DIRAC Services

A. Tsaregorodtsev,
CPPM-IN2P3-CNRS

EGI Community Forum, Bari, 11 November 2015
Plan

- DIRAC Project brief overview
- DIRAC services
- DIRAC4EGI service
- Conclusions
DIRAC provides all the necessary components to build ad-hoc grid infrastructures **interconnecting** computing resources of different types, allowing **interoperability** and simplifying **interfaces**. This allows to speak about the DIRAC **interware**.
Interware: WMS

- Pilot based Workload Management
  - High user job efficiency
  - Suitable for usage with heterogeneous resources
  - Allowing application of community policies

- Pilot 2.0 framework
  - Modular for easy customizations for different environments, communities, etc
Pilots 2.0 framework

- **Modularity:** pilots are designed as a configurable sequence of commands

- **Extensions:** each community can provide custom commands for specific operations
  - Environment checks
  - Software installation
  - Monitoring reports
LHCb example: “pilots to fly in all the sky's”

still available from everywhere, as backup

Web Server

DIRAC pilot factory

DIRAC Jobs queue

DIRAC VMs factory

WLCG

setup interware

Get job to run

setup SW

DIRAC CE

setup interware

Get job to run

setup SW

HLT Farm

setup interware

Get job to run

setup SW

VAC

setup interware

Get job to run

setup SW

BOINC

setup interware

Get job to run

setup SW

CLOUD

setup interware

Get job to run

setup SW

CVMFS
Computing resources available via pilots

- **Grids**
  - gLite/EMI: EGI (CREAMComputingElement)
  - VDT: OSG (GlobusComputingElement)
  - ARC: NDGF (ARCComputingElement)

- **Standalone clusters**
  - Using SSH/GSISSH tunnel
  - LSF, BQS, SGE, PBS/Torque, Condor(G), ...
    - E.g. Yandex computing farm
  - HTCondor ComputingElement is in the test

- **HPC centers:**
  - OAR, SLURM, using DIRAC proxy servers to move data in and out

- **Clouds**
  - VM scheduler for EC2, OCCI, Nova, libcloud
    - Amazon, OpenNebula, OpenStack, CloudStack, Stratuslab
  - VAC, Vcycle

- **BOINC**
  - IDGF
  - Standalone, e.g. LHCb@HOME
- VM Scheduler developed for Belle Data Production System
  - Dynamic VM spawning driven by the user workload
  - Discarding VMs automatically when no more needed

- Support for various cloud technologies
  - Amazon, OCCI, OpenNebula, OpenStack, CloudStack, Stratuslab

- DIRAC is an active participant of the EGI FedCloud activities
  - FedCloud sites are accessible through the DIRAC4EGI service
Support for the Per User SubProxies (PUSP)

- To be in line with the EGI developments
- Used for the DIRAC tutorial
- Can be used for scientific gateway portal users

Each PUSP user is registered in person in DIRAC

- Per User policies (group membership)
  - E.g. File Catalog access rights
- Per User Accounting
Interfaces

- Command line clients
- REST interface
  - Useful to use as a WMS engine for Application portals
    - VIP, WS-PGrade, WeNMR. Etc
- Web Interface
  - Web Portal development framework
    - Tornado CMS, ExtJS GUI
    - Secured with X509 certificates
  - Desktop GUI paradigm
    - Natural for non-expert users
  - Support of most of the user tasks (jobs, data, monitoring, management)
Web Portal: example interfaces
DIRAC is aiming at providing an abstraction of a single computer for massive computational and data operations from the user perspective.

- Logical Computing and Storage elements (Hardware)
- Global logical name space (File System)
- Desktop-like GUI
Dedicated installations for the large user communities

- LHCb, Belle II, BES III, ILC, CTA

CTA MC Production run

MC production: stable operations for > 5 months

User analysis:
- constant over all the year
- high I/O
DIRAC services are provided by several National Grid Initiatives: France, Spain, Italy, UK, China, Romania, …

- Support for small communities
- Heavily used for training and evaluation purposes

Example: France-Grilles DIRAC service

- Hosted by the CC/IN2P3, Lyon
- Distributed administrator team
  - 5 participating universities
  - 15 VOs, ~100 registered users
- In production since May 2012
  - >12M jobs executed in the last year
    - At ~90 distinct sites
In production since 2014

Partners
- Operated by EGI
- Hosted by CYFRONET
- DIRAC Project providing software consultancy

10 Virtual Organizations
- enmr.eu
- Vlemed
- fedcloud.egi.eu
- training.egi.eu
- …

Usage
- > 6 million jobs processed in the last year
• Details of Cyfronet instance
  • DIRAC services split over 4 VMs + DB server, all with regular backups.
  • 1 TB disk space for I/O sandboxes
  • Web Portal address: dirac.egi.eu
  • Contact for interested groups: dirac@mailman.egi.eu
There is a clear need for services like DIRAC for an increasing number of communities with a low expertise in (distributed) computing and with high demands for computing resources.

Important goal is to lower the threshold for scientists of these communities.

- Training is one of the main purposes of all the infrastructures deploying DIRAC services.
- Examples of training projects:
  - vo.formations.idgrilles.fr (FG-DIRAC)
  - training.egi.eu (DIRAC4EGI)
    - The next DIRAC tutorial is at the EGI Community Forum, November
  - Distributed computing student courses
  - CERN@school project using the GridPP DIRAC service
- Assistance in porting application to the (DIRAC) grids
DIRAC is providing a framework to bring together various services and computing resources in a single coherent system.

From the user perspective the whole system is seen as a single computer with an intuitive (graphical) interface.

Multi-VO DIRAC services is an excellent way to open access to distributed computing resources for non-expert user communities.

http://diracgrid.org