The GridPP CernVM

T. Whyntie*

Queen Mary University of London

HEPSYSMAN Jan' 16

Friday 15th January 2016

GridPP
UK Computing for Particle Physics

Queen Mary
University of London

Overview of the talk

- CernVM – what is it?
- Why use the CernVM for the PENGUIN?
- How the GridPP CernVM can be used
- How do I get started?
- Thanks and acknowledgments

CernVM – what is it?

- From the website, https://cernvm.cern.ch :
  - *"CernVM is a baseline Virtual Software Appliance for the participants of CERN LHC experiments. The Appliance represents a complete, portable and easy to configure user environment for developing and running LHC data analysis locally and on institutional and commercial clouds (OpenStack, Amazon EC2, Google Compute Engine), independently of Operating System software and hardware platform (Linux, Windows, MacOS). The goal is to remove a need for the installation of the experiment software and to minimize the number of platforms (compiler-OS combinations) on which experiment software needs to be supported and tested."*

- What you get with the Virtual Machine (VM):
  - *An SL6 image that can be run on any host OS with suitable VM software;*
  - *Out-of-the-(Virtual)-box access to the CernVM-FS (a.k.a. CVMFS);*
  - *The option to contextualise via user- or experiment-defined **Context**.*

Why use the CernVM for the PENGUIN?

- PENGUIN = **P**rogramme for the **E**ngagement of **N**ew **G**ridPP **U**sers in **I**ndustry and **N**on-physics applications. (Obviously.)

- A consistent user interface/working environment:
  - *Not a technical requirement as such (though it does help with libraries, compiling software locally, etc.) – it's more about providing remote support to new users. If the VM is the same, the same instructions will apply and (more importantly) the errors should be reproducible.*
  - *Independent of the Guest OS (Windows, Mac, etc.), at the cost of needing enough memory to run a guest VM on their host machine.*

- No need for access to a university-based, grid-ready cluster:
  - *Removes a barrier for industry and other external users.*

- Provides instant access to the CernVM-FS (a.k.a. CVMFS):
  - *More on this later!*

GridPP
UK Computing for Particle Physics

Queen Mary
University of London

How the GridPP CernVM can be used

- Installing and running a GridPP CernVM:
  - *Download the CernVM image and run on the host machine's VM software;*
  - *Apply the GridPP CernVM context and pair the instance (see UserGuide).*
- Using the GridPP CernVM as a (sort-of) virtual grid node:
  - *Compiling and running user software in a grid node environment(?);*
  - *Check for extra libraries that need to be added to one's CVMFS repository;*
  - *All before you touch the grid (which should, in theory, save time…).*
- Installing the GridPP DIRAC User Interface:
  - *Relatively trivial to do – see the UserGuide for more information.*
- Accessing experiment/VO/utility software via CVMFS:
  - *e.g. GANGA, grid UI tools… even the DIRAC UI ultimately?*
  - *(HUGE BONUS: independent of grid activities, can access and run e.g. cernatschool.org software anywhere – useful for schools!)*
- Uploading software to the CernVM-FS via gsi tools.
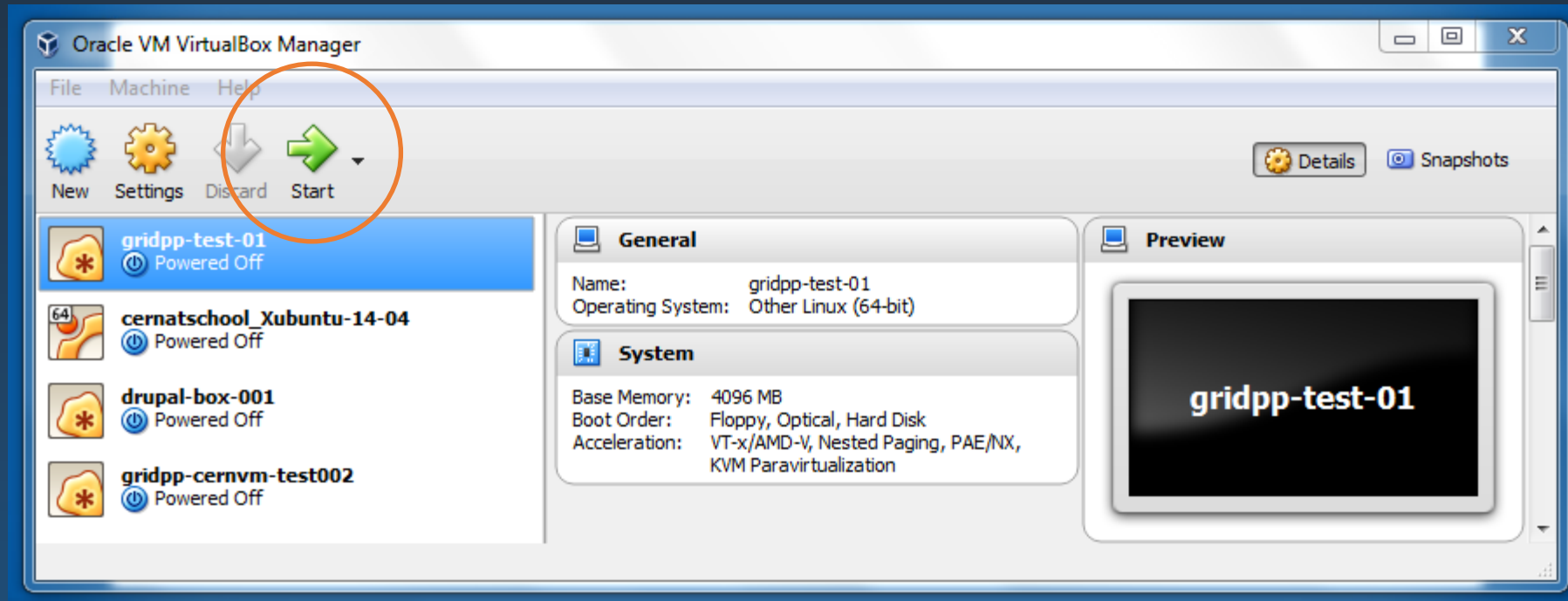
Starting a GridPP CernVM from the Oracle VM VirtualBox Manager
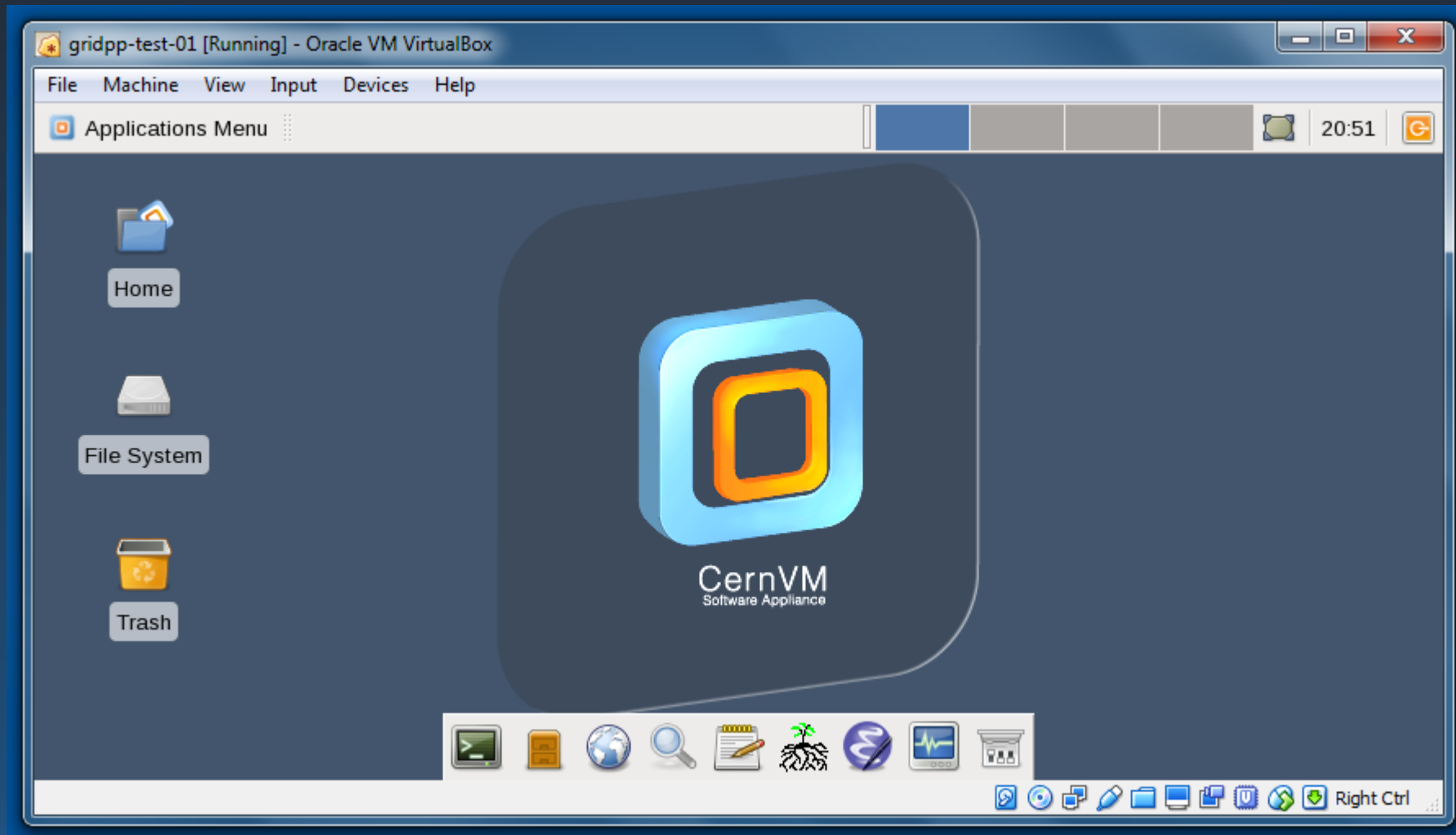
The GridPP CernVM in action (note the ROOT icon...!)
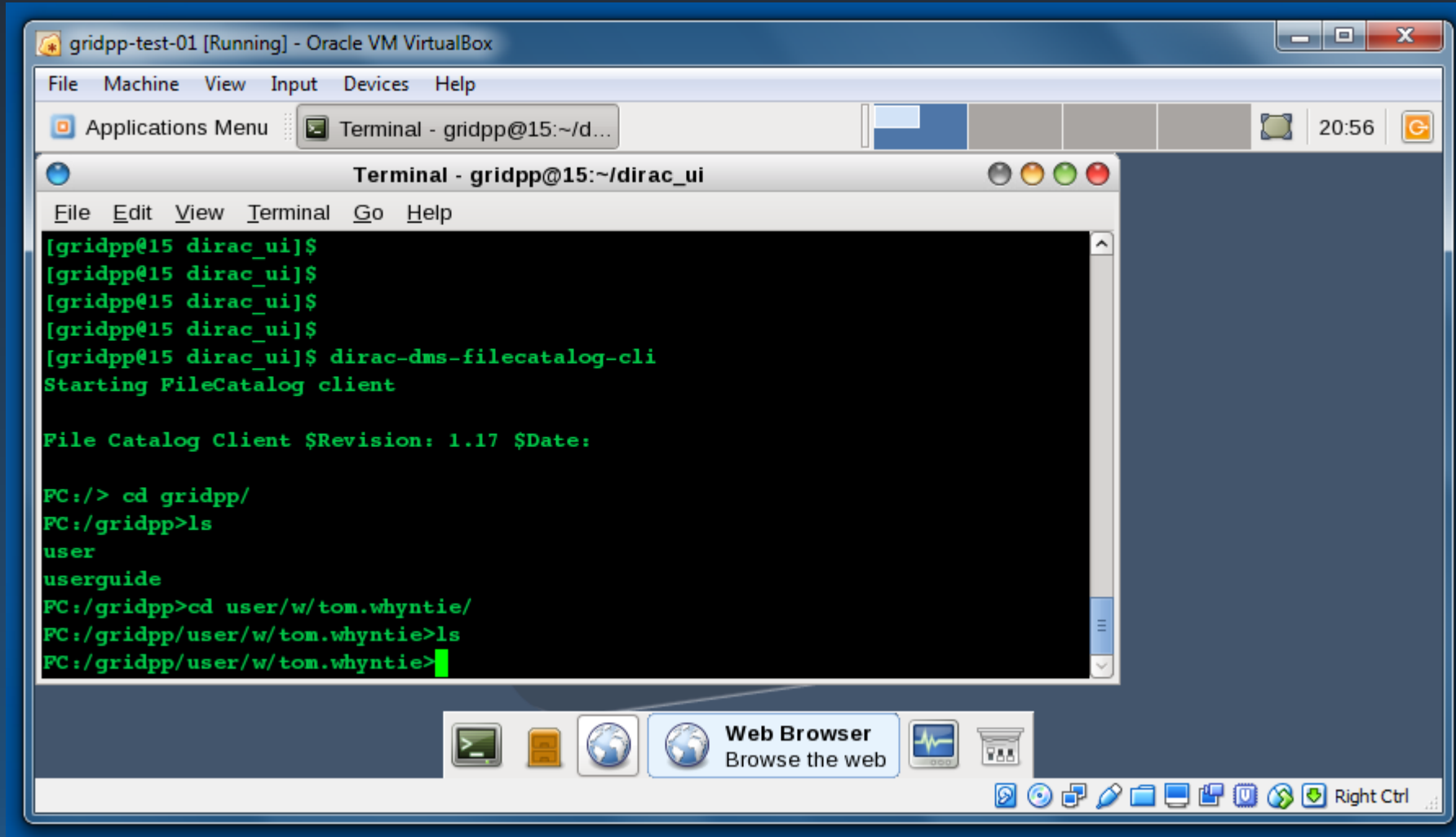
Using GridPP DIRAC via the CernVM (here, the DIRAC File Catalog Client Interface)

How do I get started?

- Consult the GridPP UserGuide!
  - *https://www.gridpp.ac.uk/userguide*
  - *Specifically, this part will take you through the CernVM setup process:*
    - https://www.gridpp.ac.uk/userguide/gridpp-cernvm/gridpp-cernvm.html
  - *And the GridPP DIRAC UI setup is covered here:*
    - https://www.gridpp.ac.uk/userguide/dirac-create-ui/dirac-create-ui.html

- Apologies for any out-of-date links/instructions on the Wiki…

Thanks and acknowledgments

- The GridPP DIRAC team (Imperial).
- CernVM and the CernVM-FS: Catalin (RAL STFC).
- Ewan for pointing out CernVM, Matt D. for `grid.cern.ch` hints.
- Steve J et al. – invaluable feedback on the UserGuide so far!
- The "Incubator VO" managers.

Backup slides

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

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

GridPP
UK Computing for Particle Physics

Queen Mary
University of London

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.
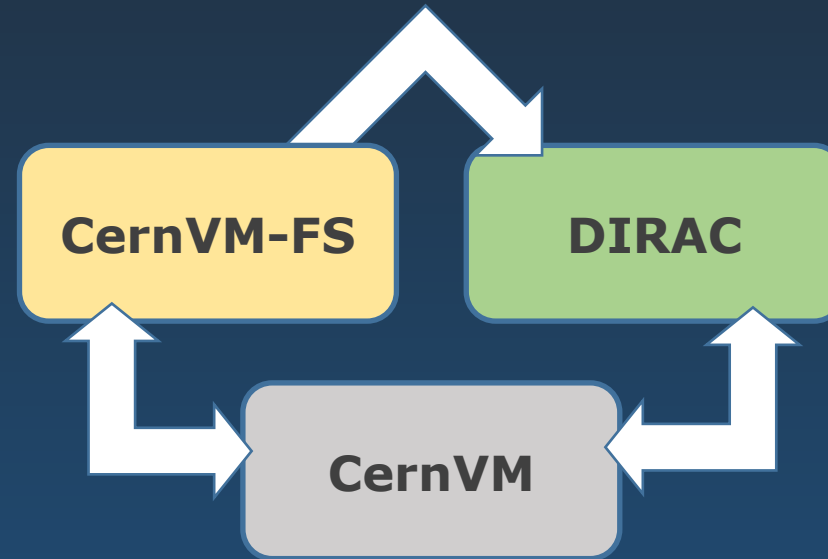
CernVM-FS

DIRAC

CernVM

The CernVM group has developed a generic SL6 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.

GridPP
UK Computing for Particle Physics

Queen Mary
University of London

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.

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.)

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.

GridPP
UK Computing for Particle Physics

Queen Mary
University of London