DIRAC framework for distributed computing

L. Arrabito[1], J. Bregeon[1], P. Gay[2], V. Hamar[3], F. Hernandez[3], S. Pop[4], A.Tsaregorodtsev[5]
[1]LUPM-IN2P3/CNRS, Montpellier
[2]Université de Bordeaux
[3]CC-IN2P3/CNRS, Lyon
[4]Creatis CNRS, Lyon
[5]CPPM-IN2P3-CNRS, Marseille
Journées SUCCESS 2017, 16th-17th October, Grenoble

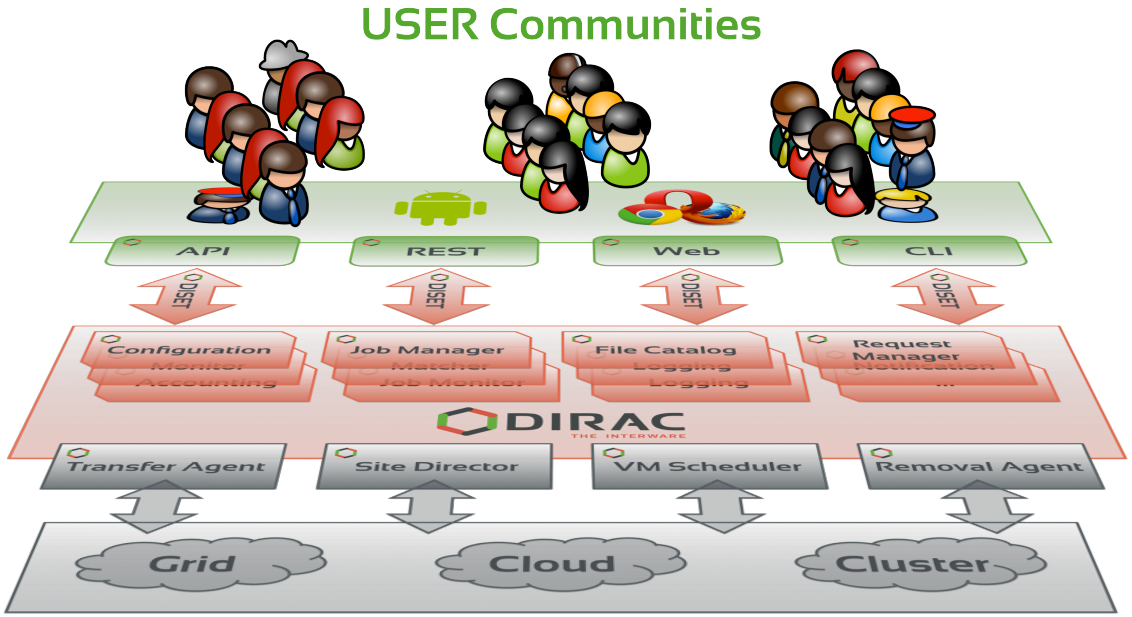
DIRAC
THE INTERWARE

Plan

- DIRAC overview
- Managing computing and storage resources
- Users communities
- DIRAC as a service
- Conclusions

The DIRAC interware

- A software framework for **distributed computing**
- Intermediates between **users and resources**
- Allows **interoperability** simplifying **interfaces**



The project

- Originally developed to support the production activities of the LHCb experiment at CERN (~10 years ago)
- Today is a general purpose software, targeting several large scientific communities
- Open source project developed by user communities for users
  - http://github.com/DIRACGrid/DIRAC
  - Publicly documented, active assistance forum, yearly users workshops, open developers meetings
- The DIRAC consortium as representing body
  - CERN, CNRS, University of Barcelona, KEK, IHEP, University of Montpellier as members
- In 2017 started as IN2P3 project to support further software generalization
  - CPPM, CC-IN2P3, LUPM, Creatis/CNRS, University of Bordeaux

The actors


USER Communities
API
REST
Web
CLI
DISET
Configuration
Job Manager
File Catalog
Request Manager
DIRAC
Transfer Agent
Site Director
VM Scheduler
Removal Agent
Grid
Cloud
Cluster
Resources

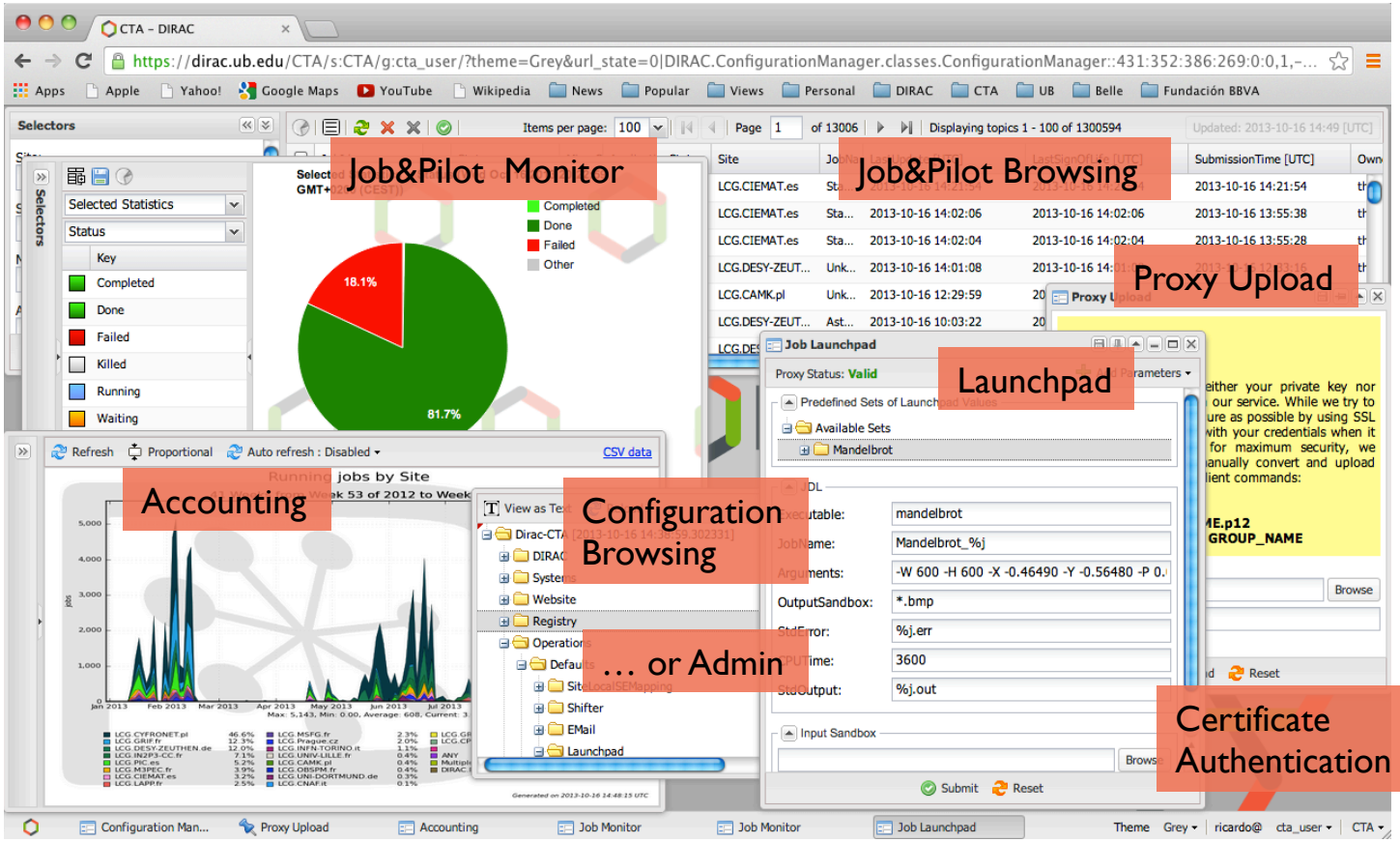
Main DIRAC components

- Workload Management System (WMS)
  - Job brokering with Pilot Jobs
  - Interoperability with different types of computing resources
- High level workflows Management System (Transformation System)
  - Support for automated massive data production and processing
- Data Management System
  - Storage management (access to various storage systems)
  - Data bookkeeping (File Catalog)

- These components should be installed on dedicated services to which clients connect

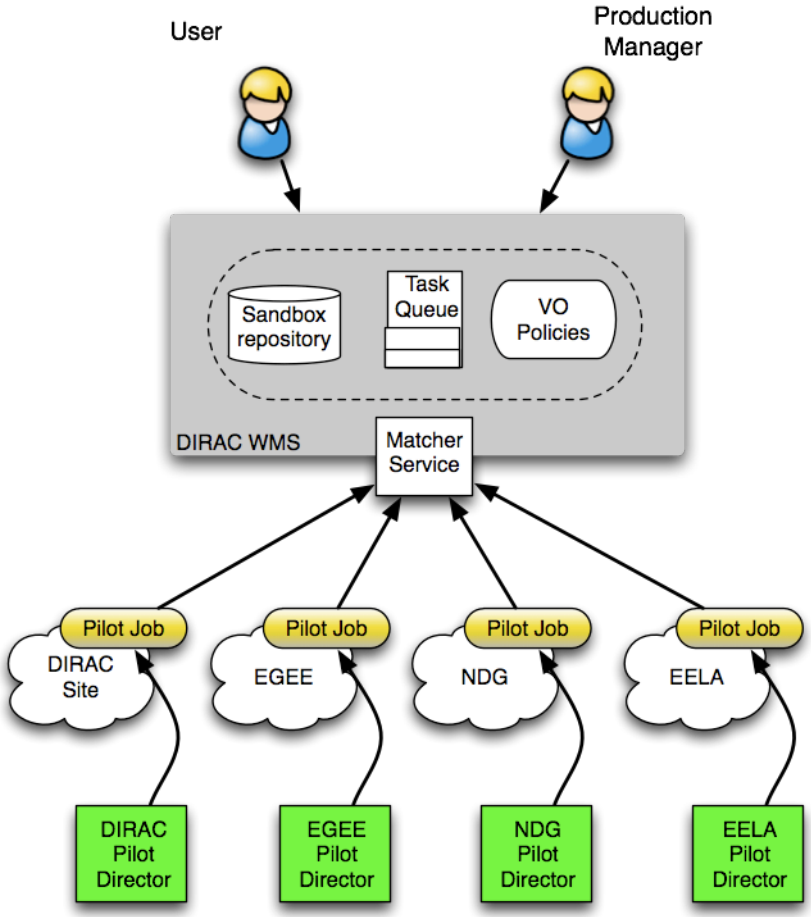
Interfaces

- Command line tools, Python API, RESTful interface
- Web portal



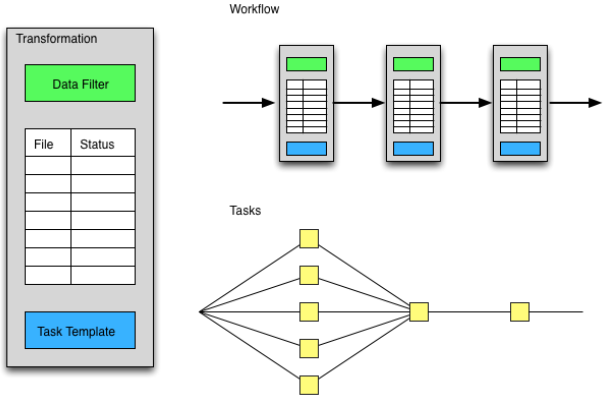
Workload Management System

- Implementation of Pilot Jobs
  - Introduced by the DIRAC project, now widely used in various WMS
  - Efficient usage of resources
- Transparent access to different types of resources
  - grids, clouds, clusters, etc.



Transformation System

- Conceived for **Production Managers**
- Enables automated **workflow execution** handling **large datasets**, *e.g.*:
  - Data Processing -> Merging -> Data Analysis -> Data Replication
- Workflows as chains of ‘data transformations’
- Transformation = input data filter + recipe to create tasks (jobs or data operations)
- Transformations are created by the Production Manager
- Tasks are automatically created through as soon as input data are registered in the system
  - **Fully data-driven system**



DIRAC
THE INTERWARE

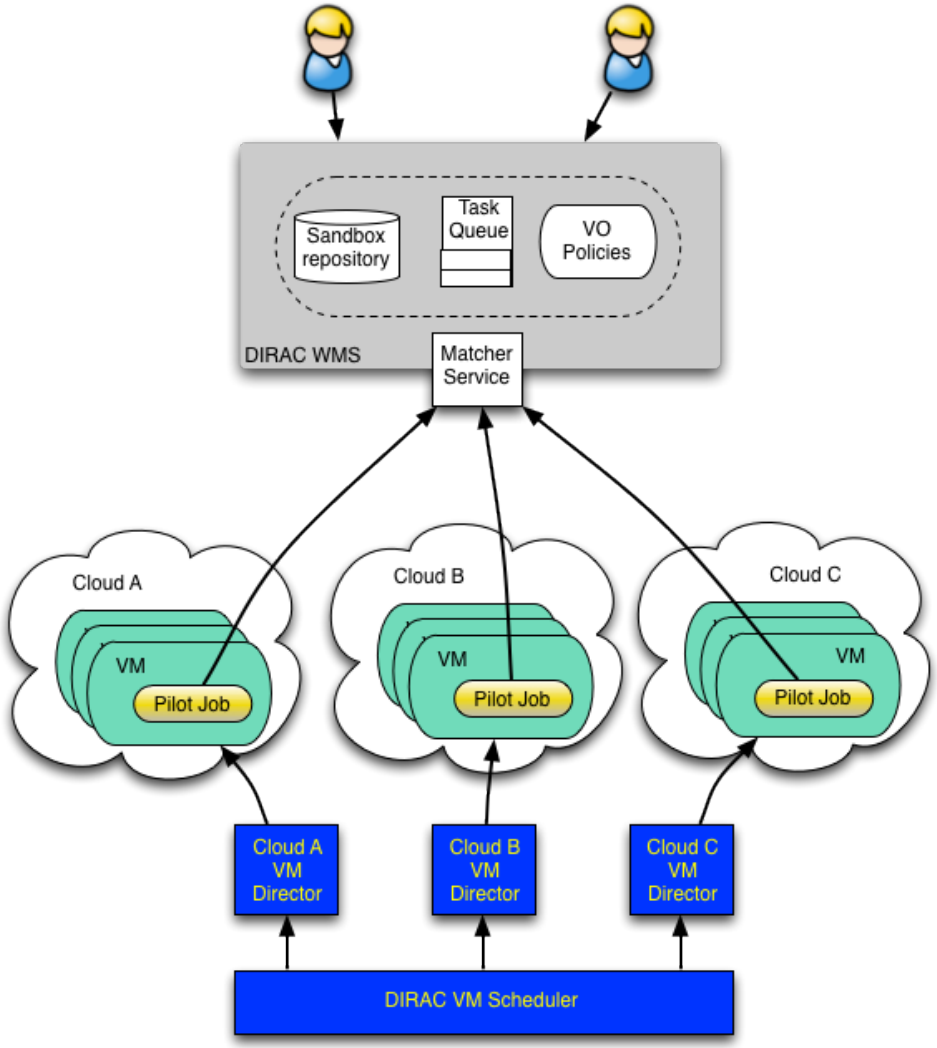
Resources available via the DIRAC system

Computing Resources: HTC

- DIRAC was initially developed with the focus on accessing conventional Grid computing resources
  - It fully supports *multiple grid middleware* and infrastructures
    - EGI, WLCG, OSG, NorduGRID, etc
  - Other types of grids can be supported
    - As requested by users
- Standalone clusters
  - Access through SSH/GSISSH tunnel
  - Batch systems supported: LSF, SGE, PBS/Torque, Condor, OAR, SLURM
- BOINC Volunteer resources

Computing resources: Cloud

- VMDIRAC module
- Allows transparent access to various private or public cloud resources
  - Apache-libcloud
  - Rocci
  - EC2
  - Others are in the works
    - OCCI, Google, Azur, IBM, …
    - Preferring RESTful interfaces
- Manages the whole VMs life cycle
  - Creation, Monitoring, Discarding



Computing Resources: HPC

- Multiple HPC centers are available for large scientific communities
  - E.g., HEP experiments started to have access to a number of HPC centers
    - Using traditional HTC applications
    - Filling in the gaps of empty slots
    - Including HPC into their data production systems
- Advantages of federating HPC centers
  - More users and applications for each center - better efficiency of usage
  - Elastic usage: users can have more resources for a limited time period
- Collaboration with HPC centers to integrate them into a common framework under the DIRAC WMS control
  - France: Aix Marseille University
  - Russia: Dubna, NNGU, others
  - China: IHEP HPC center, Beijing

Computing Resources: HPC

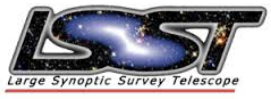
- Unlike grid sites, HPC centers are not uniform
  - Different access protocols
    - Different user authentication methods
  - Different batch systems
  - Different connectivity to outside world
- DIRAC work in progress to overcome these differences
  - Support HPC access protocols (SSH, GSISSH, ARC, OAR, SLURM)
  - Methods for remote control of user payloads and data at the HPC centers
    - Site proxy/gateway services
    - User data import and export

User communities

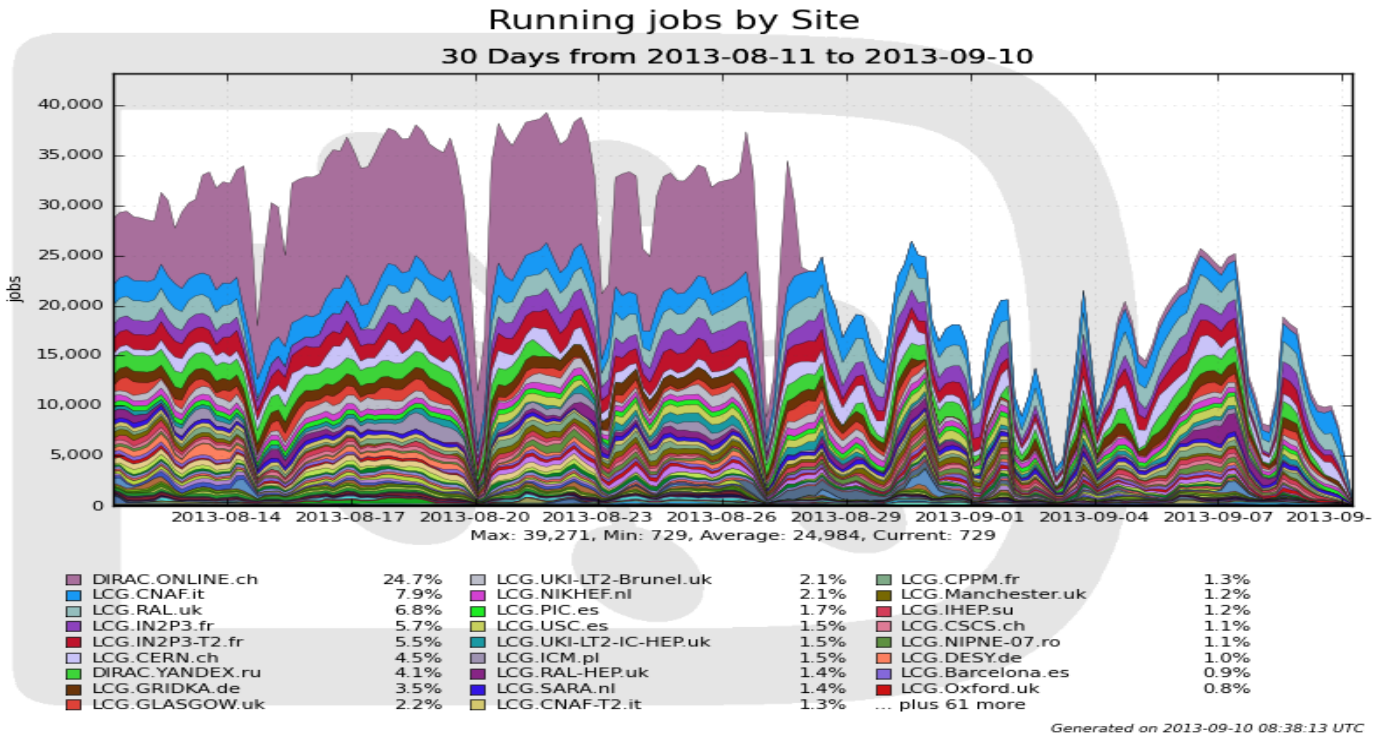
- High Energy Physics

- Astrophysics

- And many others



LHCb Collaboration



- More than 100K concurrent jobs in ~120 distinct sites
  - This is not the limited by the system capacity, but by the available resources
  - In needed, further optimizations to increase the capacity are possible
    - Hardware, database optimizations, service load balancing, etc.

DIRAC dedicated installations

- Belle II Collaboration, KEK
  - First use of clouds (Amazon) for data production
- ILC/CLIC detector Collaboration, Calice VO
  - DIRAC File Catalog was developed to meet the ILC/CLIC requirements
- BES III, IHEP, China
  - Dataset management developed for the needs of BES III
- CTA
  - Contributing to Transformation System development enabling data-driven workflows
- Geant4
  - Validation of MC simulation software releases
- DIRAC evaluations by other experiments
  - LSST, Auger, TREND, Daya Bay, Juno, ELI, NICA, …
  - Evaluations can be done with general purpose DIRAC services

DIRAC
THE INTERWARE

EGI ACCOUNTING PORTAL

Normalised CPU time [units 1K.SI2K.Hours] by DATE and VO												
DATE	alice	atlas	belle	biomed	cms	compchem	ilc	lhcb	virgo	vo.cta.in2p3.fr	Total	%
Nov 2015	83,043,071	213,187,021	29,633,040	2,992,249	107,998,028	812,409	3,051,240	44,495,710	365,193	5,203,790	**490,781,751**	**8.60%**
Dec 2015	81,681,064	167,642,164	30,755,315	2,771,463	81,200,999	1,197,402	10,250,775	42,772,247	4,370	9,643,804	**427,919,603**	**7.50%**
Jan 2016	100,472,899	212,596,116	8,254,706	2,221,994	99,768,667	2,869,544	3,904,455	32,614,451	329,113	8,746,790	**471,778,735**	**8.27%**
Feb 2016	80,340,391	202,531,157	48,965	1,312,309	100,330,129	1,220,127	2,704,948	44,547,976	1,962,465	5,563,528	**440,561,995**	**7.72%**
Mar 2016	108,810,699	172,663,251	3,412,262	2,286,939	75,113,354	1,623,540	2,049,130	83,154,401	1,917,611	1,539,919	**452,571,106**	**7.93%**
Apr 2016	111,707,745	211,516,946	496,969	1,622,314	67,855,621	1,970,394	3,051,624	78,821,567	3,517,152	3,079,316	**483,639,648**	**8.47%**
May 2016	88,434,699	229,055,135	457,771	3,055,283	64,161,648	3,990,478	4,366,309	70,550,242	11,311,493	669,299	**476,052,357**	**8.34%**
Jun 2016	91,963,895	220,222,321	10,039,317	1,375,916	104,040,606	1,755,334	2,097,169	66,545,602	2,558,741	1,103,183	**501,702,084**	**8.79%**
Jul 2016	113,408,142	187,198,001	3,614,046	2,152,445	104,373,741	1,614,892	1,596,155	65,898,735	8,005,698	7,794,153	**495,656,008**	**8.69%**
Aug 2016	88,278,412	212,942,846	34,225	6,500,219	51,366,225	3,474,177	5,538,912	72,803,805	2,919,127	5,410,036	**449,267,984**	**7.87%**
Sep 2016	88,164,653	309,040,532	7,314,602	514,897	90,018,815	2,602,763	3,297,430	106,365,999	1,770,213	6,487,567	**615,577,471**	**10.79%**
Oct 2016	68,902,764	167,532,717	1,528,430	467,733	82,329,281	1,301,416	5,324,702	71,019,670	2,752,272	104,325	**401,263,310**	**7.03%**
Total	**1,105,208,434**	**2,506,128,207**	**95,589,648**	**27,273,761**	**1,028,557,114**	**24,432,476**	**47,232,849**	**779,590,405**	**37,413,448**	**55,345,710**	**5,706,772,052**	
Percentage	**19.37%**	**43.91%**	**1.68%**	**0.48%**	**18.02%**	**0.43%**	**0.83%**	**13.66%**	**0.66%**	**0.97%**		

- 5 out of Top-10 EGI communities used heavily DIRAC for their payload management in the last year
  - lhcb, belle, biomed, ilc, vo.cta.in2p3.fr
  - compchem will likely join the club

Getting started

- For a full evaluation
  - Install a server instance dedicated to your community
- The easiest
  - Connect to a running **DIRAC service** (see next slides)
    - Just need to register and install the DIRAC client on your laptop
- Doc/Support
  - Documentation
    - http://diracgrid.org/
  - User Forum
    - https://groups.google.com/forum/?hl=en#!forum/diracgrid-forum

DIRAC
THE INTERWARE

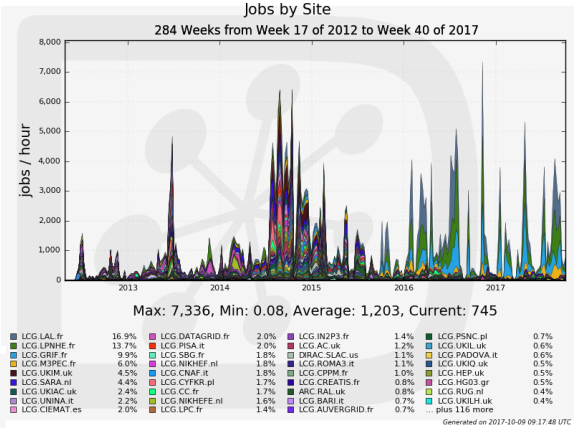
DIRAC as a service

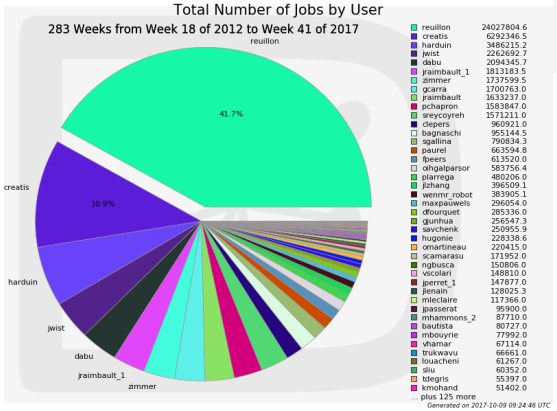
DIRAC as a service

- DIRAC framework was updated to support multi-VO installations
  - Allows to provide better support for several small user communities with a single DIRAC instance
  - Keeps the costs of operating DIRAC under control
- Several services provided by national grid infrastructure
  - FG-DIRAC, France
    - http://www.france-grilles.fr/catalogue-de-services/fg-dirac/
  - GridPP, UK
    - https://www.gridpp.ac.uk/services/gridppdirac/
- DIRAC4EGI
  - http://dirac.egi.eu/DIRAC/
  - Starting from 2018 DIRAC becomes a Core Service of EGI
  - Serving both Grid and FedCloud resources

FG-DIRAC service

- Joint effort to provide FG-DIRAC service
  - Hosted by CC-IN2P3
  - Distributed team of service administrators
    - 5 participating institutes (CC-IN2P3, CPPM, Creatis, U. Bordeaux, LUPM)
- Usage
  - 21 Virtual Organizations
    - robot users
      - VIP/GateLab Biomed
  - About 12 million jobs processed every year





Conclusions

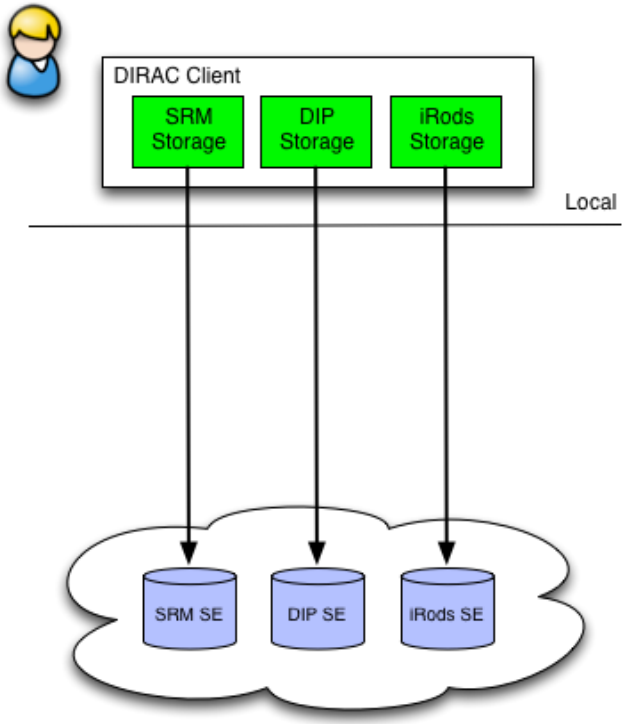
- Distributed computing is no more something exotic, it is used in a daily work by users in various scientific domains
- DIRAC provides a framework for building distributed computing systems aggregating multiple types of computing and storage resources
- Several large scientific collaborations adopted DIRAC for their production systems. Multiple evaluations are ongoing
- Increasing number of projects providing *'DIRAC as a service'* (multi-community)
- In 2018 DIRAC will become an EGI core service, replacing gLite WMS
- DIRAC can help users to get started in the world of distributed computing and discover its full potential

http://diracgrid.org

Backup

Storage plugins

- Storage element abstraction with a client implementation for each access protocol
  - DIPS, SRM, XROOTD, RFIO, etc
  - gfal2 based plugin gives access to all protocols supported by the library
    - HTTP, DCAP, WebDAV, S3, …
- Each SE is seen by the clients as a logical entity
  - With some specific operational properties
  - SE's can be configured with multiple protocols

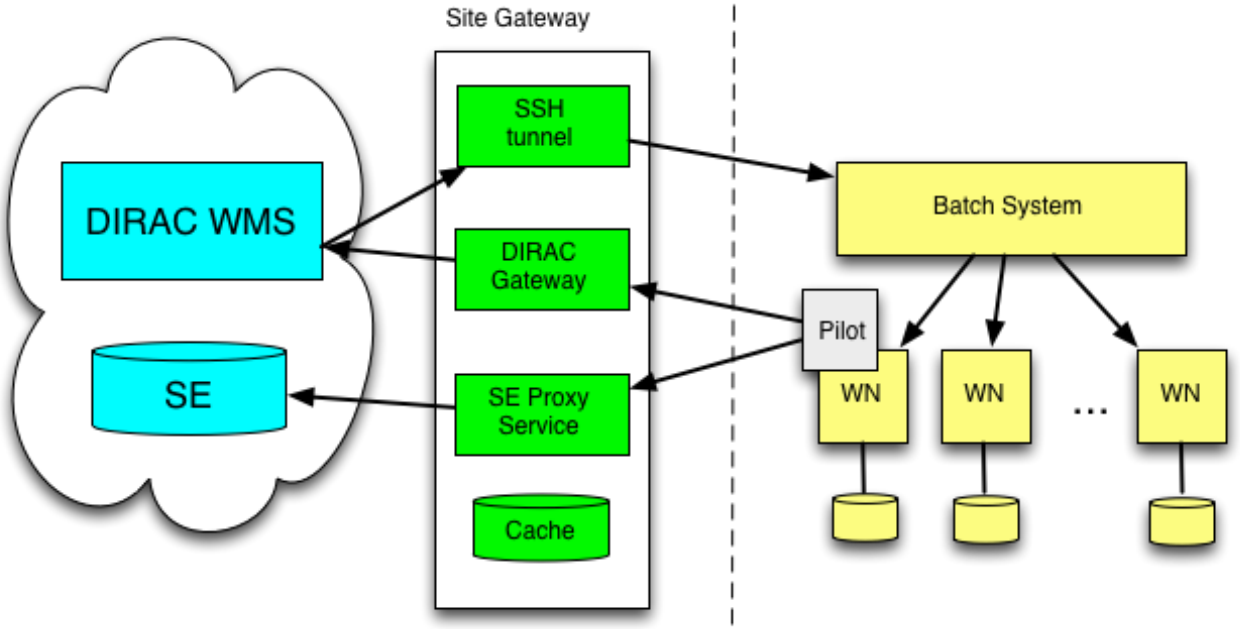


Managing VM life cycle

- *VM creation through a CloudDirector (similar to grid jobs)*
  - *Based on Task Queue status*
    - If there are waiting user payloads
    - VM properties corresponding to payload requirements
- *VM contextualization*
  - On the fly installation of DIRAC, CVMFS, …
  - Starting as many pilots as they are cores (single core jobs)
- *Starting the VM Monitor Agent*
  - Monitor and report the VM state, VM heartbeats
  - Halt the VM in case of no activity
  - Getting instructions from the central service, e.g. to halt the VM
- *VM Scheduler* orchestrates spawning and halting virtual machines depending on the Task Queue status, Accounting history
  - Necessary for fair sharing of cloud resources
  - Work in progress

Accessing HPC resources

- Pilot submitted to the batch system through an (GSI)SSH tunnel
- Pilot communicates with the DIRAC service through the Gateway proxy service
- Output upload to the target SE through the SE proxy

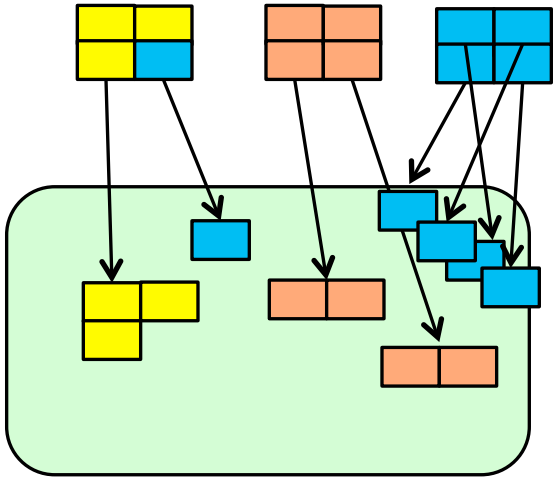


Computing Resources: HPC

- HPC resources allow a rich description with respect to to traditional grid resources
  - DIRAC work in progress to develop a more elaborated model of their description with the corresponding payload matching mechanisms
  - Worker node micro-management
    - Single-core and multi-core applications
    - Multi-processor, multi-node applications are in the works

Computing Resources: HPC

- Multi-core job scheduling
- Pilots with partitionable internal slots
  - M-core Pilots pull N-core jobs (N<=M) until internal slots used up
  - Pilot is standard-size, can be whole-node, 4-node, 8-node….
  - Optimizing CPU efficiency



DIRAC4EGI service

- In “best effort” production since 2014
- Partners
  - Operated by EGI
  - Hosted by CYFRONET
  - DIRAC Project providing software, consultancy
- 10 Virtual Organizations
  - enmr.eu, vlemed, eiscat.se
  - fedcloud.egi.eu
  - training.egi.eu
- Usage
  - Workload Management solution
    - > 6 million jobs processed in the last year
  - Data Management solution
    - E.g. Eiscat 3D
- Starting from 2018 DIRAC becomes a Core Service of EGI
  - WMS replacement
  - Serving both Grid and FedCloud resources

DIRAC4EGI activity snapshot