Journal of Physics: Conference Series

PAPER • OPEN ACCESS

GridPP DIRAC: Supporting non-LHC VOs on LHC centric resources

To cite this article: D Bauer and S Fayer 2017 *J. Phys.: Conf. Ser.* **898** 052003

IOP Conf. Series: Journal of Physics: Conf. Series **898** (2017) 052003 doi:10.1088/1742-6596/898/5/052003

GridPP DIRAC: Supporting non-LHC VOs on LHC centric resources

D Bauer and S Fayer

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

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

Abstract.

To allow non-LHC communities access to the primarily LHC dominated resources of the grid, the GridPP consortium in the UK maintains a multi-VO DIRAC service for this user-base.

After an extensive testing phase, this service has been in production for the last two years and has been fully integrated into the user communities' workflows. We report on the approaches taken by the VOs and the insights gained from these.

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.

The DIRAC project (Distributed Infrastructure with Remote Agent Control) [2] was originally designed as a workload management system for the LHCb experiment [3]. It has since been extended to allow support for multiple VOs on a single DIRAC server. It is also used by other collaborations including Belle2 [4], BESIII [5], ILC [6] and many others.

2. GridPP DIRAC

The DIRAC workload management system (see 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. DIRAC also provides a file catalogue (DFC) with metadata support. The GridPP DIRAC [7] instance is a standard multi-VO DIRAC installation with an additional automatic configuration module. It is monitored by both an internal Nagios server and a series of automated job submission tests developed by GridPP [8][9].

The GridPP DIRAC instance can be accessed via a number of different client interfaces. Apart from direct access with the DIRAC command line tools the most popular options are the DIRAC Python API and Ganga [10], a front-end for job definition.

All VOs currently use CVMFS [11] to distribute their software, however this is not a prerequisite for using the GridPP DIRAC instance.

3. User Communities

We currently support 16 VOs with 6 of them using the GridPP DIRAC service for their primary computing provision (see figure 2). The VOs are small (<100 users), typically have

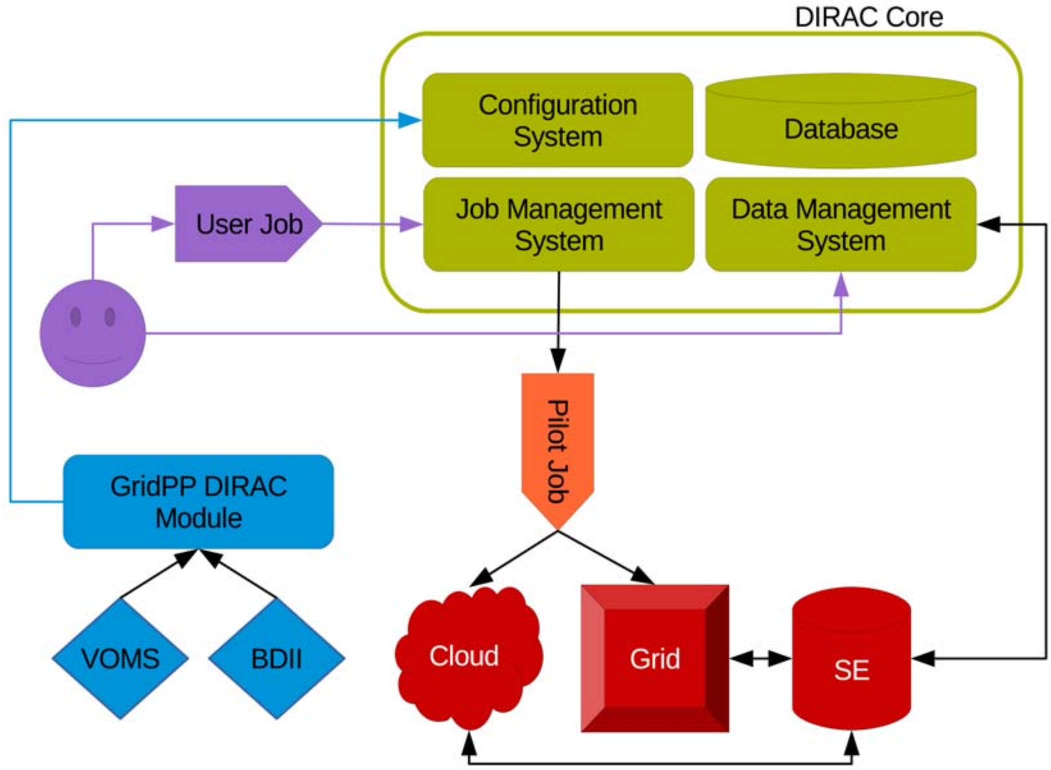

Figure 1. Schematic of the DIRAC framework

UK involvement and have no dedicated computing support. Their members are typically familiar with the concept of a batch system, but not necessarily with the concept of distributed computing. Three of the currently active VOs were completely new to the grid.

In the following section we review the experiment specific approaches to integrating grid resources using the GridPP DIRAC framework.

4. Small VOs and their use of GridPP DIRAC service

4.1. Large Synoptic Survey Telescope

The Large Synoptic Survey Telescope (LSST) [12] is currently being constructed in Chile and is expected to generate 200PB of data over its 10 year lifetime. The current UK contribution is running software for shape classification of galaxies on the data taken by a predecessor, the Dark Energy Survey [13]. The data comprise 100 million galaxies divided into 30,000 files. The classification of each galaxy is an independent measurement which uses 10-20s of CPU time. All of the data for a specific galaxy is contained within a single data file. This data layout and access scheme is very similar to most particle physics experiments.

The UK LSST collaboration had no previous experience in grid computing. They settled on using Ganga for job submission and the DIRAC file catalogue for data indexing. It took around 40 FTE days of work to complete the initial set-up and run through a full workflow.

4.2. COMET

COMET [14] is a fixed-target experiment at the J-PARC accelerator complex looking for coherent μ to e transitions. The experiment is still under construction so their current use of the provided computing resources is for simulation. The detector simulation software is

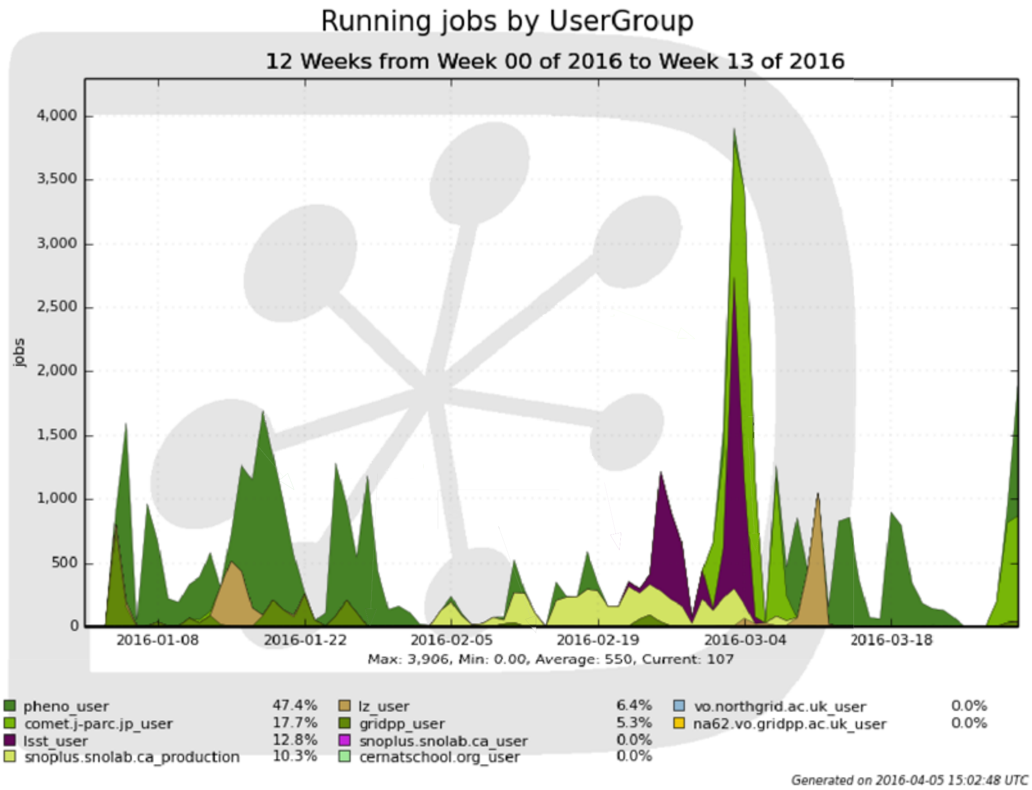

Figure 2. GridPP DIRAC usage by VO. This shows the typical small VO usage pattern, submitting jobs in bursts rather than making continuous usage of the Grid.

a framework based around GEANT4 [15] like many other particle physics experiments. Using their pre-existing Python expertise they developed a customised job submission framework based around the DIRAC API.

4.3. Lux Zeplin

Lux Zeplin (LZ) [16] is a dark matter search experiment based in the USA. Another experiment that is still under construction, it mainly runs GEANT4 based detector simulations. The output ROOT files are stored on a central grid-based storage element and indexed in the DIRAC file catalogue. Up to now the simulations have been co-ordinated with experiment specific python scripts. However they have just brought a new ganga-based custom job submission system online; this allows users to request production of datasets and allows managers to approve and start them via a web interface. This custom web interface was developed with computing expertise from outside of the VO supplied by the Imperial College HEP group. The development time was approximately one FTE month.

4.4. gridpp (Umbrella VO)

The GridPP VO is an umbrella VO for UK researchers whose projects are not covered by an existing VO but for whom the overhead of commissioning an experiment specific VO is too great compared to the amount of resources required.

An example experiment using the GridPP VO is the Geant Human Oncology Simulation Tool (GHOST) [17]. This project only required very basic job submission and management functionality and therefore the DIRAC command line tools were sufficient. The GridPP VO is also used for systematic infrastructure testing using the DIRAC API.

4.5. pheno
The pheno [18] VO is run by the phenomenology group based at Durham university in the UK. They use Ganga to submit home-grown Monte Carlo jobs via DIRAC. As they had previous experience with grid computing, they required very little computing support.

4.6. SNO+
SNO+ [19] are a neutrino experiment based in Canada's SNOLAB. They are an established experiment who have been using the grid for a number of years and are now migrating from the traditional WMS-based submission method. The workflows run are Monte Carlo production based around GEANT4 and experiment specific software submitted using Ganga as a frontend; they also submit a limited amount of user analysis work using the same method. They are using the classic UMD LFC [20] to index their data which is stored on a number of grid SEs, primarily around the UK. They were also the first experiment to introduce non-British sites into the GridPP DIRAC system, including some using the HTCondorCE services.

5. Observations and Conclusions
DIRAC was conceived as a single VO framework and not all features are truly multi-VO, such as the listing of available sites and enabling sites on a VO-by-VO basis. Feature requests by users (e.g. HTCondorCE support) can also take a long time to be implemented. Both of these are primarily caused by a lack of manpower rather than technical issues. Users have also remarked on the particular unhelpfulness of error messages, however this is not necessarily due to the DIRAC middleware but rather the error support in the underlying tools and is probably beyond the scope of this project to fix.

Having supported several small VOs with getting started on DIRAC, it seems that in most cases several weeks worth of expert effort is required to reach the level of a full production service. This is a reasonable amount of time given the complexity of the system and the diversity of the requirements of the VOs. Given proper expectation management this does not prove to be an insurmountable problem. Experience has shown that there are advantages in steering the VOs towards a couple of proven approaches, even if these are insufficiently documented.

In conclusion, we have successfully run a multi-VO DIRAC server in the UK for two years which has proven useful to allow small VOs access to LHC-centric resources.

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] A Tsaregorodtsev et al, "DIRAC: a community grid solution", J. Phys.: Conf. Ser. 119 062048, 2008
[3] LHCb Collaboration: `http://lhcb.web.cern.ch/lhcb/` [Retrieved: 2016-11-17]
[4] Belle 2: `https://www.belle2.org/` [Retrieved: 2016-11-16]
[5] BESIII Experiment: `http://bes3.ihep.ac.cn/` [Retrieved: 2016-11-16]
[6] ILC - International linear collider: `https://www.linearcollider.org/ILC` [Retrieved: 2016-11-16]
[7] GridPP DIRAC: `https://dirac.gridpp.ac.uk/` [Retrieved: 2016-11-16]
[8] GridPP DIRAC SAM Tests: `http://www.gridpp.ac.uk/php/gridpp-dirac-sam.php?action=view` [Retrieved: 2016-11-16]
[9] DIRAC Test Results: `http://pprc.qmul.ac.uk/~lloyd/gridpp/gridtests/diractest.html` [Retrieved: 2016-11-16]
[10] Ganga: `https://ganga.web.cern.ch/ganga` [Retrieved: 2016-11-16]
[11] CernVM File System: `https://cernvm.cern.ch/portal/filesystem` [Retrieved: 2016-11-16]
[12] About LSST: `https://www.lsst.org/about` [Retrieved: 2016-11-16]
[13] The Dark Energy Survey: `https://www.darkenergysurvey.org` [Retrieved: 2016-11-17]
[14] COMET: `http://comet.kek.jp/Introduction.html` [Retrieved: 2016-11-16]

[15] GEANT4: https://geant4.web.cern.ch/geant4 [Retrieved: 2016-11-16]
[16] The LZ Dark Matter Experiment: http://lz.lbl.gov [Retrieved: 2016-11-16]
[17] GHOST Project: http://www.comprt.org/research/ghost-project [Retrieved: 2016-11-16]
[18] PhenoGrid: https://www.phenogrid.dur.ac.uk/ [Retrieved: 2016-11-16]
[19] SNO+: https://www.snolab.ca/science/experiments/snoplus [Retrieved: 2016-11-16]
[20] EGI Software Repository: http://repository.egi.eu/ [Retrieved: 2016-11-16]