

The GridPP DIRAC project - DIRAC for non-LHC communities

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 J. Phys.: Conf. Ser. 664 062036

(<http://iopscience.iop.org/1742-6596/664/6/062036>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 155.198.8.192

This content was downloaded on 21/09/2016 at 15:51

Please note that [terms and conditions apply](#).

You may also be interested in:

[The GridPP DIRAC project: Implementation of a multi-VO DIRAC service](#)
D Bauer, D Colling, R Currie et al.

The GridPP DIRAC project - DIRAC for non-LHC communities

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

HEP Group, Department of Physics, Imperial College London, London, SW7 2AZ, UK

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

Abstract.

The GridPP consortium in the UK is currently testing a multi-VO DIRAC service aimed at non-LHC VOs. These VOs (Virtual Organisations) are typically small and generally do not have a dedicated computing support post. The majority of these represent particle physics experiments (e.g. NA62 and COMET), although the scope of the DIRAC service is not limited to this field. A few VOs have designed bespoke tools around the EMI-WMS & LFC, while others have so far eschewed distributed resources as they perceive the overhead for accessing them to be too high. The aim of the GridPP DIRAC project is to provide an easily adaptable toolkit for such VOs in order to lower the threshold for access to distributed resources such as Grid and cloud computing. As well as hosting a centrally run DIRAC service, we will also publish our changes and additions to the upstream DIRAC codebase under an open-source license. We report on the current status of this project and show increasing adoption of DIRAC within the non-LHC communities.

1. Introduction

GridPP [1] is a collaboration of physicists and computer scientists from leading UK universities, Rutherford Appleton Laboratory (RAL) and CERN. It was originally created to provide computing support for the LHC experiments but its scope has since expanded to include non-LHC physics communities and other research disciplines. Initially the LHC experiments used the EMI-WMS and LFC for their workload and data management on the Grid. However over the years the LHC experiments have slowly moved away from these products to using their own custom software. Due to the lack of use by any large experiments, the EMI-WMS and LFC [2] are likely to become unsupported in the near future. Small experiments with little or no dedicated Grid computing support which had been relying on these products now must find a replacement. After assessing alternative technologies it was decided that the DIRAC framework [3] would be suitable for their requirements.

2. GridPP DIRAC

The DIRAC project (Distributed Infrastructure with Remote Agent Control) [3] was originally designed as a workload management system for the LHCb experiment. It is also used by other collaborations including Belle2 [4], BESIII [5], ILC [6] and many others. Although it was initially meant to serve a single VO, the DIRAC consortium aimed to keep it VO neutral and extended its functionality to support multiple VOs.



The DIRAC workload management system (Figure 1) is a pilot based framework. These lightweight agents are sent to multiple resources, and once in a running state they pull in a job payload. This mechanism greatly improves resource usage efficiency compared to direct submission. DIRAC also provides a file catalog (DFC) with metadata support; this is an improvement over the EMI-LFC which had no metadata capability.

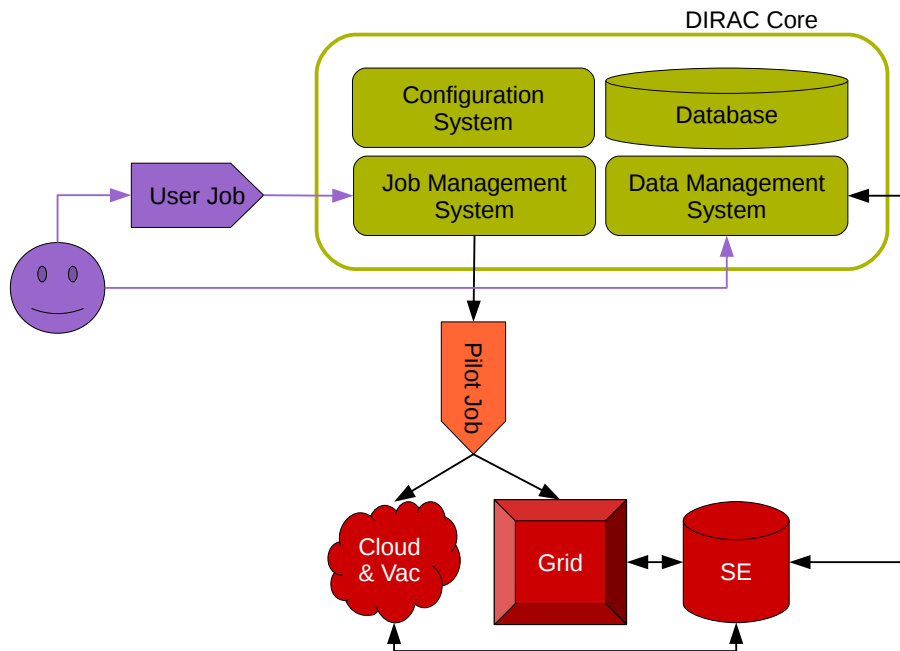


Figure 1. Schematic of the DIRAC framework

2.1. DIRAC set-up and maintenance

The DIRAC software is designed to be easily installed. The base framework is written in Python consisting of a number of core & optional modules. To simplify the installation procedure all of the dependencies are included in an externals bundle which is unpacked automatically. This avoids the need to install any extra packages on the base system and can therefore be installed by any non-administrative user. The installation directory is completely self contained, this allows different users to make use of different configurations on the same machine. As-is typical for such a complex system, the configuration of the DIRAC server is non-trivial. As expected for a product still in very active development, the documentation is not yet complete. To facilitate the user interface (UI) installation we have provided a step-by-step guide on a wiki. This simple set of instructions and checklist reduces the need of expert knowledge.

2.1.1. GridPP resource definition and management The definition of Grid sites and their resources have been created manually, based on the configuration obtained from the DIRAC service used by the LHCb collaboration. We added some new UK Grid sites and modified the existing ones as necessary. Maintaining the list of resources is error prone so currently only the minimum required subset is provided; users and roles are also only added on request. We are

aware that this solution is not ideal due to its limited scalability, so we have started extending the existing DIRAC modules to automate the maintenance of these lists [7].

2.1.2. Server monitoring We have enabled DIRAC specific Site Availability Monitoring (SAM) tests developed by GridPP [8] on our DIRAC server. These tests are run periodically several times a day to verify that the sites are able to run DIRAC payload jobs. LCG Grid sites, Vac [9] and cloud sites are monitored. The results are published on the web [10].

3. Small VOs and their use of GridPP DIRAC service

We support a dozen communities in our DIRAC instance. Some of them are “technical” VOs, like the gridpp VO which are mainly used by the Grid experts to test various services DIRAC offers or run the SAM tests, the others are groups of physicists or students from non-LHC experiments. Here we briefly describe some of them and their analyses done on our server. They fall into 3 groups, which:

- have done their entire analysis using DIRAC (CERN@school)
- have designed their own, experiment specific software and are using or planning to use DIRAC in their work (NA62)
- have just started assessing the use of DIRAC as their Grid gateway (COMET).

3.1. CERN@school VO

The CERN@school project makes High Energy Physics technology available to students at schools. It gives participants the opportunity to be part of a national collaboration of students, teachers and academics, analyzing data from particle physics detectors. CERN@school provides a Web portal, which allows access to data from the Langton Ultimate Cosmic ray Intensity Detector (LUCID) [11][12].

The VO successfully used the GridPP DIRAC service to perform their analysis. They made extensive use of the DIRAC file catalog including its metadata functionality. The metadata information was added to raw and generated data and was subsequently used in the analysis.

3.2. NA62

NA62 [13] is a CERN based fixed target experiment designed to study rare K meson decays.

The Monte Carlo (MC) simulation for the experiment is performed at NA62 enabled sites in the UK and Italy. The data are then copied to the storage elements at CERN and RAL. In order to achieve this a custom made job submission system has been developed; this system allows shift crews to submit and monitor job progress and data transfers. The job submission can be done using either the EMI WMS or GridPP DIRAC service. After a job is complete, a request to the web service based file transfer controller is made. This decouples job submission tasks from the file transfer tasks, so file copies are performed asynchronously. NA62 are currently investigating the possibility to use the DIRAC Data Management System (DMS) which also offers an asynchronous file transfer mechanism, ideally using the EMI File Transfer Service (FTS) system as a backend.

3.3. COMET

COMET is an experiment searching for neutrino-less μ conversions at J-PARC [14]. It is expected to produce data sets of the order of 20TB. Monte Carlo production will take place at a number of sites globally and the results will need to be transferred to dedicated storage sites at other locations. The COMET experiment plans to use DIRAC to manage data sets (both real data and MC) by providing a universal Grid-accessible file catalog. Once the file catalog

is operational then DIRAC will be considered for use as the job submission framework for both MC production and user analysis.

4. Conclusions

The GridPP DIRAC instance deployed at Imperial College London is a viable substitute for the EMI-WMS and LFC services for small experiments. Both the DIRAC workload and data management systems offer an enhanced set of features which are likely to be of specific use to small VOs.

We invite any interested communities to browse our documentation wiki [15].

References

- [1] D. Britton et al. “GridPP: the UK Grid for particle physics” *Phil. Trans. R. Soc. A* June 28, 2009 367 pp 2447-2457
- [2] EMI Middleware Initiative - Products: <http://www.eu-emi.eu/products> [Retrieved: 2015-05-15]
- [3] A Tsaregorodtsev et al, “DIRAC: a community grid solution”, *J. Phys.: Conf. Ser.* 119 062048, 2008
- [4] Belle 2: <http://belle2.kek.jp> [Retrieved: 2015-05-15]
- [5] BESIII Experiment: <http://bes3.ihep.ac.cn> [Retrieved: 2015-05-15]
- [6] ILC - International linear collider: <https://www.linearcollider.org/ILC> [Retrieved: 2015-05-15]
- [7] The GridPP DIRAC project: Implementation of a multi-VO DIRAC service. These proceedings.
- [8] A McNab, “Vac and Vcycle”, preGDB, 9 Sep 2014, CERN
- [9] A. McNab, F. Stagni and M Ubeda Garcia “Running Jobs in the Vacuum” *Journal of Physics: Conference Series* 513 (2014) 03206
- [10] GridPP DIRAC SAM Tests: <http://www.gridpp.ac.uk/php/gridpp-dirac-sam.php?action=view> [Retrieved: 2015-05-15]
- [11] T. Whyntie & M. A. Harrison, “Simulation and analysis of the LUCID experiment in the Low Earth Orbit radiation environment”, *J. Phys Conf. Ser.* 513 022038 (2014) - proceedings of CHEP 2013
- [12] T. Whyntie & B. Parker, “Investigating the inverse square law with the Timepix hybrid silicon pixel detector: a CERN@school demonstration experiment”, *Phys. Educ.* 48 344 (2013)
- [13] The NA62 Experiment: Prospects for the $K^+ \rightarrow \pi^+ \nu \bar{\nu}$ Measurement. Proceedings of Science PoS (KAON13) 032.
- [14] COMET technical design report: http://comet.kek.jp/Documents_files/PAC-TDR-2014/PAC-Review-20141110.pdf [Retrieved: 2015-05-15]
- [15] DIRAC - GridPP Wiki: <https://www.gridpp.ac.uk/wiki/Dirac> [Retrieved: 2015-05-15]