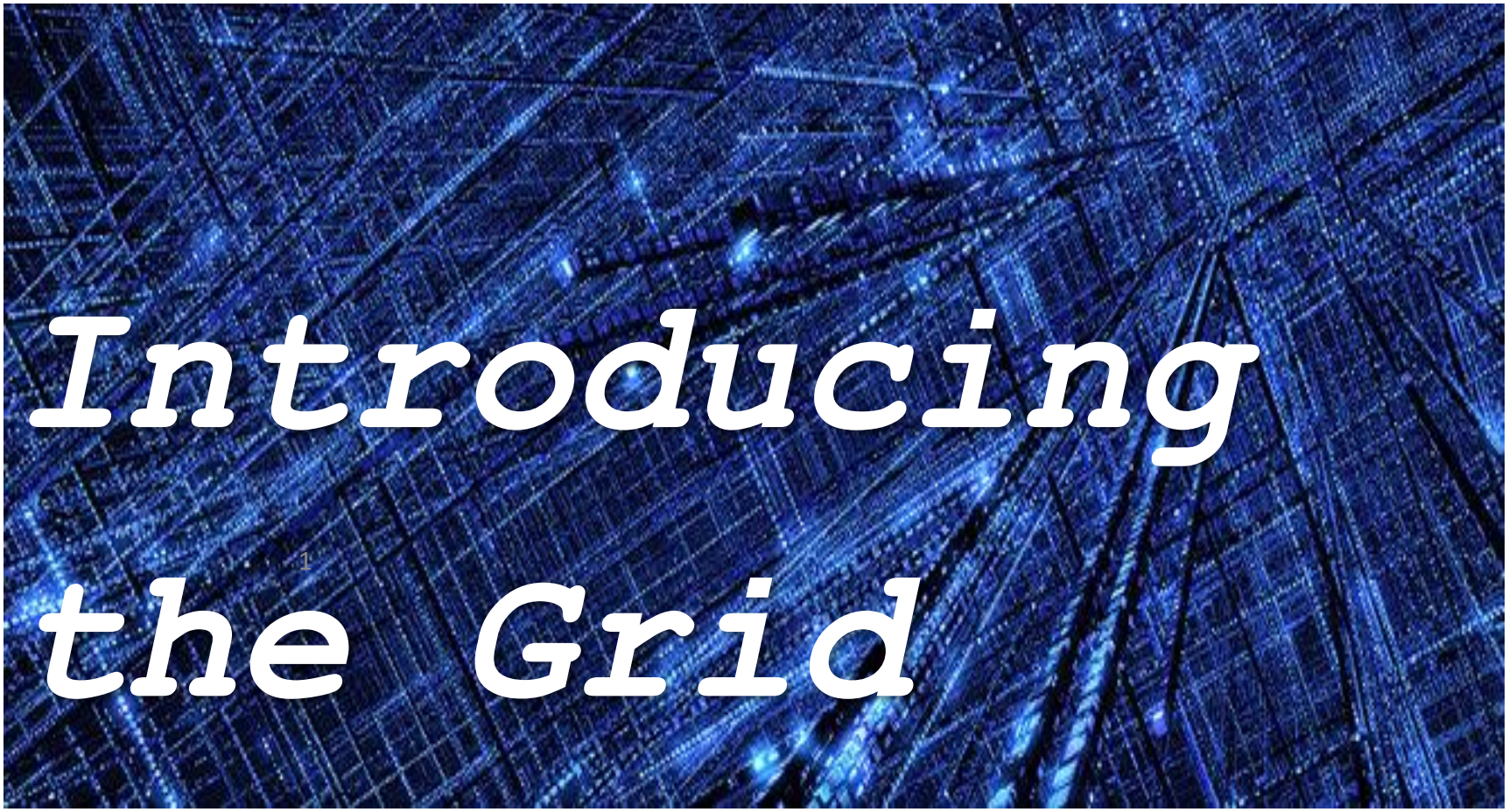


A. Sartirana

Plan of part 1.

I. Introducing the Grid

- what is it;
- the VOs;
- architecture;
- examples.

II. Introducing WLCG

- LHC computing;
- activity.

III. WLCG Architecture

- sites
- jobs workflow;
- data management;
- information system;
- authentication;
- VO's applications.

Introducing the grid.

“ The term «The Grid» was coined in the **mid-1990s** to denote a (then) proposed **distributed computing infrastructure** for advanced science and engineering. ”

The Anatomy of the Grid. I. Foster, C. Kesselman, S. Tuecke.

I. Foster

C. Kesselman

- Born during «Building a Computational Grid» workshop at ANL (8-10 Sep 1997)
  - http://www.crpc.rice.edu/newsletters/fal97/news_grid.html;
- some **fundamental texts:**
  - The Grid: Blueprint for a New Computing Infrastructure. I.F, C.K.
  - Grid Services for Distributed System Integration. I.F., C.K., J. Nick, S. Tuecke. Computer, 35(6), 2002.
  - The Anatomy of the Grid: Enabling Scalable Virtual Organizations. I.F., C. K.C., S. Tuecke. International J. Supercomputer Applications, 15(3), 2001.

What is it for/about.

- **Coordinated resource sharing** and problem solving in dynamic, multi-institutional virtual organizations
  - sharing is not only file exchange but rather **direct access to computers**, **software**, **data**, and other resources;
  - it is, necessarily, **highly controlled**, with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs. All of which may **vary dynamically**;

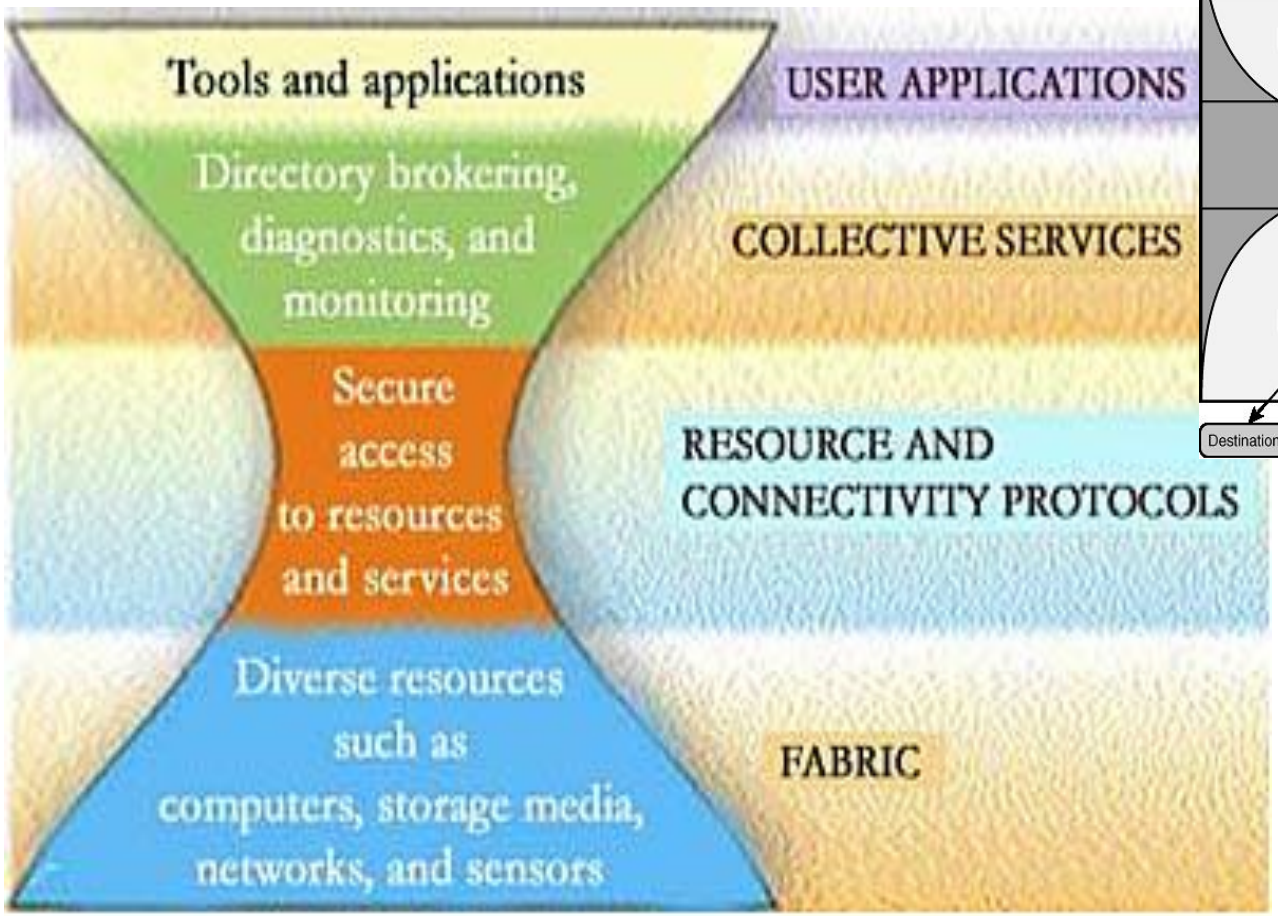
- **Virtual Organizations** (VO's)
  - a **group** of mutually distrustful participants with **varying degrees of prior relationship** (perhaps none at all) that **want to share resources** in order to perform some task.

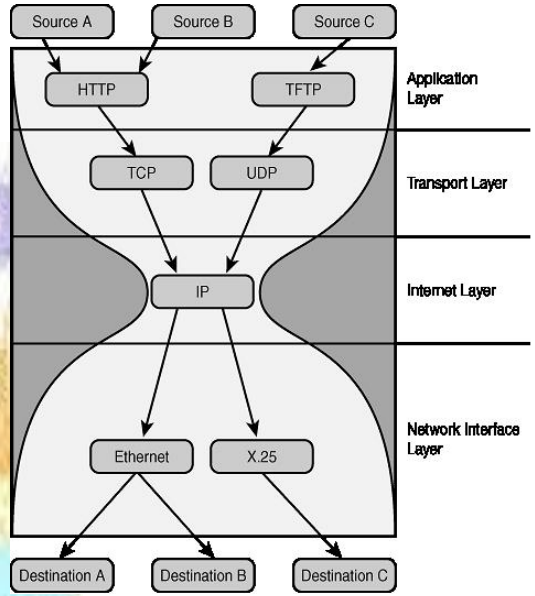
Grid architecture.

- A stack of **common protocols (Middleware)** organized according to the "hourglass model":
  - common protocols are **necessary for interoperability;**
  - standards-based open architecture which facilitates extensibility, interoperability, portability, and code sharing;
  - "hourglass" architecture: **inspired by internet arch.**

- Hourglass model:
  - the **narrow neck** of the hourglass defines a small set of **core abstractions and protocols** (e.g., TCP and HTTP in the Internet),
  - onto which many **different high-level behaviors** can be mapped (the top of the hourglass),
  - and which themselves can be mapped onto many **different underlying technologies** (the base of the hourglass).

The hourglass.





The hourglass.

	Multidisciplinary Simulation	Ray Tracing
Collective (application-specific)	Solver coupler, distributed data archiver	Checkpointing, job management, failover, staging
Collective (generic)	Resource discovery, resource brokering, system monitoring, community authorization, certificate revocation	
Resource	Access to computation; access to data; access to information about system structure, state, performance.	
Connectivity	Communication (IP), service discovery (DNS), authentication, authorization, delegation	
Fabric	Storage systems, computers, networks, code repositories, catalogs	

Examples.

Open Science Grid

The **Open Science Grid Consortium** is an organization that administers a worldwide grid of technological resources called the Open Science Grid, which facilitates distributed computing for scientific research. Founded in 2004, the consortium is composed of **service and resource providers, researchers from universities and national laboratories, as well as computing centers across the United States**. Members i**ndependently own and manage the resources which make up the distributed facility**, and consortium agreements provide the framework for technological and organizational integration.

…

The OSG facilitates **access to distributed high throughput computing for research in the US**. The resources accessible through the OSG are contributed by the community, organized by the OSG, and governed by the OSG consortium. In the last 12 months, we have provided more than 800 million CPU hours to researchers across a wide variety of projects.

…

http://www.opensciencegrid.org

Examples.

EUROPEAN MIDDLEWARE INITIATIVE

The **European Middleware Initiative** (EMI) is a **computer software platform for high performance distributed computing**. It is developed and distributed directly by the EMI project. It is the **base** for other **grid middleware** distributions used by scientific research communities and distributed computing infrastructures all over the world especially in **Europe, South America and Asia**. EMI supports broad scientific experiments and initiatives, such as the Worldwide LHC Computing Grid (for the Large Hadron Collider).
The EMI middleware is a cooperation among three general purpose grid platforms, the **Advanced Resource Connector, gLite and UNICORE and the dCache storage software**.
…
EMI **improves the existing middleware services and harmonizes them**, realizing a common framework with the result of rendering the middleware to be simpler and easier to use, reducing at the same time the interoperability problems faced by the distributed computing infrastructure providers. Thanks to EMI, it was possible to get the results in a previously unthinkable time and to contribute to the discovery of one of the most 'wanted' particles: a Higgs boson.

http://www.eu-emi.eu

Examples.

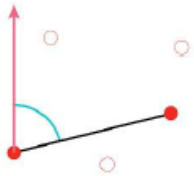
World Community Grid enables anyone with a computer, smartphone or tablet to **donate** their **unused computing power to advance cutting-edge scientific research on topics related to health, poverty and sustainability.** Through the contributions of **over 650,000 individuals and 460 organizations**, World Community Grid has supported **25 research projects** to date, including searches for more effective treatments for cancer, HIV/AIDS and neglected tropical diseases. Other projects are looking for low-cost water filtration systems and new materials for capturing solar energy efficiently.

…

Started in 2004, World Community Grid is a **philanthropic initiative of IBM** Corporate Citizenship, the corporate social responsibility and philanthropy division of IBM. Through Corporate Citizenship, IBM donates its technology and talent to address some of the world's most pressing social and environmental issues.

http://www.worldcommunitygrid.org/

Introducing WLCG.



A key tool to study physics

The most sophisticated data-taking & analysis system ever built for science, providing near real-time access to LHC data.

Global collaboration

42 countries
170 computing centres
2 million jobs run every day

The **Worldwide LHC Computing Grid** (WLCG) is a global computing infrastructure whose mission is to **provide computing resources to store, distribute and analyze the data generated by the Large Hadron Collider** (LHC), making the data equally available to all partners, regardless of their physical location. WLCG is the world's largest computing grid. It is **supported by many associated national and international grids across the world,** such as European Grid Initiative (Europe-based) and Open Science Grid (US-based), as well as many other regional grids. WLCG is co-ordinated by CERN. It is **managed and operated by a worldwide collaboration** between the experiments (ALICE, ATLAS, CMS and LHCb) and the participating computer centres. It is reviewed by a board of delegates from partner country funding agencies, and scientifically reviewed by the LHC Experiments Committee.

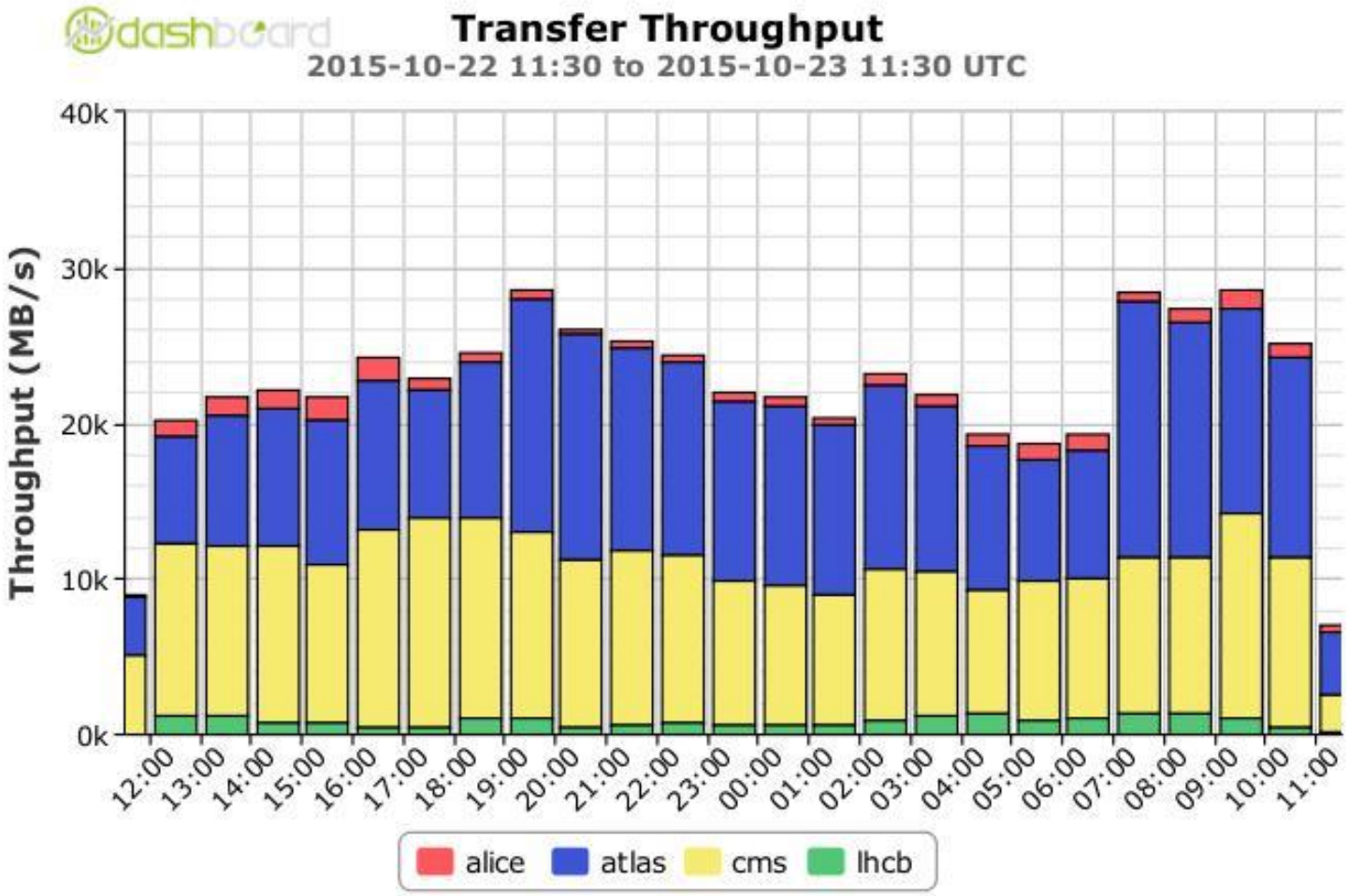
http://wlcg.web.cern.ch/

LHC (CMS) computing.

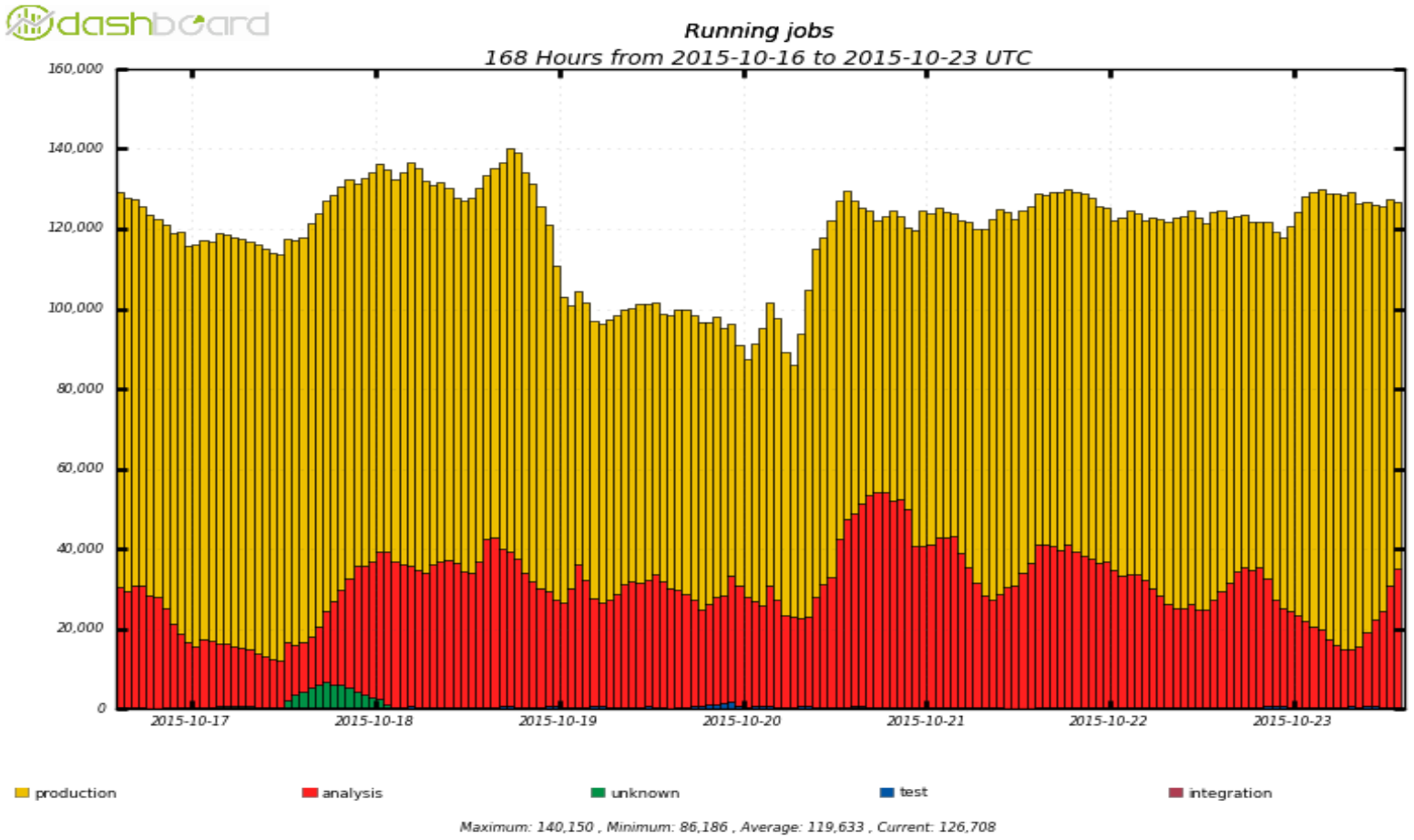
Data pours out of the LHC detectors at a blistering rate. Even after filtering out 99% of it, there is still around **30 petabytes of data per year to deal with**. That's 30 million gigabytes, the equivalent to nearly 9 million high-definition (HD) movies.
The scale and complexity of data from the LHC is unprecedented. This data **needs to be stored, easily retrieved and analyzed by physicists all over the world**. This requires massive storage facilities, global networking, immense computing power, and, of course, funding.
CERN does not have the computing or financial resources to crunch all of the data on site, so in 2002 it turned to grid computing to **share the burden with computer centres around the world.**
The result, the **Worldwide LHC Computing Grid** (WLCG), is a distributed computing infrastructure arranged in tiers – giving a community of over 10,000 physicists near real-time access to LHC data. The WLCG builds on the ideas of grid technology initially proposed in 1999 by Ian Foster and Carl Kesselman (link is external).

http://wlcg-public.web.cern.ch/about

WLCG activity.

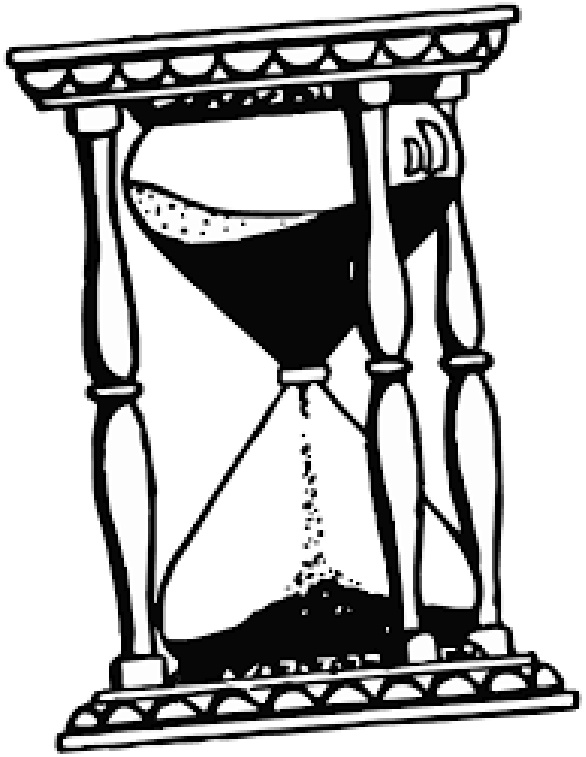


WLCG activity.



WLCG architecture.

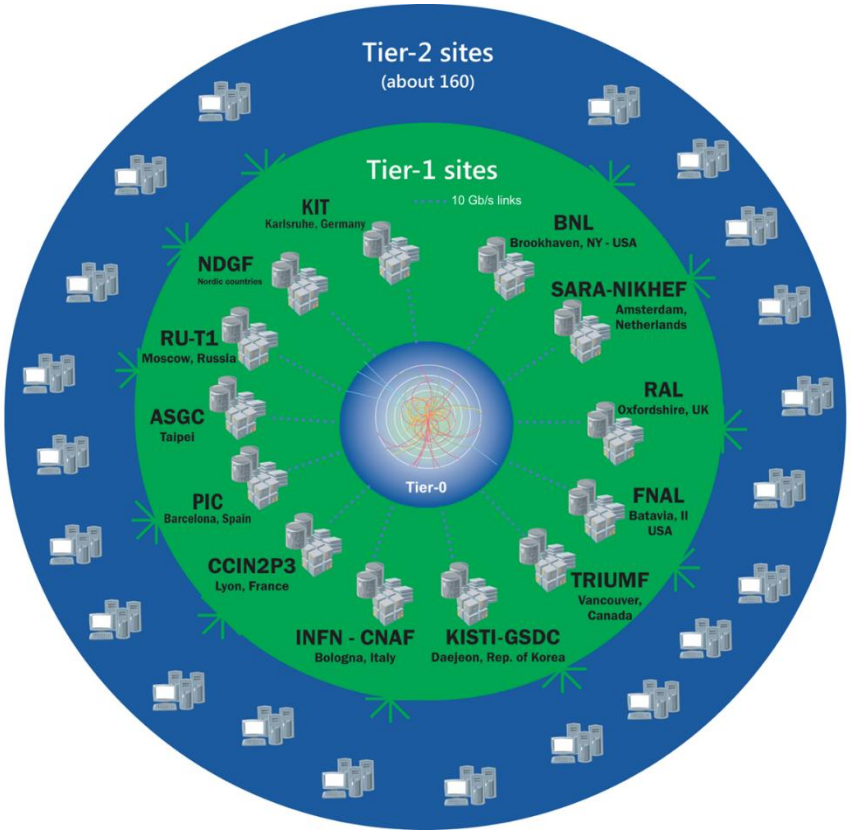
Let's have a look to the WLCG architecture from the bottom to the top of the hourglass…



WLCG sites.

Tier-0

This is the **CERN Data Centre**, which is located in Geneva, Switzerland and also at the **Wigner Research Centre for Physics** in **Budapest**, Hungary over 1200km away. The two sites are connected by two dedicated 100 Gbit/s data links. All data from the LHC passes through the central CERN hub, but CERN provides less than 20% of the total compute capacity.
Tier 0 is responsible for the **safe-keeping of the raw data** (first copy), **first pass reconstruction**, **distribution** of raw data and reconstruction output to the Tier 1s, and reprocessing of data during LHC down-times.

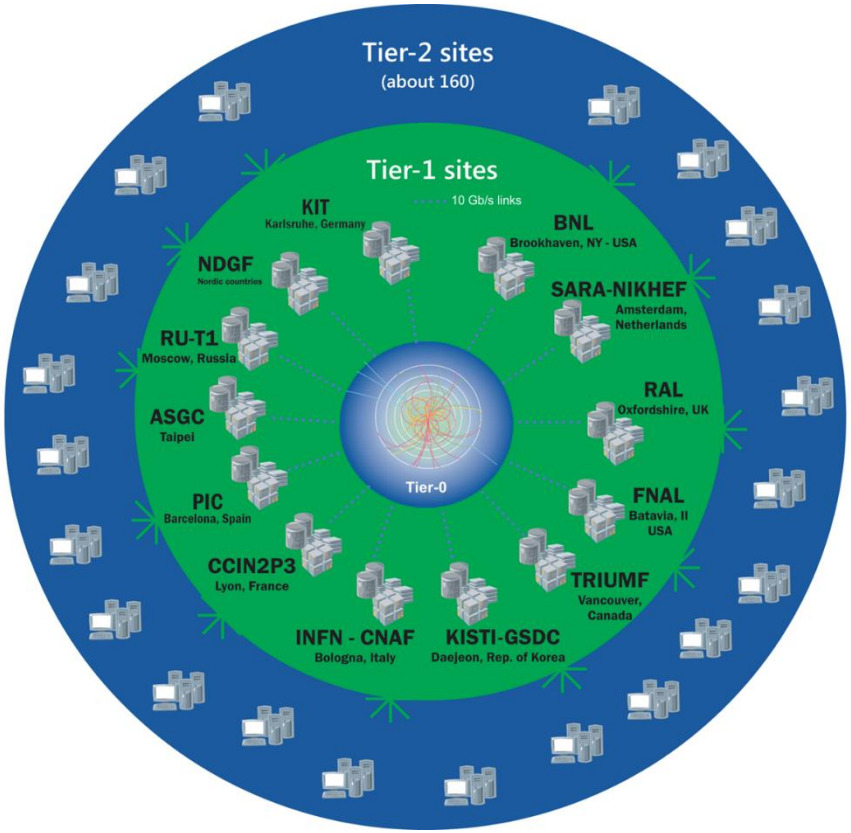


http://wlcg-public.web.cern.ch/tier-centres

WLCG sites.

Tier-1's

These are **thirteen large computer centres** with sufficient **storage capacity** and with **round-the-clock support** for the Grid. They are responsible for the **safe-keeping** of a proportional **share of raw and reconstructed data, large-scale reprocessing** and safe-keeping of corresponding output, **distribution of data to Tier 2s** and **safe-keeping** of a share of **simulated data** produced at these Tier 2s.



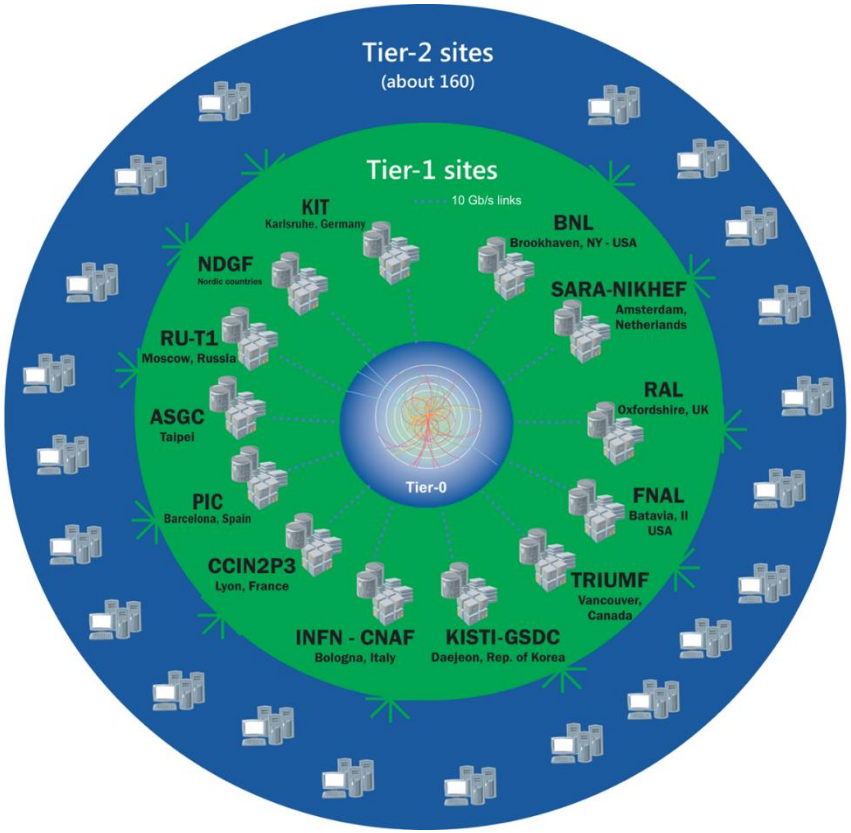
http://wlcg-public.web.cern.ch/tier-centres

WLCG sites.

Tier-2's

The Tier 2s are typically universities and other scientific institutes, which can **store** sufficient **data** and **provide** adequate **computing power** for **specific analysis tasks**. They handle analysis requirements and proportional share of simulated event production and reconstruction.

There are currently around 160 Tier 2 sites covering most of the globe.

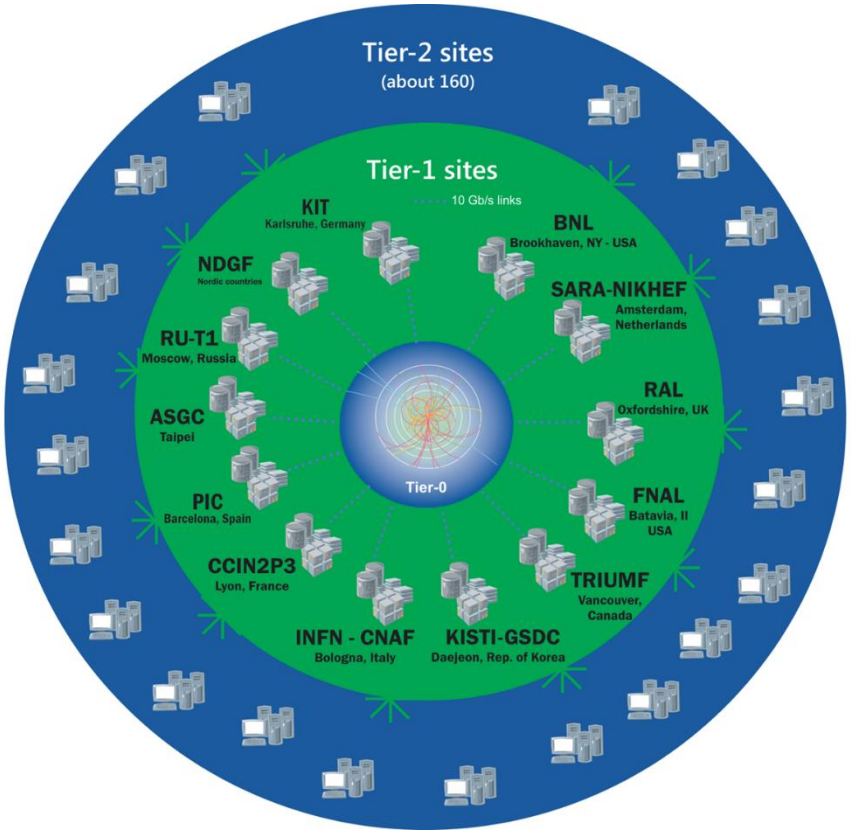


http://wlcg-public.web.cern.ch/tier-centres

WLCG sites.

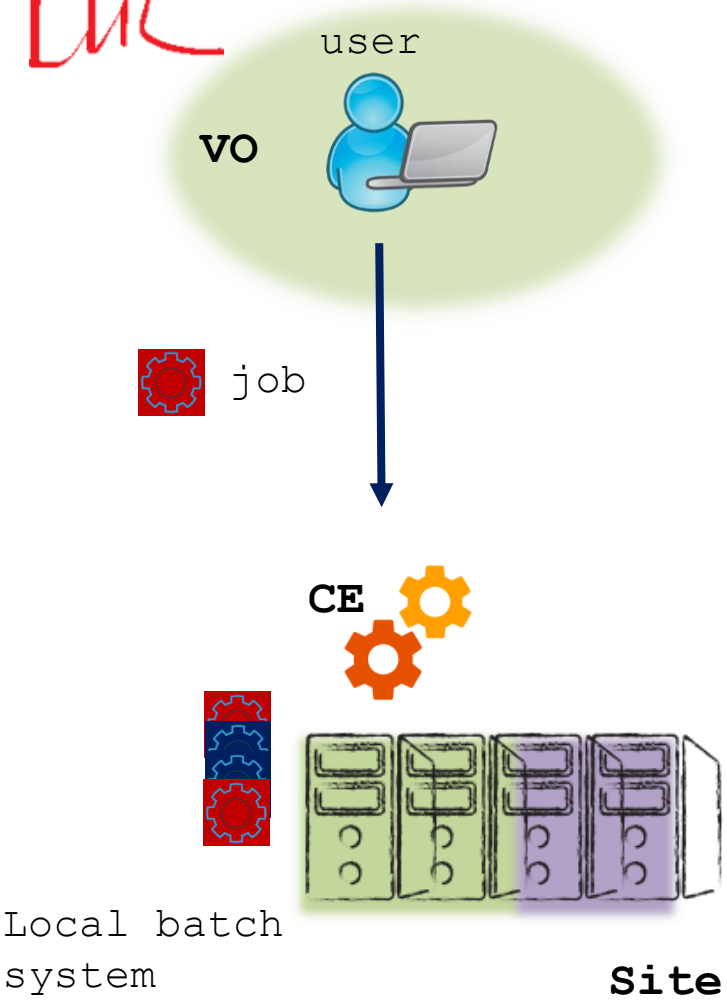
Tier-3's

Individual scientists will access these facilities through **local** (also sometimes referred to as Tier 3) **computing resources**, which can consist of local clusters in a University Department or even just an individual PC. There is **no formal engagement** between WLCG and Tier 3 resources.



http://wlcg-public.web.cern.ch/tier-centres

Grid jobs workflow.

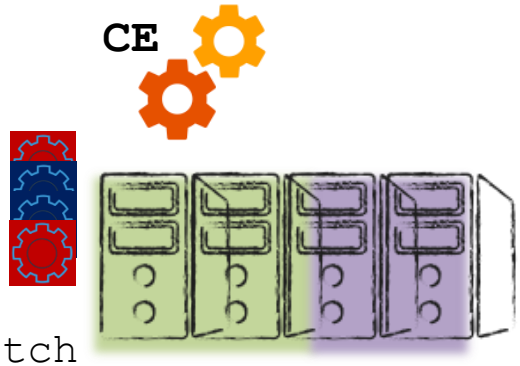


- **Computing Element:**
  - **middleware** layer on top of a local batch system;
  - standard protocol for **different batch technologies;**
  - deals with **authentication** and **authorization,** map the grid user to local group;
  - **local** batch system implements **shares and priorities.**

(Resource Layer)

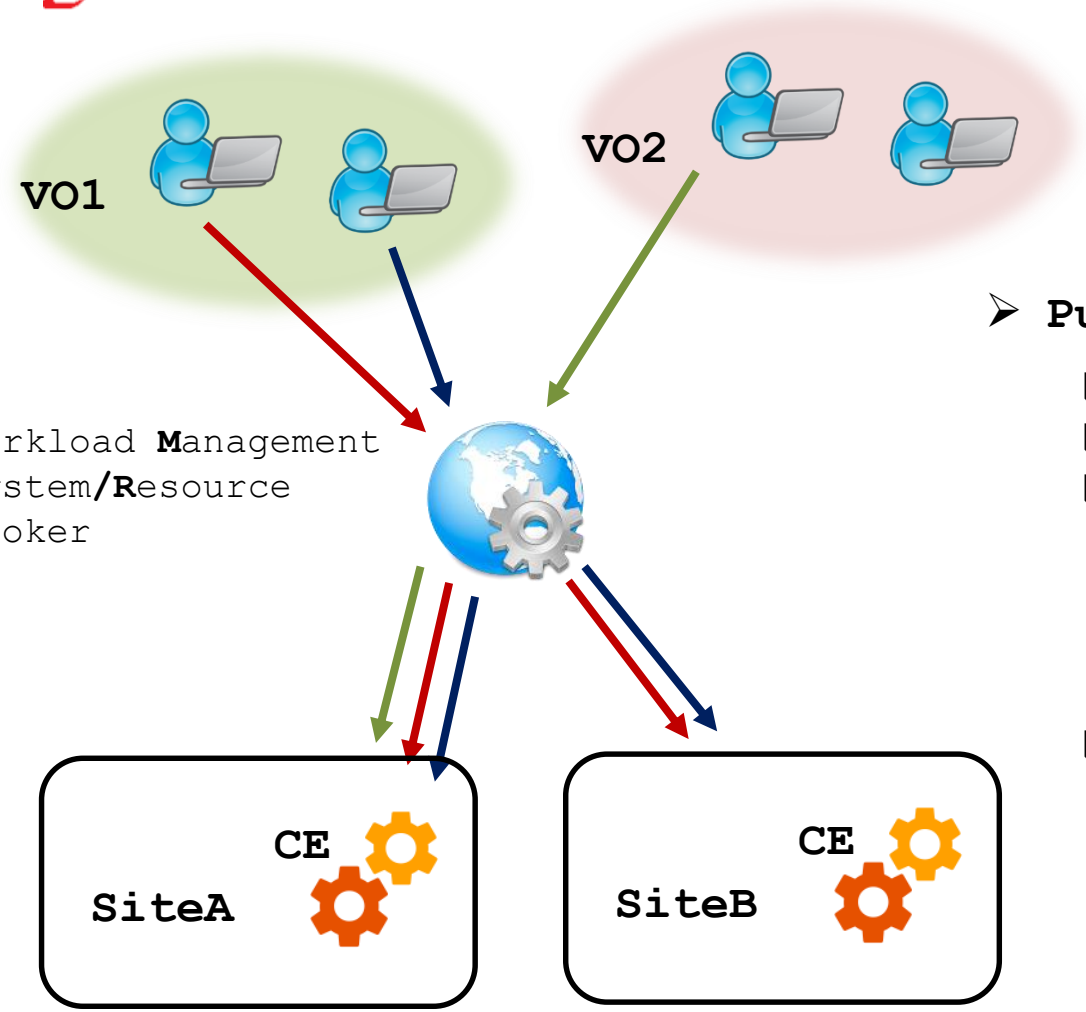
Grid jobs workflow.

To ease notation, from now on…



Local batch system

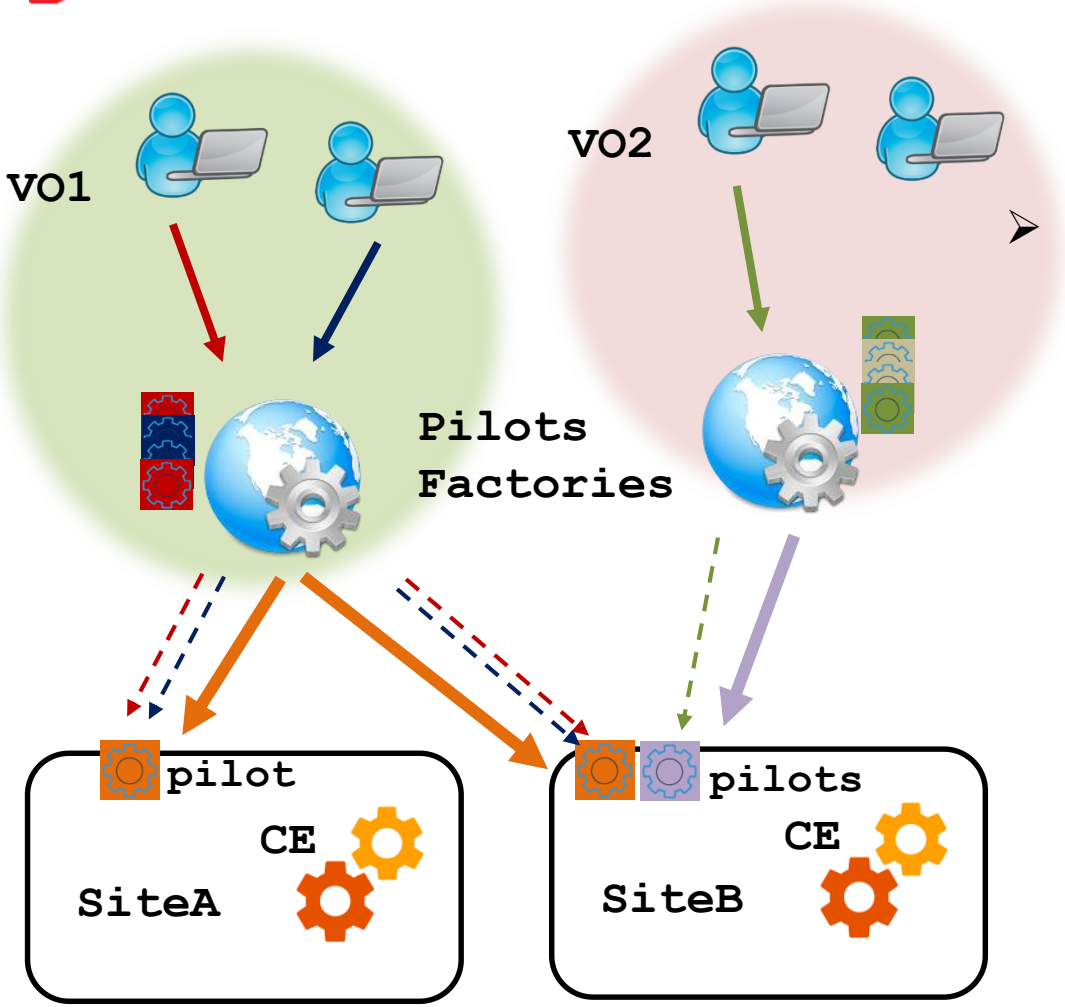
Grid jobs workflow.



(Collective Layer)

- **Push mode:**
  - **baseline** (Run-1);
  - **WMS** does **match-making;**
  - **local sites scheduling** does **fine grained** (intra-VO) policies: not optimal for manage complex policies;
  - users **jobs run directly** on the nodes: non optimal slots usage time.

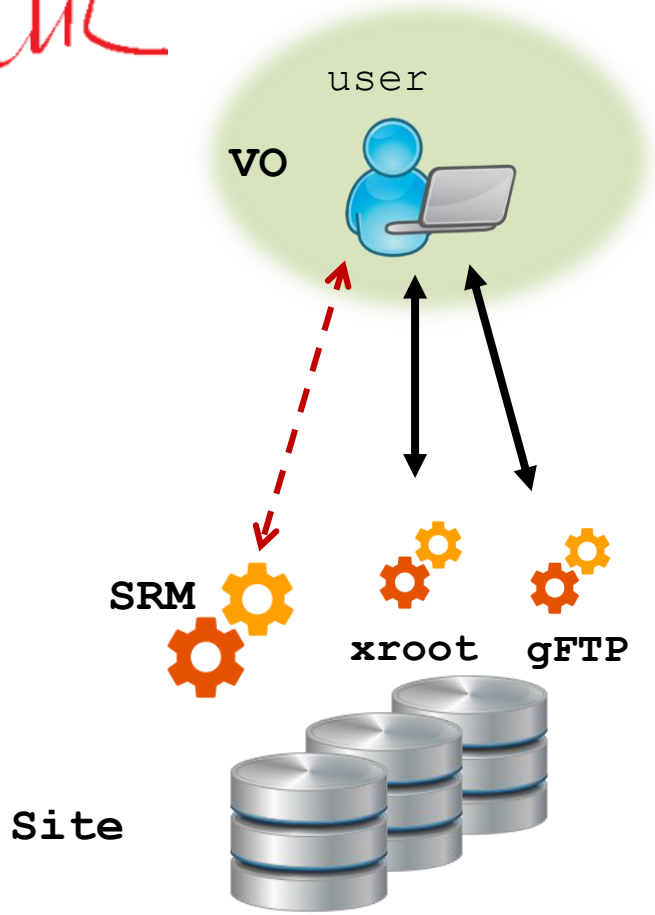
Grid jobs workflow.



(Collective Layer)

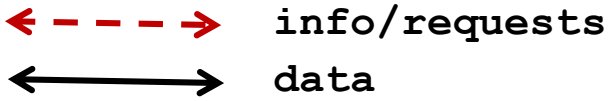
- **Pull mode:**
  - in prod for Run-2;
  - **pilot factory** does **match-making** and **fine-grained** (intra-VO) **scheduling**;
  - submits pilot agents to the site. Local scheduling for coarse grained policies;
  - **pilots pull users jobs** from the pilot factory: usage time optimized.

Data management.



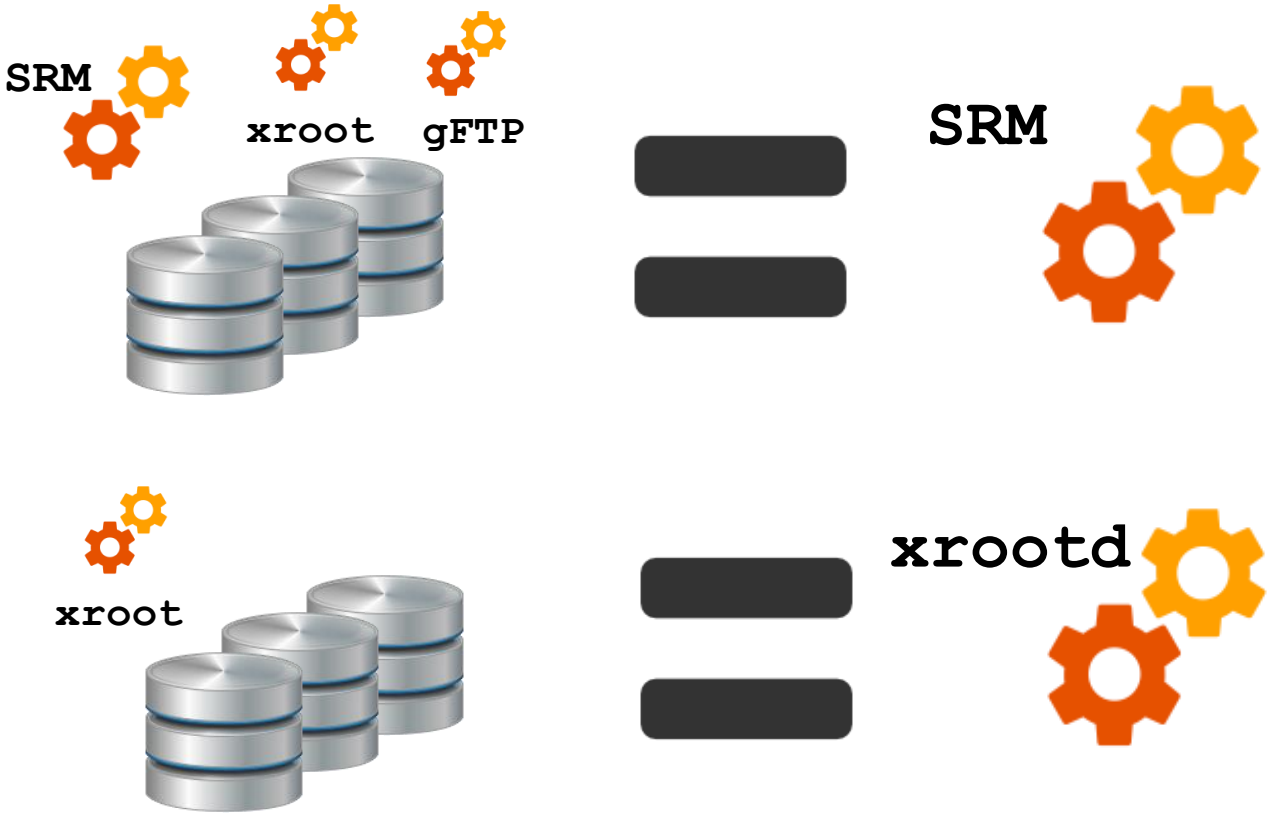
- **SRM based storage:**
  - SRM defines a middleware over the different storage technologies;
  - serfs metadata and redirect data request to other services: gFTP, (gsi)http, xrootd;
  - (gsi)http, xrootd and gridFTP are middleware services over data serving;

(Resource Layer)

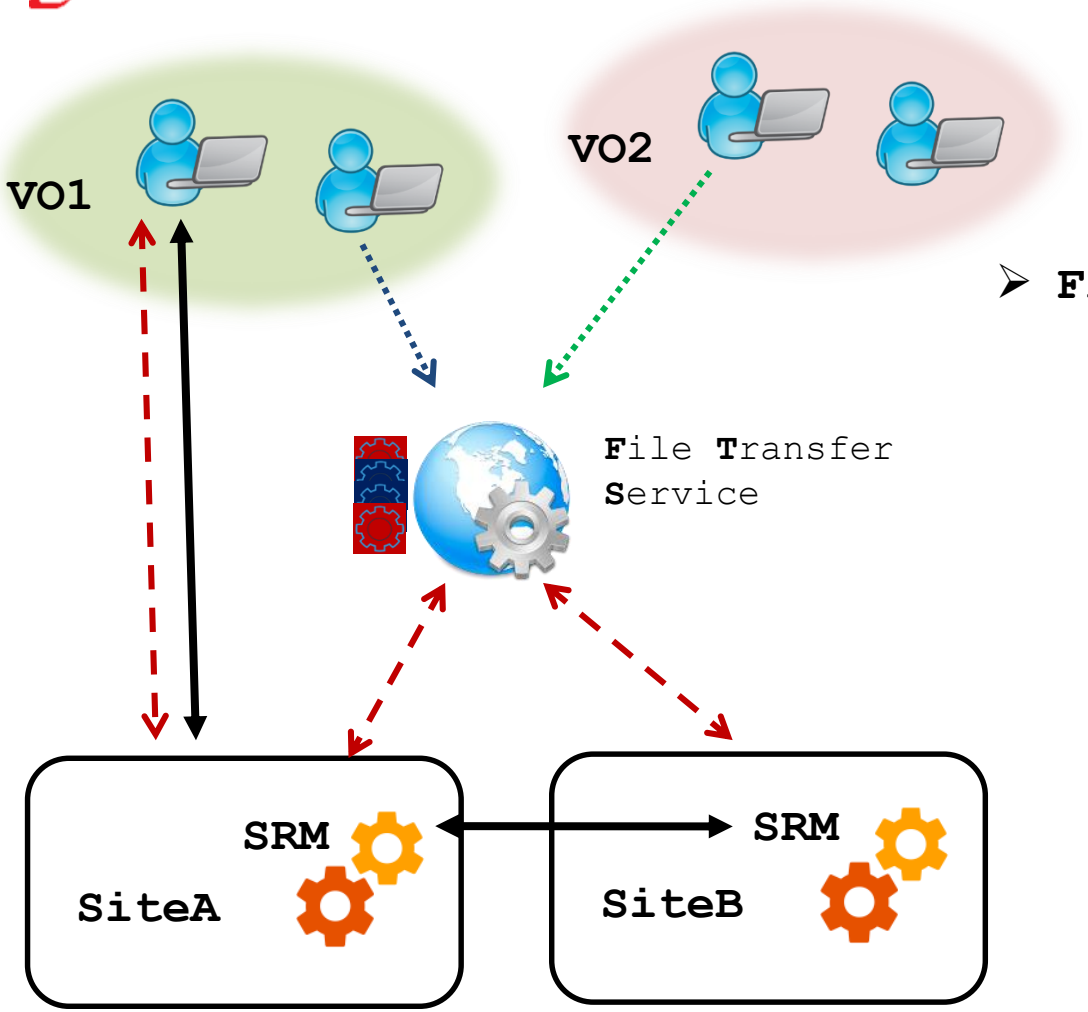


Data management.

To ease notation, from now on…



Grid jobs workflow.



(Collective Layer)

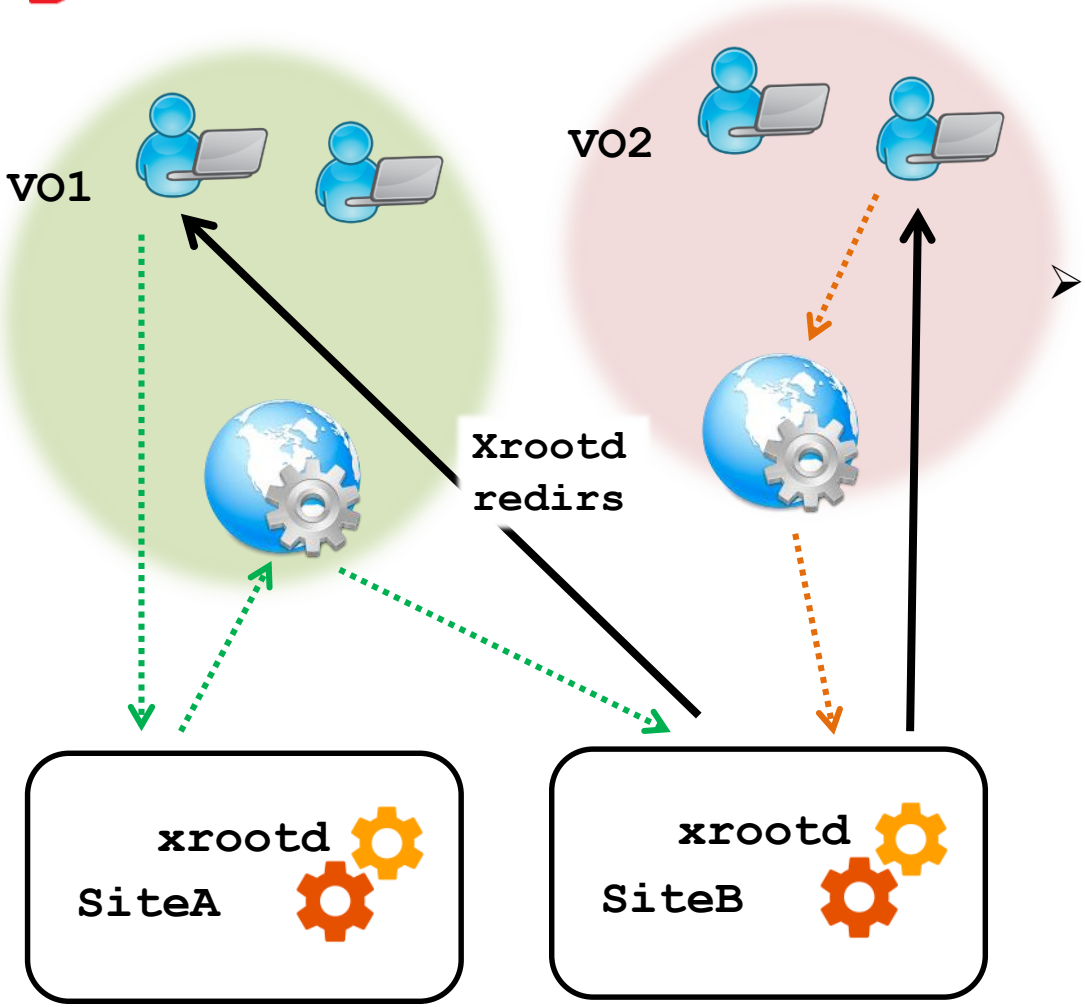
- **File Transfer Service:**
  - manages **site to site** transfers at **file level;**
  - the **user submits** a file **request to the FTS queue** and FTS deals with the SRMs to perform it;

info/requests

data

transf. req.

Grid jobs workflow.



(Collective Layer)

- **Xrootd federations**
  - global (read) **access** to files via a **hierarchy of redirectors;**
  - **integrated** in **root** and in the **exp's software;**
  - soon http and gsiftp federations as well.

data

request

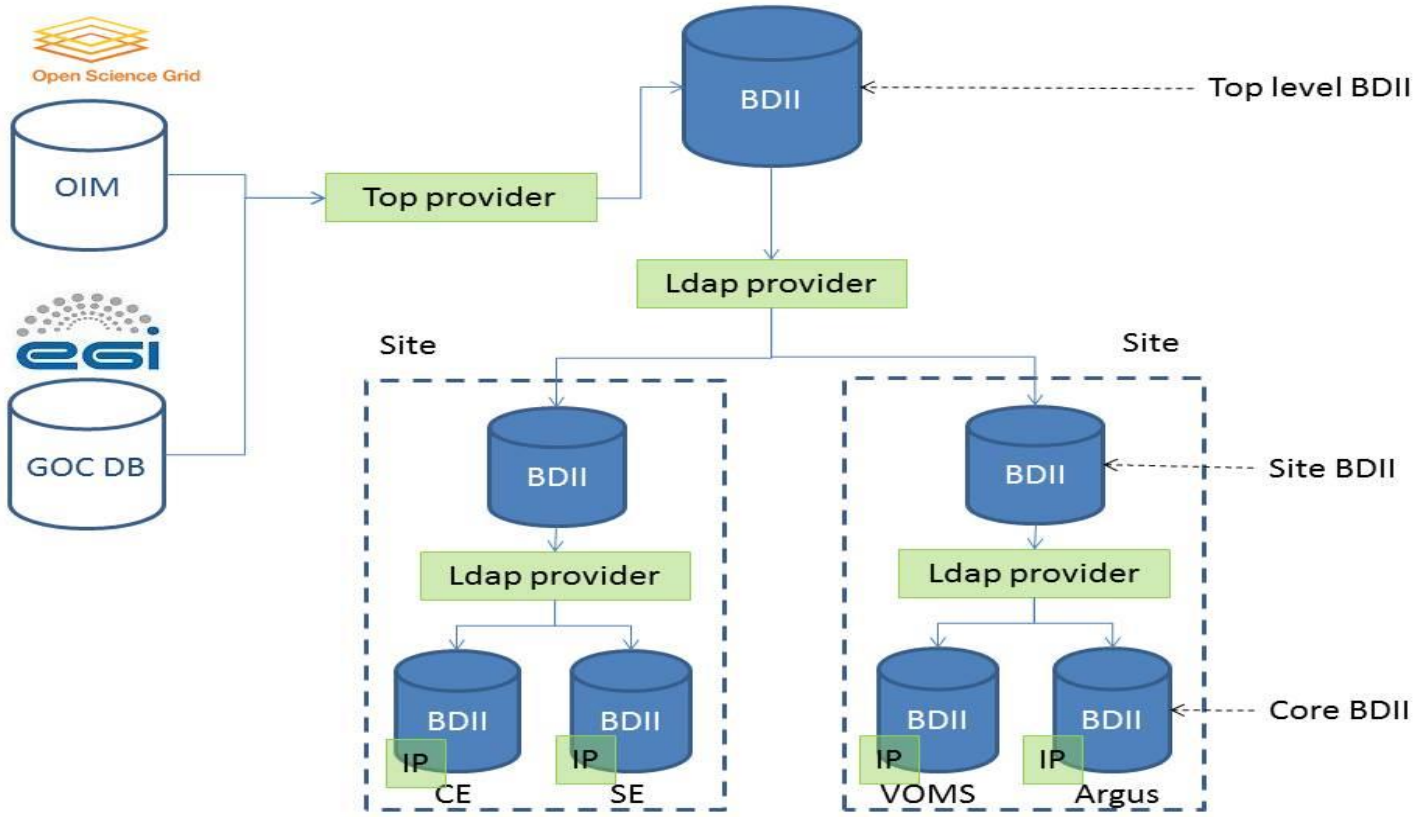
Information system.

Service	Provided by	Content	Protocol/format
REBUS	WLCG	List of WLCG federations and sites Repository for pledge information (federation level) Installed capacities (from BDII)	HTTPS, JSON+XML+CSV
GOCDB	STFC, EGI	Administrative info for EGI sites List of all site services Downtime information	HTTPS, XML
OIM	OSG	As above but for OSG	HTTPS, XML
BDII	WLCG, EGI	Detailed information for sites and services	LDAP, LDIF
Info provider	Middleware developers	Generate the information in the BDII	LDIF

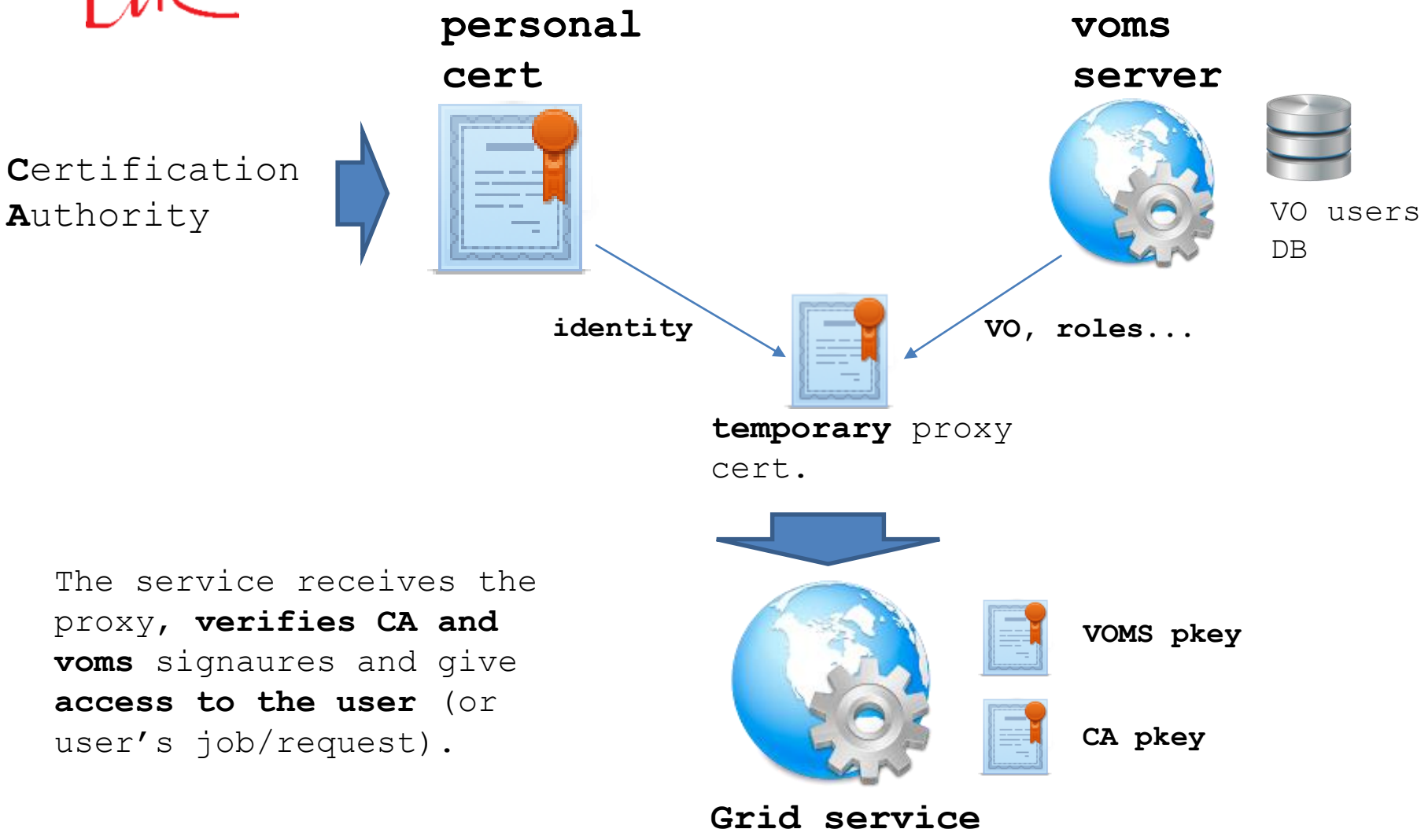
Collective

Resource

Information system.



Authentication.



The service receives the proxy, **verifies CA and voms** signaures and give **access to the user** (or user's job/request).

VO (CMS) applications.

SiteDB: CMS specific informations about sites, people and resources.

DBS, TMDB and DAS: informations about data: location of replicas, arganisation in datasets/fileblocks, physics related info (Run, lumi, etc.).

PhEDEx: site to site data transfers at dataset/fileblock level, tracking of sites occupation.

CRAB
Cms Remote Analysis Builder

CRAB: user tool for performing analysis on grid. Connects grid submission to CMS specific data management (DBS/TMDB) and site's info (SiteDB).

Questions?

Backup slides

Grid and HTC.

- **High Throughput Computing** (HTC): computing paradigm that focuses on the efficient execution of a **large number of loosely-coupled tasks** [*]
  - opposed to High Performance Computing which focuses on the fast execution of highly-coupled tasks;
  - typical metric is **jobs per month or year** (rather than FLOPS);
  - given the **minimal parallel communication** and large requirements -> large distributed resources;

- **HTC and Grid** are today often **used as synonyms**
  - in HTC we naturally have wide **heterogeneous communities using large heterogeneous distributed resources** (with complex sharing policies);
  - this is a **(THE) typical Grid Computing use case**!

[*] https://wiki.egi.eu/wiki/Glossary_V1#High_Throughput_Computing