Introduction

DIRAC Tutorial

Plan

- DIRAC Project
- DIRAC interware
- DIRAC as a Service
- Tutorial plan

Massive computations: HEP

- LHC experiments pioneered the massive use of computational grids
  - 10s of PBytes of data per year
  - 100s of thousands CPUs in 100s of centers
  - 10s GB/sec network transfers
  - 100s of users from 100s of institutions

- CERN Director General Rolf Heuer about the Higgs discovery:

"It was a global effort and it is a global success. The results today are only possible because of the extraordinary performance of the accelerators, including the infrastructure, the experiments, and the *Grid computing*."

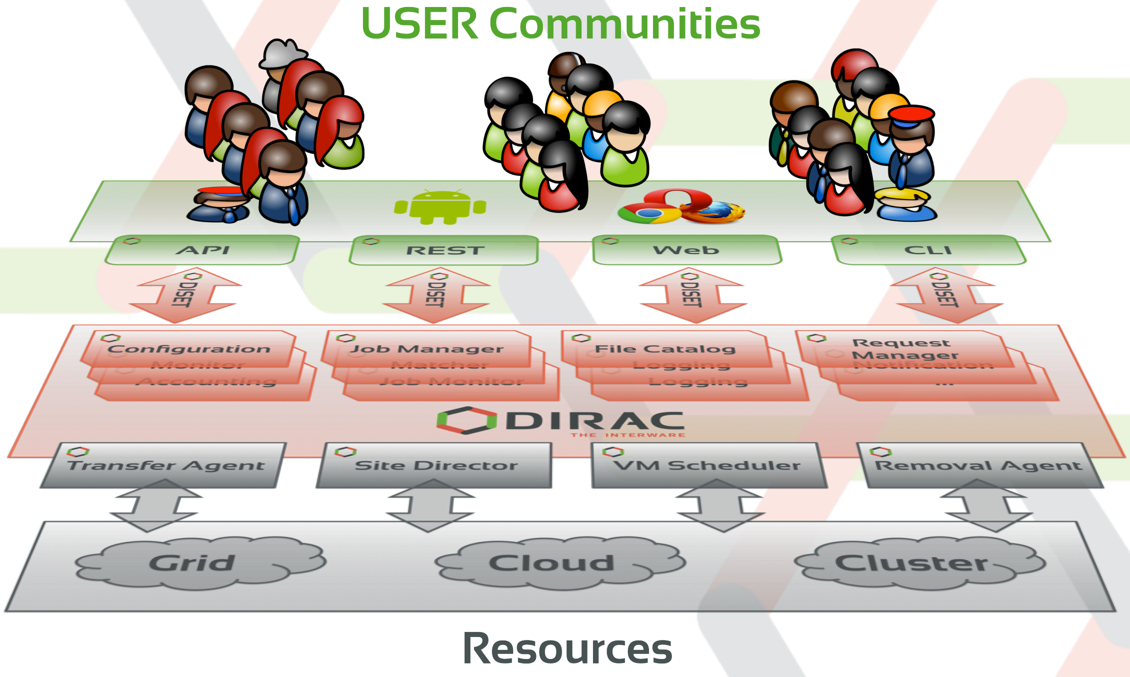
- Other domains are catching up quickly with the HEP experiments
  - Life sciences, earth sciences, astrophysics, social sciences, etc

Massive computations

- The computing expertise level in non-HEP scientific domains is relatively lower
  - Grouped around well known applications and scientific portals
  - Moving existing applications to run in distributed environments is still difficult
- Convenient tools for small research groups with no local gurus are clearly needed
- All LHC experiments developed their own middleware
  - PanDA, AliEn, glideIn WMS, PhEDEx, DIRAC, …
  - WMS with pilot jobs, intelligent data management, software distribution, …
- Experience of the LHC experiments in using distributed computing infrastructures should now be made available for non-LHC user communities

Interware

- DIRAC provides all the necessary components to build ad-hoc distributed computing infrastructures interconnecting resources of different types, allowing interoperability and simplifying interfaces. This allows to speak about the DIRAC ***interware***.



Towards general purpose middleware

- Several new experiments expressed interest in using this software relying on its proven functionality

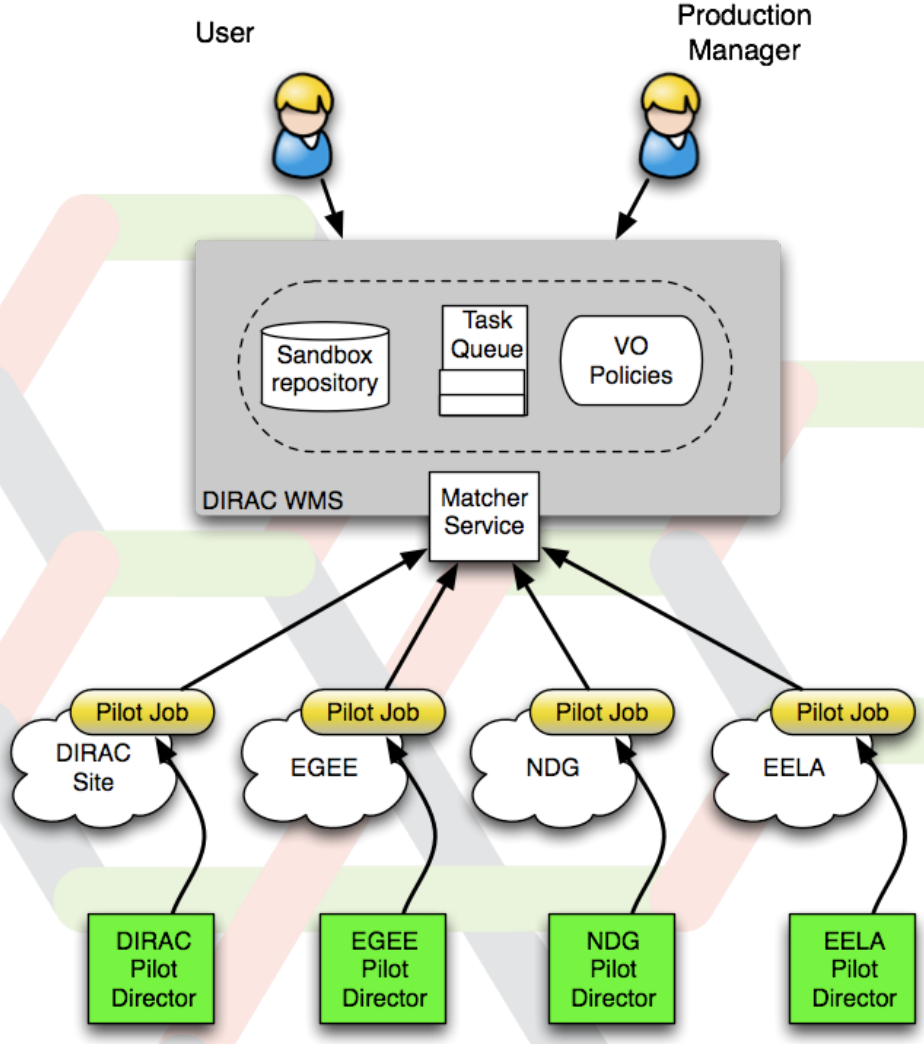
- DIRAC Consortium founded in 2013
  - CNRS, University of Barcelona and CERN
    - University of Montpellier and IHEP, Beijing joint later
  - Development, maintenance and promotion of the DIRAC software
  - Consortium is the software copyright holder (GPL v3)

- This work made it possible to offer general-purpose DIRAC services to any scientific community

Workload Management

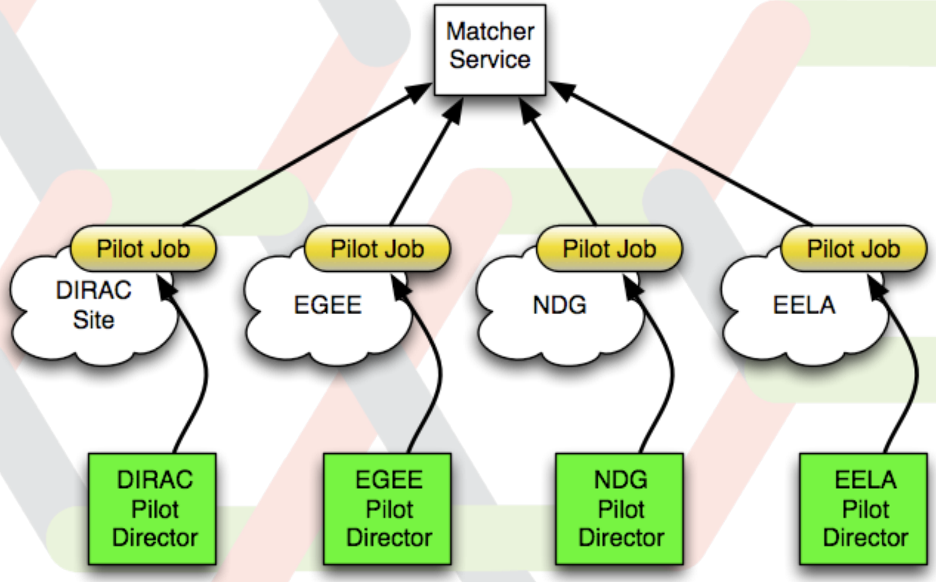
DIRAC WMS

- Jobs are submitted to the DIRAC Central Task Queue with credentials of their owner (VOMS proxy)
- Pilot Jobs are submitted by specific Directors to a Grid WMS with credentials of a user with a special Pilot role
- The Pilot Job fetches the user job and the job owner's proxy
- The User Job is executed with its owner's proxy used to access SE, catalogs, etc



WMS: using heterogeneous resources

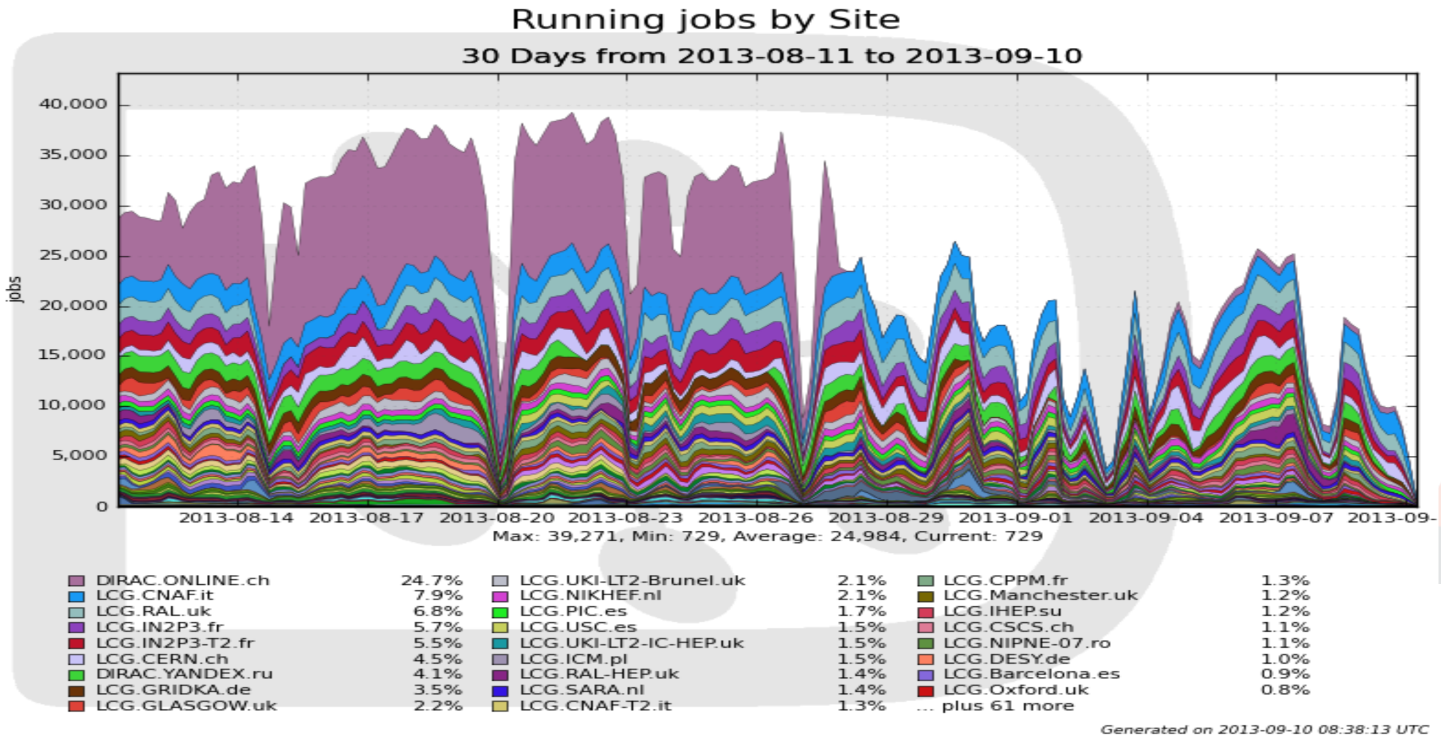
- Including resources in different grids and standalone clusters is simple with Pilot Jobs
  - Needs a specialized Pilot Director per resource type
  - Demonstrated with various grid sites, clouds, etc
  - Users just see new sites appearing in the job monitoring



Resources provisioning

- DIRAC middleware facilitates access to various types of resources
  - gLite and ARC middleware based grids (EGI, NDGF, etc)
  - Standalone clusters
    - Simple SSH accessible account is sufficient to include the site
  - Clouds (Amazon, OpenStack, OpenNebula, OCCI compliant)
    - Automatic virtual machine scheduling
  - Desktop Grid
    - Based on BOINC technology
    - Support for multiple platforms with virtualization

LHCb DIRAC performance



- DIRAC performance in production
  - Up to 50K concurrent jobs in ~120 distinct sites
  - 10 mid-range central servers hosting DIRAC services
  - Further optimizations to increase capacity are possible

Data Management

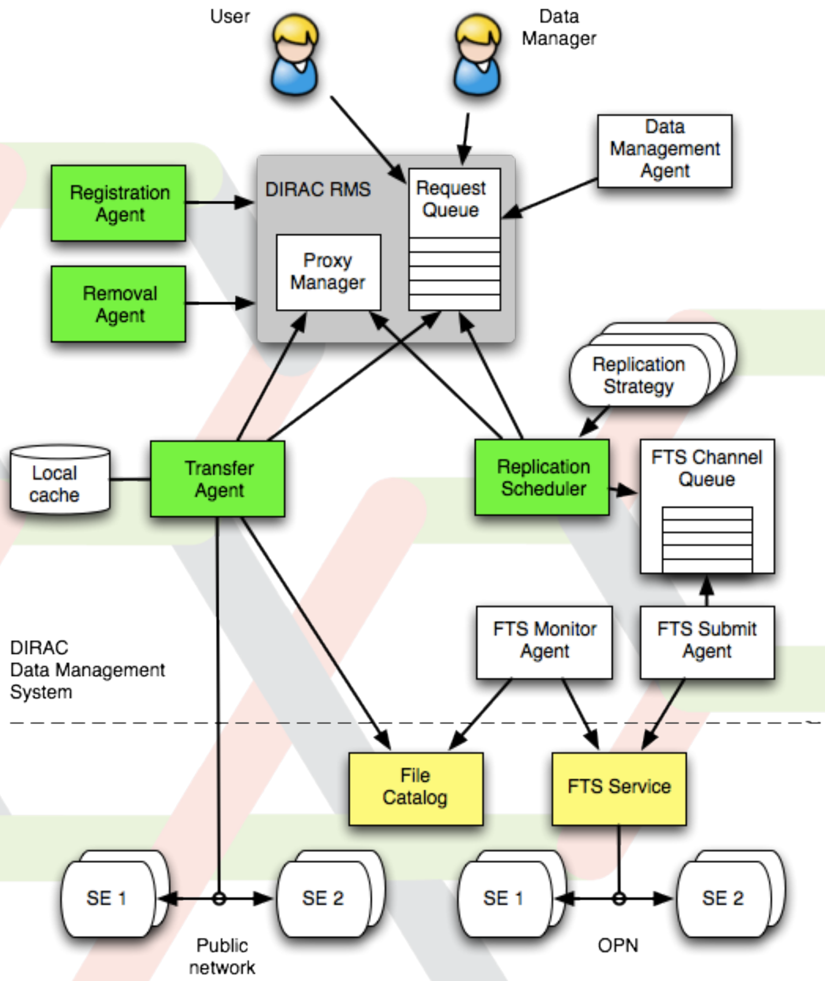
Data Management components

- Storage Elements
  - gLite/EGI Storage Elements
  - DIRAC Storage Elements
  - iRods Storage Elements
  - More Storage Elements can be included
    - (F,SF,HT,BBF)TP servers
    - S3, CDMI, WebDav

- File Catalogs
  - LCG File Catalog (LFC)
    - *de facto* standard grid catalog
  - DIRAC File Catalog
    - Support for the User Metadata
    - Support for data provenance
  - More Catalogs can be included
    - LHCb has developed several specific catalogs in the same framework
  - Different catalogs can be used together

Data Management services

- Based on the Request Management System
- Asynchronous data operations
  - transfers, registration, removal
- Two complementary replication mechanisms
  - Transfer Agent
    - user data
    - public network
  - FTS service
    - Production data
    - Private FTS OPN network
    - Smart pluggable replication strategies

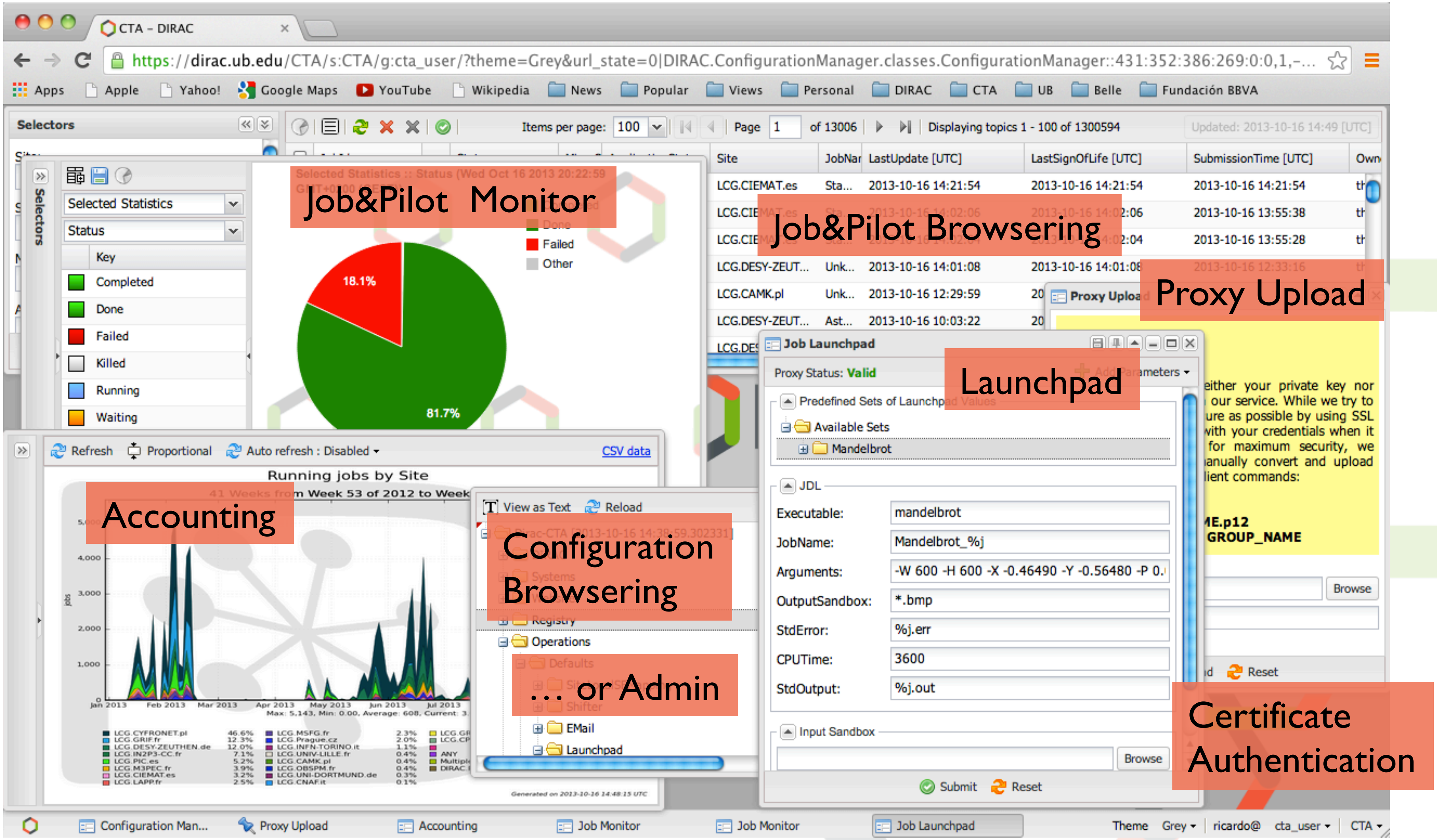


User Interfaces

DIRAC: Secure Web Portal

- Focus on the Web Portal as the main user tool for interactions with the grid
- Intuitive desktop application like interface
  - Ajax, Tornado, ExtJS Javascript library
- Monitoring and control of all activities
  - User job monitoring and manipulation
  - Data manipulation and downloads
  - DIRAC Systems configuration and management
- Secure access
  - Standard grid certificates
  - Fine grained authorization rules

Web Portal



Job&Pilot Monitor

Job&Pilot Browsering

Proxy Upload

Launchpad

Accounting

Configuration Browsering

… or Admin

Certificate Authentication

https://dirac.ub.edu/CTA/s:CTA/g:cta_user/?theme=Grey&url_state=0|DIRAC.ConfigurationManager.classes.ConfigurationManager::431:352:386:269:0:0,1,-...

Items per page: 100 | Page 1 of 13006 | Displaying topics 1 - 100 of 1300594 | Updated: 2013-10-16 14:49 [UTC]

Site	JobName	LastUpdate [UTC]	LastSignOfLife [UTC]	SubmissionTime [UTC]
LCG.CIEMAT.es	Sta...	2013-10-16 14:21:54	2013-10-16 14:21:54	2013-10-16 14:21:54
LCG.CIEMAT.es	Sta...	2013-10-16 14:02:06	2013-10-16 14:02:06	2013-10-16 13:55:38
LCG.CIEMAT.es	Sta...	2013-10-16 14:02:04	2013-10-16 14:02:04	2013-10-16 13:55:28
LCG.DESY-ZEUT...	Unk...	2013-10-16 14:01:08	2013-10-16 14:01:08	2013-10-16 12:33:16
LCG.CAMK.pl	Unk...	2013-10-16 12:29:59	20	
LCG.DESY-ZEUT...	Ast...	2013-10-16 10:03:22	20	

Selected Statistics :: Status (Wed Oct 16 2013 20:22:59 GMT+0000 (UTC))

Key
🟩 Completed
🟩 Done
🟥 Failed
□ Killed
🟦 Running
🟧 Waiting

Running jobs by Site

41 Weeks from Week 53 of 2012 to Week ...

Max: 5,143, Min: 0.00, Average: 608, Current: 3...

Generated on 2013-10-16 14:48:15 UTC

Job Launchpad

Proxy Status: Valid

Predefined Sets of Launchpad Values

Available Sets

Mandelbrot

JDL

Executable: mandelbrot

JobName: Mandelbrot_%j

Arguments: -W 600 -H 600 -X -0.46490 -Y -0.56480 -P 0.

OutputSandbox: *.bmp

StdError: %j.err

CPUTime: 3600

StdOutput: %j.out

Input Sandbox

...either your private key nor ... our service. While we try to ...ure as possible by using SSL ...with your credentials when it ... for maximum security, we ...anually convert and upload ...lient commands:

...E.p12 ...GROUP_NAME

DIRAC interfaces

- Command line interface
  - `dirac-xxx` style commands mimicking standard Grid commands
    - More than 200 commands for all the purposes
  - COMDIRAC subproject provides an intuitive shell-like set of commands
    - Inspired by iRods iCommands interface
    - Those will be mostly used during the tutorial
- Python API for power users
  - Allows to program specific applications to work with the DIRAC services

DIRAC Framework

DIRAC middleware

- Services oriented architecture (SOA)
- DIRAC has a well defined architecture
  - Services
    - passive components reacting to client request
    - Keep their state in a database
  - Light distributed agents
    - permanently running components, animating the whole system
  - Clients
    - User interfaces
    - Agent-service, service-service communications
- Technologies
  - Python, MySQL, OpenSSL

DIRAC Framework

- All the communications between the distributed components are secure
  - DISET custom client/service protocol
    - Focus on efficiency
    - Control and data communications
  - X509, GSI security standards
  - Fine grained authorization rules
- The framework allows to easily build these components concentrating on the business logic of the applications
  - Making use of rich base services: Configuration, Security, Monitoring, support for asynchronous operations, Accounting, etc

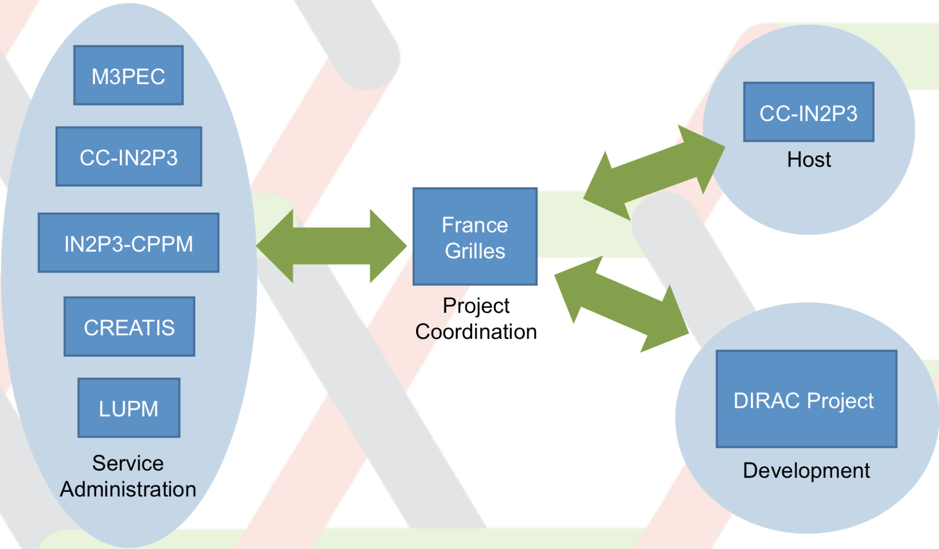
DIRAC as a Service

DIRAC as a service

- DIRAC client is easy to install
  - Part of a usual tutorial
- DIRAC services are easy to install but
  - Needs dedicated hardware for hosting
  - Configuration, maintenance needs expert manpower
  - Monitoring computing resources is a tedious every-day task
- Small user communities can not afford maintaining dedicated DIRAC services
  - Still need easy access to computing resources
- Large grid infrastructures can provide DIRAC services for their users.

France-Grilles

- **DIRAC** services are provided by several National Grid Initiatives: France, Spain, Italy, UK …
- Example: France-Grilles DIRAC service
  - Hosted by the CC/IN2P3
  - Distributed administrator team
    - 5 participating universities
  - 20 VOs, ~150 registered users
  - In production since May 2012
    - >30 millions jobs
- DIRAC 4 EGI service is available for users in Europe
  - Also similar services in UK, Italy, Russia Chine, Romania, etc



Service User Support

- Heavily used for the grid tutorials
  - Using resources of the VO france-formation

- Support for users, applications
  - Forum for experience dissemination
  - Help in porting applications to the grid
  - Help new communities to try out DIRAC for their production systems

    □ Fermi-LAT, Glast

    □ LSST

    □ CTA

    □ …

Summary

DIRAC has most of the features of a "standard" Grid middleware stack

Power users will see extra support:

- Massive job execution
- Data operations

Developers can easily add new functionalities specific for their applications

Community administrators get tools to apply community policies

- User and group priorities, quotas

Site administrators can easily include their resources

- Easy addition of new resources without bulky installation
- Easy user management with only one "VO user"

The DIRAC project is in full development

- More new exciting features to come – stay tuned !
- Your contributions are welcome

DIRAC Tutorial plan

Getting Started: DIRAC client installation, getting ready user credentials, preparing the client environment

Job execution mechanics: Basic job operations with Web Portal explained

Job manipulation tools: Submission, monitoring, getting results

Basic data management operations: data uploading, downloading, replication

Advanced job operations: jobs with input and output data, bulk job submission

Advanced data operations: managing metadata

Emphasis on exercises

Agenda with all the presentations:

https://indico.in2p3.fr/event/11764

DIRAC Tutorial setup

- DIRAC service installation at *IN2P3 Computing Center, Lyon*
  - The service used for the tutorial is permanent, will stay in place afterwards*:*
  - ***http://dirac.france-grilles.fr***
- Resources
  - EGI sites
    - Computing Elements
  - Storage Elements
    - One DIRAC SE (DIRAC-USER)
    - One iRods SE (MCIA-irods)
    - Several SRM EGI SE's
- VO vo.formation.idgrilles.fr
  - DIRAC group ***dirac_user***

- Tutorial materials are available here
  - https://github.com/DIRACGrid/DIRAC/wiki/DIRAC-Tutorials
  - https://github.com/DIRACGrid/COMDIRAC/wiki

The course tutors team

- The course will be given by the members of the DIRAC Project team (http://diracgrid.org)
- Tutors:
  - Andrei Tsaregorodtsev – DIRAC Project Coordinator, CPPM/Marseille
  - Vanessa Hamar – Responsible for the DIRAC production infrastructure in France-Grilles NGI, CC/IN2P3
  - Pierre Gay – T2 and FG-DIRAC administrator, Bordeaux University
  - Sorina Camarasu – biomed VO and FG-DIRAC administrator, CREATIS/CNRS