@GridPP @twhyntie

The GridUser Toolkit: Notes from the GridPP New User Engagement Programme

T. Whyntie*, †

* *Queen Mary University of London*; † *Langton Star Centre*

#GridPP34, QMUL

Thursday 30th April 2015

GridPP
UK Computing for Particle Physics

Queen Mary
University of London

Overview of the talk

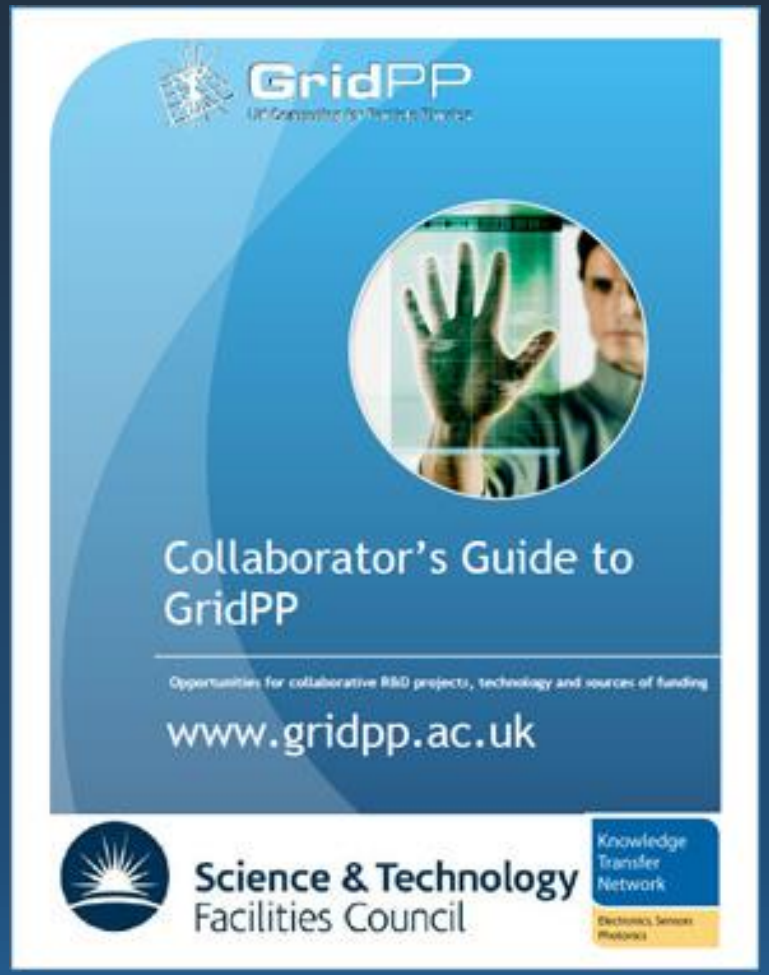
- Introduction
  - *The goals of, and motivation behind, the New User Engagement Programme.*
- Barriers to grid access
  - *The technical and non-technical barriers faced by users.*
- The GridUser Toolkit
  - *DIRAC, CVMFS, CERN VMs - and the documentation and user support for them.*
- Work to date and outlook
- Thanks and acknowledgments

Introduction

The goals of, and motivation behind, the New User Engagement Programme

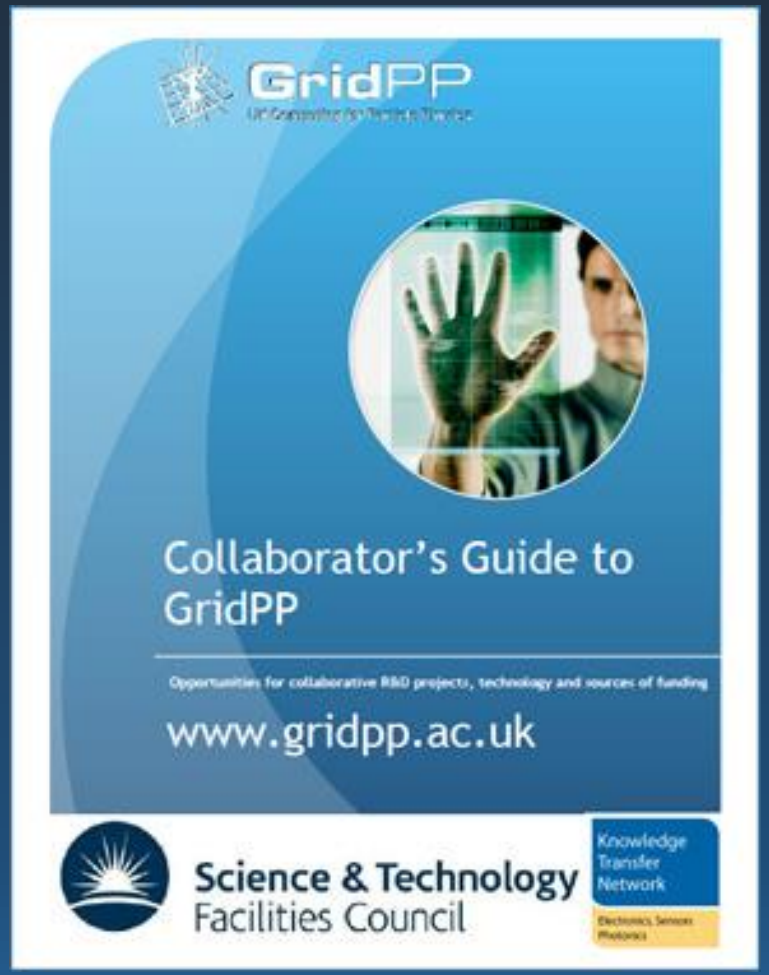
Goals

- We would like to make it as easy as possible to get onto the grid.
- Endpoint:
  - *A new user should be able to setup and run their own grid workflow;*
  - *Submit and monitor jobs, upload, process and store data, retrieve useful output.*
- Caveat:
  - *We shouldn't provide users solutions…*
  - *…but the tools needed to make their own.*
  - *(We can provide examples though!)*



Motivation

- Historic: make the grid available to all:
  - *GridPP's commitment – 10% of resources available to non-LHC users;*
- Engage new user communities:
  - *Non-HEP science; interdisciplinary work;*
  - *SMEs – industry partnerships;*
  - *Schools and hard-to-reach audiences;*
  - *"Pathways to Impact" – RCUK.*
- New partners for GridPP5 and beyond?



Barriers to grid access

The technical and non-technical barriers faced by users

Removing the barriers to grid access

- The grid is a very powerful resource:
  - *Developed **by** and **for** skilled users.*
  - *All have benefited from the training and resources provided by STFC et al.*
  - *New users will need help to overcome technical and non-technical hurdles.*
- Constraints:
  - *Only 0.5 FTE for "Dissemination and Engagement"; ad-hoc support at Tier-2s;*
  - *Can't provide dedicated user support;*
  - *Can't provide a dedicated "user space" (e.g. a cluster for new users).*

Technical Barriers

- T1) **The User Interface (UI)**: how the user interacts with the grid.
  - *T1.1) **The user's system**: local machine or via (university) cluster?*
  - *T1.2) **The grid UI**: the software used to interact with the grid.*
    - T1.2.1) Accessing a Workload Management System (WMS);
    - T1.2.2) Uploading, managing and retrieving data.
- T2) Software deployment
  - *T2.1) **Preparing custom software for deployment to the grid**: users will have their own software frameworks used in their workflows – and will need to be able to run these on the grid.*
  - *T2.2) **Installing software on grid worker nodes**: once prepared, this software needs to be made available to worker nodes on enabled sites.*

Non-Technical Barriers

- N1) Getting a grid certificate: this takes time and coordination (human interaction). Unavoidable but necessary for security purposes etc.
- N2) **Virtual Organisations (VOs):**
  - *N2.1) Joining a VO: if a suitable VO already exists. If not…*
  - *N2.2) Creating a VO: for new user communities to exploit grid resources.*
  - *N2.3) Enabling a VO at a given (GridPP) site:*
- N3) **Service Level Agreements (SLAs)**: determining the requirements of new users, ensuring they can be met, and how they evolve over time.
- N4) **Documentation and user support**: making sure (new) users can achieve the goal of implementing and running their workflow(s).

The GridUser Toolkit

DIRAC, CVMFS, CERN VMs - and the documentation and user support for them

Distributed Infrastructure with Remote Agent Control

A software framework for distributed computing with grid resouces.

See http://diracgrid.org/

DIRAC

OVERVIEW

CernVM File System
A network file system for delivering experiment software in a scalable, fast, reliable way via http.

See the website here.

CernVM-FS

Distributed Infrastructure with Remote Agent Control
A software framework for distributed computing with grid resouces.

See http://diracgrid.org/

DIRAC

OVERVIEW

CernVM File System
A network file system for delivering experiment software in a scalable, fast, reliable way via http.

See the website here.

Distributed Infrastructure with Remote Agent Control
A software framework for distributed computing with grid resouces.

See http://diracgrid.org/

CernVM-FS

DIRAC

CernVM

OVERVIEW

CernVM
CernVM is a baseline Virtual Software Appliance for the participants of CERN LHC experiments.

IMPLEMENTATION

RAL hosts a CernVM-FS Stratum 0 for non-LHC VO software repositories.

See the wiki page.

GridPP DIRAC server hosted and supported by Imperial College.

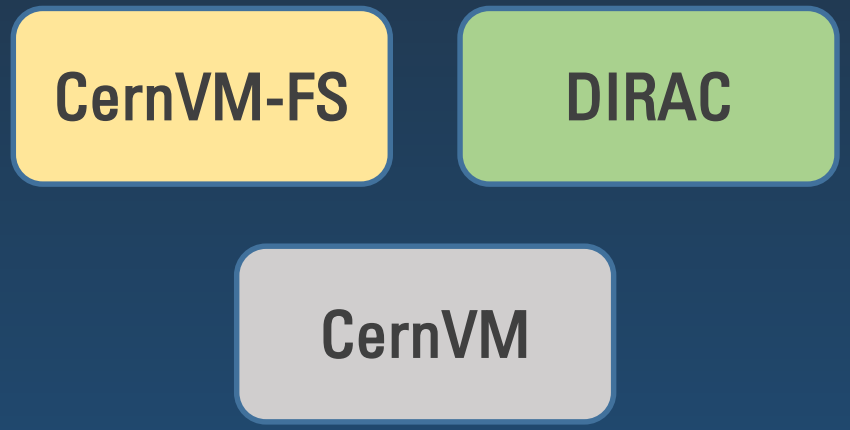
CHEP '15 R. Currie et al. talk here.

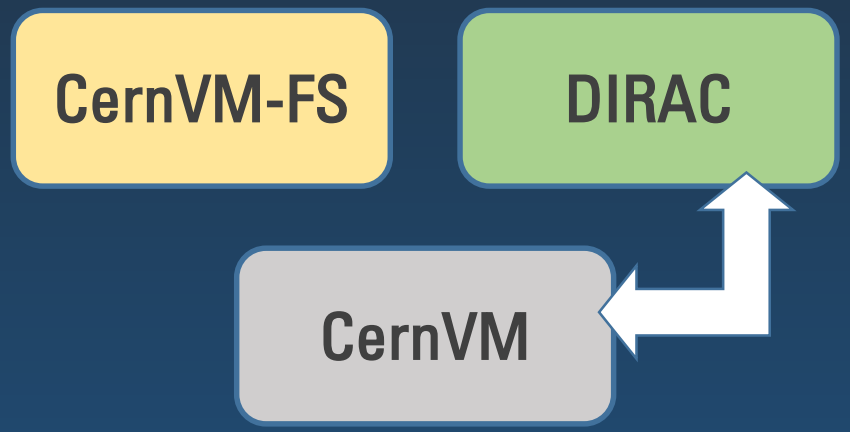
CernVM-FS

DIRAC

CernVM

The CernVM group has developed a generic SLC6 VM that can be contextualised by users. Users can run this image as a guest on their own local machine host system via e.g. VirtualBox.


CernVM-FS
DIRAC
CernVM

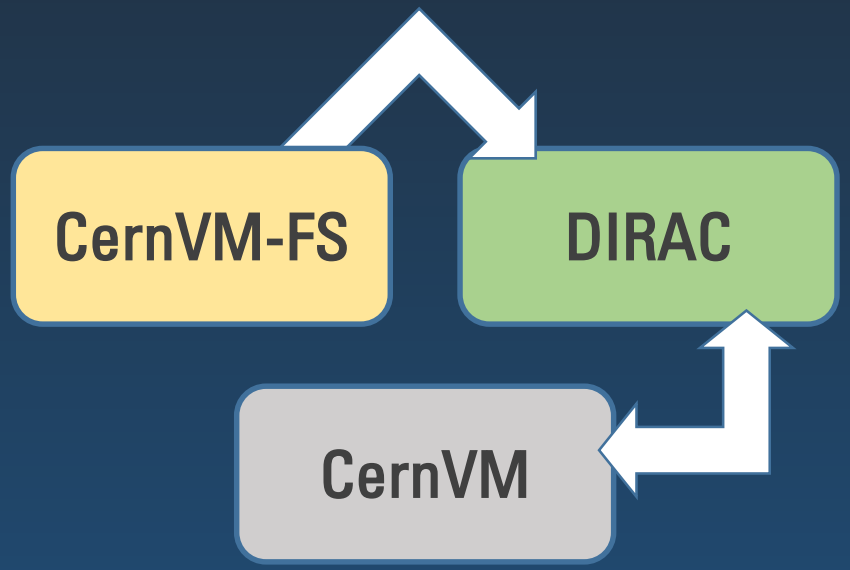


The DIRAC UI can be installed and run by a user from a suitably contextualised CernVM.

The DIRAC UI also contains the DIRAC Python API libraries.

(Ganga can also be installed and run from the same CernVM.)

User software in (custom) CernVM-FS repositories can run in jobs managed by the Imperial DIRAC server.



The DIRAC UI can be installed and run by a user from a suitably contextualised CernVM.

The DIRAC UI also contains the DIRAC Python API libraries.

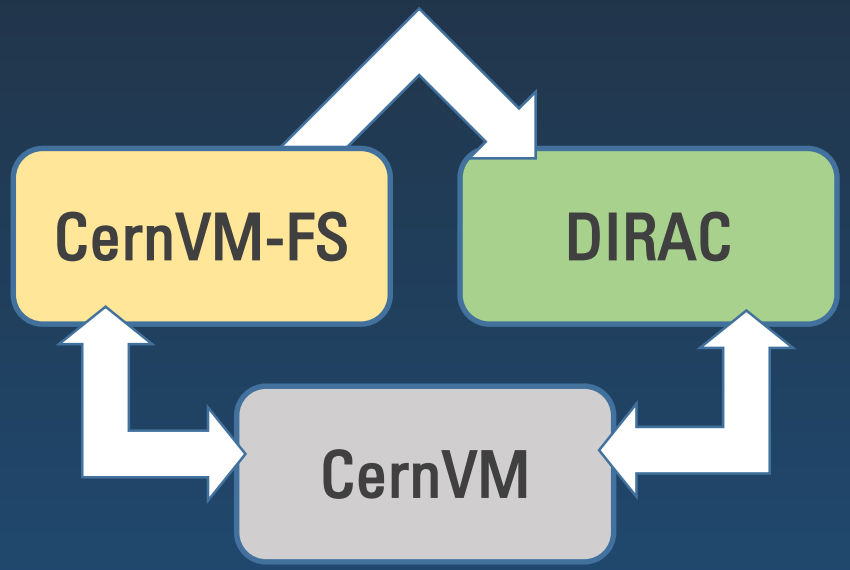
(Ganga can also be installed and run from the same CernVM.)

User software in (custom) CernVM-FS repositories can run in jobs managed by the Imperial DIRAC server.

Users can build their software (executables and libraries) on their local GridPP CernVM ready for deployment to the grid.

Users can upload their software to their custom CernVM-FS repository from a GridPP CernVM using the gsi* tools in the repository `/cvmfs/grid.cern.ch`

Custom CernVM-FS repository software can be accessed from a contextualised GridPP CernVM.



The DIRAC UI can be installed and run by a user from a suitably contextualised CernVM.

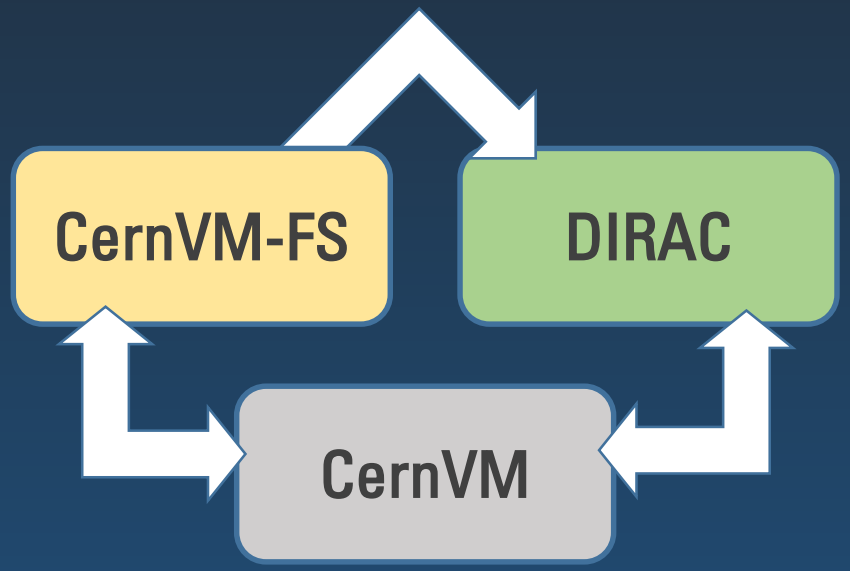
The DIRAC UI also contains the DIRAC Python API libraries.

(Ganga can also be installed and run from the same CernVM.)

User software in (custom) CernVM-FS repositories can run in jobs managed by the Imperial DIRAC server.

Users can build their software (executables and libraries) on their local GridPP CernVM ready for deployment to the grid.

Users can upload their software to their custom CernVM-FS repository from a GridPP CernVM using the gsi* tools in the repository `/cvmfs/grid.cern.ch`



The DIRAC UI can be installed and run by a user from a suitably contextualised CernVM.

The DIRAC UI also contains the DIRAC Python API libraries.

(Ganga can also be installed and run from the same CernVM.)

Custom CernVM-FS repository software can be accessed from a contextualised GridPP CernVM.

This, for me, is the game-changer.

This essentially gives the user a **grid worker node** that can be used immediately – *without a grid certificate*.

The DIRAC, Ganga and other Useful Software can be deployed via CernVM-FS for local CernVMs.

Removing barriers – DIRAC

- DIRAC provides a standardised, well-supported user interface for non-LHC VOs (i.e. for those who cannot develop their own) – T1.2.
  - *Job submission and management via the DIRAC WMS - T1.2.1.*
  - *Data management via the DIRAC File Catalog (DFC) – T1.2.2.*
  - *Interface via the command line, web portal or Python API.*
  - *Also interfaces with Ganga (see previous talk).*
- The Imperial DIRAC server is configured for multiple VOs – N2.3.
  - *Removes the burden of arranging this for the end user…*
  - *…but does need to be done by the Imperial team (case for automation etc.).*

Removing barriers – CernVM-FS

- Installing software on grid worker nodes becomes trivial – T2.2.
  - *Once a site is configured to support a VO's CernVM-FS repository (N2.3).*
- Installing software on a user's CernVM system becomes trivial – T1.1.
  - *When the CernVM is contextualised appropriately, DIRAC/Ganga/etc. can be accessed – even without a grid certificate (N1).*

Removing barriers - CernVM

- Users have instant access to a standardised, grid-ready system – T1.1.
  - *Not even a CERN account is required to use CernVM images – N1.*
  - *GridPP is developing a custom context for use with GridPP resources.*
  - *Solves the problem of delays caused by N1, N2 and N3 issues – something to play with "out of the (Virtual)Box" (other host platforms are available).*
- In the worker node model, users can prepare their workflow software on a CernVM – T2.1 – and upload it to the RAL Stratum 0 – T2.2.
  - *In the Cloud model, users can prepare a complete (Cern)VM ready for running on the grid via Vcycle etc.*
- Removes the need to document and support multiple user systems by providing a single, standardised GridPP CernVM – N4.

Documentation and User Support

Guiding principle: *minimise human interaction*.

- "Unit testing":
  - *As we all do when writing software, incorporate tests for success for users;*
  - *Implemented as "New User Checklists" on the wiki at present.*
- Sharing code:
  - *Code should be made available to aid collaboration and share best practice;*
  - *We use GitHub – see http://github.com/GridPP*
  - *Python is the* de facto *scripting language – not least because of the DIRAC API.*
- Worked examples:
  - *Show, don't tell – give users examples they can follow that will work.*
  - *cernatschool.org used as the technology demonstrator VO in this respect.*

Work to date and outlook

- It all works – we have demonstrated the proof of concept.
  - *CERN@school: essentially using DIRAC as a production service. Enabled at four Tier-2 sites.* `/cvmfs/cernatschool.gridpp.ac.uk` *repo in use both on the grid and on a GridPP CernVM context, as well as* `/cvmfs/grid.cern.ch`
  - *UCLan "GalDyn" (Galaxy Dynamics): using DIRAC and CernVM with test jobs. Software compiled and running on the GridPP CernVM. Next step – deploy with CernVM-FS and run their workflow on the grid.*
  - *LSST, LIGO, QMUL Proteomics: engaged (thanks Jeremy and Pete).*
- Further documentation required:
  - *CernVM-FS (for users and site admins);*
  - *Virtual Organisations and SLAs – policies and procedures.*
- Integration with Ganga…

Thanks and acknowledgments

- DIRAC: the Imperial team (see excellent CHEP ‘15 talks here and here).
- CernVM-FS: Catalin (RAL STFC)
- CernVM: Ewan for pointing out CernVM, Matt D. for `grid.cern.ch` hints.
- Regional VO managers: Alessandra, Daniela, Duncan.
- Tier-2 sysadmins: Sam S., Dan T., Steve J.

Thanks and acknowledgments

- DIRAC: the Imperial team (see excellent CHEP '15 talks here and here).
- CernVM-FS: Catalin (RAL STFC)
- CernVM: Ewan for pointing out CernVM, Matt D. for `grid.cern.ch` hints.
- Regional VO managers: Alessandra, Daniela, Duncan.
- Tier-2 sysadmins: Sam S., Dan T., Steve J.

*This is all the result of a collaboration of experts working together to leverage new technologies - **the essence of GridPP**.*

@GridPP @twhyntie

Thanks for listening – any question?

T. Whyntie*, †

* *Queen Mary University of London*; † *Langton Star Centre*

#GridPP34, QMUL

Thursday 30th April 2015

GridPP
UK Computing for Particle Physics

Queen Mary
University of London