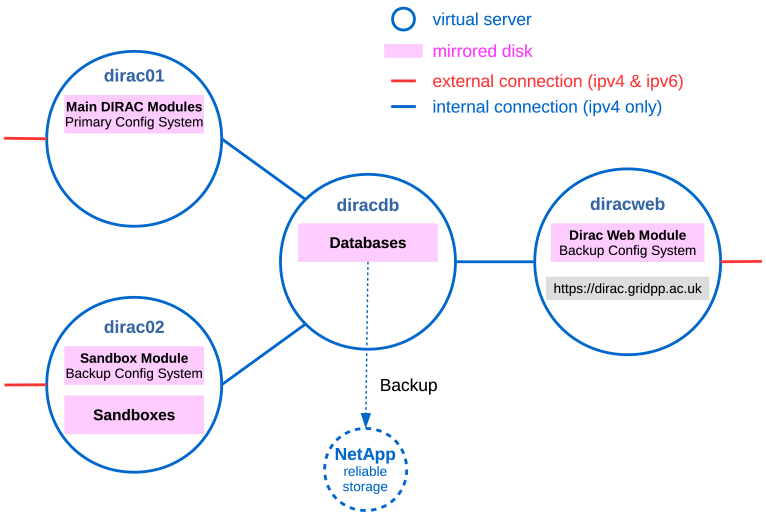

Figure 1. This figure displays the VM server setup used to deploy the DIRAC instance at Imperial College. This infrastructure has been designed with reliability and redundancy in mind. In order to achieve this the service has been split across multiple machines, each with a dedicated role.

The DIRAC model for submitting jobs allows for the user to submit a description of the job to be run in addition to a sandbox containing data required by the job. This requires both an external interface for communicating with the DIRAC server in addition to a shared storage resource for handling user sandboxes. DIRAC also provides a data management service in the form of the DIRAC file catalogue which allows for data to be managed on remote storage elements.

3. Production Setup

The production DIRAC service offered at Imperial College has been designed for maximal reliability and redundancy. One of the goals of offering this service at Imperial is to reduce the complexity of supporting many users by the GridPP consortium. In addition to this the DIRAC service is intended to run as an automated service requiring minimal intervention from the administrator.

In order to achieve these goals the DIRAC service is divided across several virtual machines (VM) as shown in Figure 1. The main DIRAC service is hosted on the `dirac01`, this machine contains the main plugins and manages the full DIRAC service. The various databases used by this service are hosted on the `diracdb` which is externally backed up. User sandboxes are managed by the `dirac02` which has storage allocated for sandbox data using the sandbox module. All User jobs can be viewed/managed the DIRAC web server hosted on the `diracweb` machine which is accessible at: `https://dirac.gridpp.ac.uk`.

In order to submit jobs to this DIRAC service the `dirac01` and `dirac02` machines are both accessible through external connections running IPv4 and IPv6 as is the interactive web portal (`diracweb`). Internal communication between the servers are managed through internal IPv4 connections only with the system hosting the databases (`diracdb`) not directly exposed to external connections.

4. The GridPP DIRAC Module

After some success in administering a vanilla DIRAC instance at Imperial College it was strongly felt that the configuration system wasn't appropriate for this use case. Changes to the DIRAC configuration are typically propagated by editing configuration files used by the DIRAC configuration service. The size and complexity of these configuration files was observed to quickly scale with the number of users and VOs which were supported by the service.

In order to build upon the features offered by vanilla DIRAC a configuration module was written for use at Imperial College. This additional 'GridPP' module is designed to communicate with VOMS and BDII services to automatically update DIRAC resources. The design of this module closely follows DIRAC coding conventions which should allow for wider adoption of this module by other communities. The main changes in the GridPP module are contained in the 'Users and Groups' and the 'AutoBDII2CS' agents. These changes are described below.

4.1. Users and Groups Agent

The Users and Groups agent is responsible for configuring the different user and group roles within the DIRAC service using information from a central VOMS service. The version of this agent within the GridPP module is based upon that used within vanilla DIRAC. The main modification to this service is that it has been designed to work with multiple-VOs. In order to do this it has been expanded to make use of a different VOMS server for each of the VO. In addition to this the agent has also been expanded to work with multiple VOMS roles as well as allowing for multiple groups within each role.

4.2. AutoBDII2CS Agent

The AutoBDII2CS agent is responsible for comparing Computing Elements (CE) and Storage Elements (SE) available in the BDII service against the elements stored in the Configuration System. The base agent in DIRAC simply advertises differences found in the BDII by sending an automated email to the administrator. The administrator then has to manually intervene and add resources to the configuration system. In order to simplify this process the agent within the GridPP module automatically updates the configuration system within DIRAC when new elements are discovered within the BDII. This agent is also capable of managing multiple SE and automatically handles creating paths to each of them.

4.3. Future Developments

In addition to the modifications described above there are future developments which are focused on the automation of this multi-VO DIRAC instance. These developments include further work into automating the addition and removal of various CE and SE elements from the DIRAC configuration system. In addition to this future work will look to add more VO-specific controls into the automated configuation offering features such as system such as site masking.

5. User Support

In order to encourage adoption of this service and to lower the barrier of entry for new users wanting to make the most of GridPP resources, a number of user based support services have been implemented. A wiki page has been setup for new users which provides a checklist of prerequisits and other useful information. This information is available at: `https://www.gridpp.ac.uk/wiki/Dirac`.

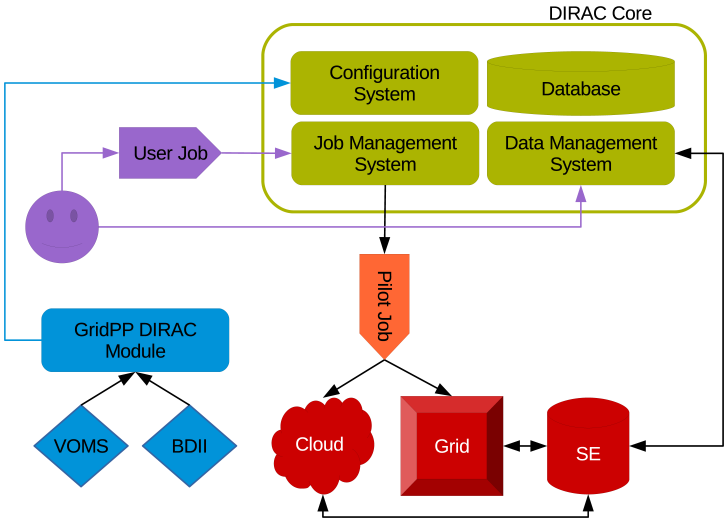

Figure 2. Figure showing the basic layout of a typical DIRAC setup with the GridPP module developed at Imperial College shown. The primary function of the GridPP module is to provide automatic configuration information for DIRAC through communication with VOMS and BDII.

In addition to this web resource an interactive support mailing list has been setup for users of this service `gridpp-dirac-users@imperial.ac.uk`. The workflow promoted and supported by GridPP is to encourage users to submit and manage DIRAC jobs through the Ganga [5] project and to use the GridPP DIRAC backend.

6. Conclusion

We have shown that with some additional configuration that DIRAC can be used sucessfully to meet the needs of the GridPP organisation. This also demonstrates how the modular design of the DIRAC framework can be built upon to be used in a complex enviroment with multiple VOs.

References

[1] Faulkner P J W et al, 2006, The GridPP collaboration, GridPP: development of the UK computing Grid for particle physics, *J. Phys. G: Nucl. Part. Phys.* **32** N1-N20
[2] Britton D et al, 2009 GridPP: the UK grid for particle physics, UK e-Science All Hands Conference, *Phil. Trans. R. Soc.* A **367** 2447-2457
[3] Tsaregorodtsev A 2003 DIRAC: distributed infrastructure with remote agent control *Proc. Conference for Computing in High Energy and Nuclear Physics* (La Jolla, California)
[4] Field L and Schulz M, 2004 Computing in High Energy Physics and Nuclear Physics 2004 Interlaken, Switzerland pp.723
[5] Moscicki J T et al, 2009 Ganga: a tool for computational-task management and easy access to Grid resources *Computer Physics Communications* Vol. **180** Issue 11