GRID COMPUTING

Miguel Afonso Oliveira

m.a.oliveira@coimbra.lip.pt

Outline

- Part I: What is GRID Computing?
- Part II: Grid Infrastructures in Portugal
- Part III: Using GRID Computing

Outline

Part I: What is Grid Computing?

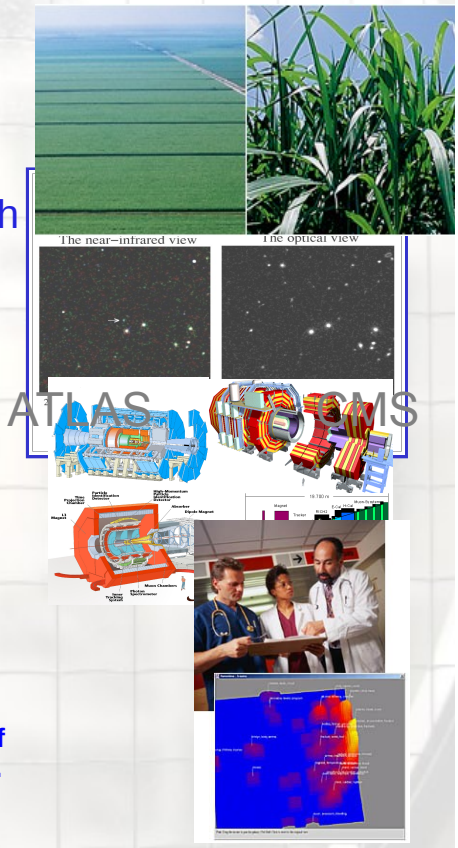
- **Introduction.**
- **European Grid Project.**
- **Grid Components: Middleware.**

Introduction

- "Grid Computing" is a concept with ~15 years:
  - **Takes distributing computing a step forward.**
- The name GRID was chosen by analogy with the "Electric Power Grid" (Ian Foster e Carl Kasselman 1999):
  - **Transparent.**
  - **Plug-in to connect to the infrastructure.**
  - **Permanent and available everywhere.**
- The GRID should provide seamless access to geographically distributed computing power and data storage.

Introduction

- Single institutions and laboratories (even the large ones) are no longer able to support the computing power and storage capacity needed for current scientific research
- Computing and data intensive sciences which are presently driving the GRID development:
  - **Physics/Astronomy:** data from different kinds of research instruments.
  - **Medical/Healthcare:** imaging, diagnosis and treatment.
  - **Bioinformatics:** study of the human genome and proteome to understand genetic diseases.
  - **Nanotechnology:** design of new materials from the molecular scale.
  - **Engineering:** design optimization, simulation, failure analysis and remote Instrument access and control.
  - **Natural Resources and the Environment:** weather forecasting, earth observation, modeling and prediction of complex systems: river floods and earthquake simulation.

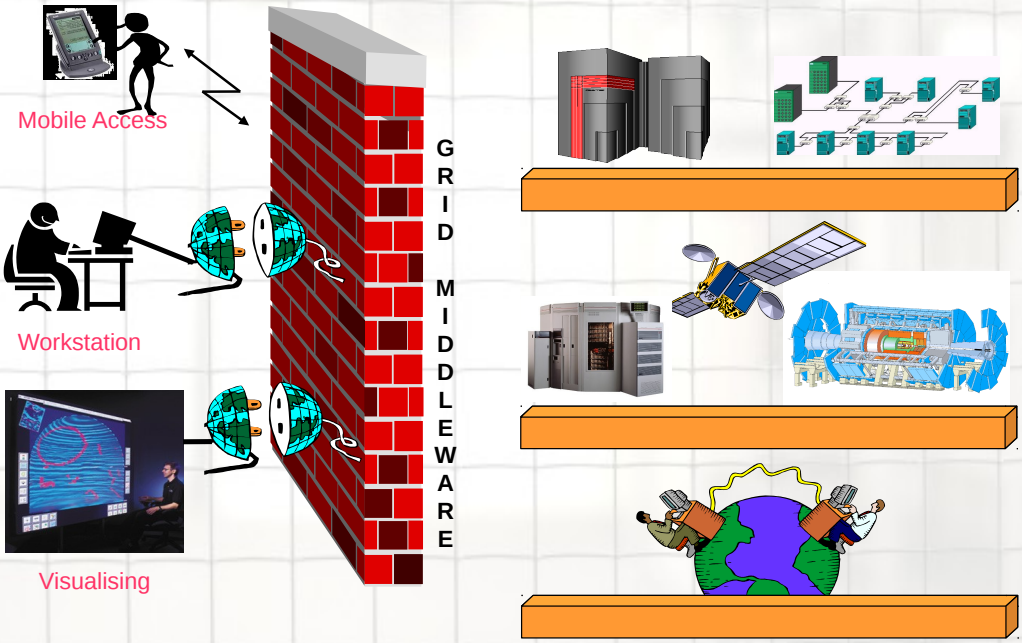


Introduction

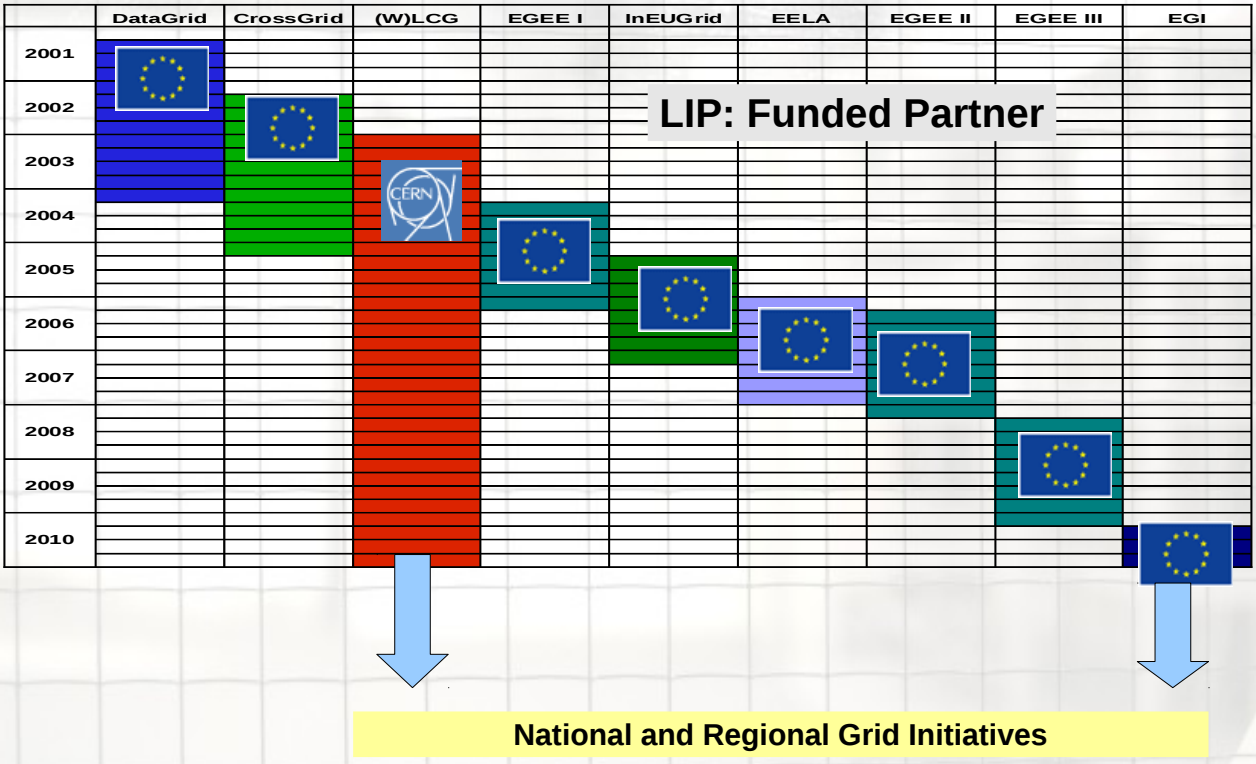
- Challenges for Grid infrastructures operation:
  - **Integration of geographically distributed resources**
  - **Resource heterogeneity:**
    - **Hardware.**
    - **Software.**
  - **Multiple administrative domains:**
    - **Authorization and access policies.**
    - **Ownership.**
    - **Security.**
  - **Multiple ways of interaction between users/applications and resources.**
  - **Dynamic nature:**
    - **Resources: addition, removal**
    - **Users/Groups.**

Introduction

The Grid Middleware is the software that should allow the virtualization and transparent access and interaction between users and applications with heterogeneous computing and storage resources which are geographically distributed.



European Grid Projects



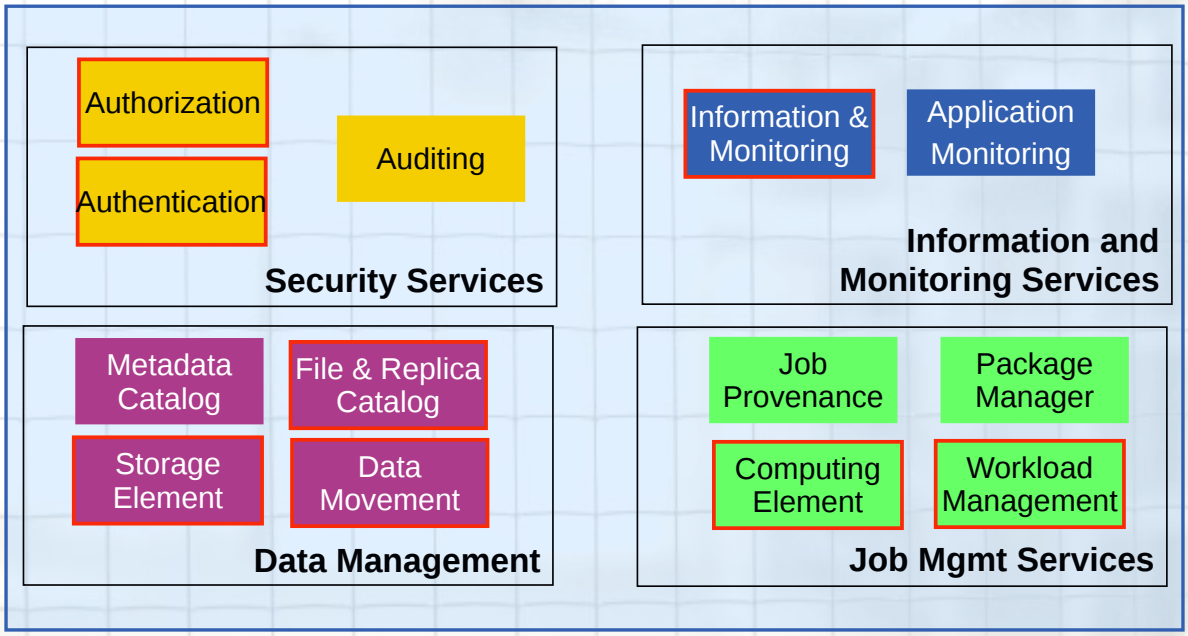
European Grid Projects

- European DataGrid (EDG)
  - **Middleware development for:**
    - **Data grids.**
    - **Large amount of computational intensive jobs.**
- World Wide LHC Computing Grid (WLCG), CERN
  - **An infrastructure for the data storage and analysis of the large LHC physics community.**
- Enabling Grid for E-sciencE (EGEE I/II/III)
  - **Multi-disciplinary European Grid infrastructure.**
- CrossGrid (CG) and Int-EU-Grid (I2G)
  - **Middleware development for:**
    - **Parallel applications**
    - **Interactive applications**

European Grid Projects

- Grid middleware (MW):
  - **gLite stack currently contains MW developed in the:**
    - **EDG, LCG, EGEE, EGI.**
    - **Globus Alliance (Globus Toolkit 4 pre-WS), through the Virtual Data Toolkit (VDT).**
    - **Condor.**
    - **MPI tools from CG and I2G, integration and support for parallel applications.**
  - **Adoption of open standards from the Open Grid Forum:**
    - **Glue Schema (v. 1.3 and 2.0): information publishing.**
    - **Storage Resource Manager (SRM v. 2.2): common set of methods to access Storage Elements.**

gLite Architecture: Main blocks



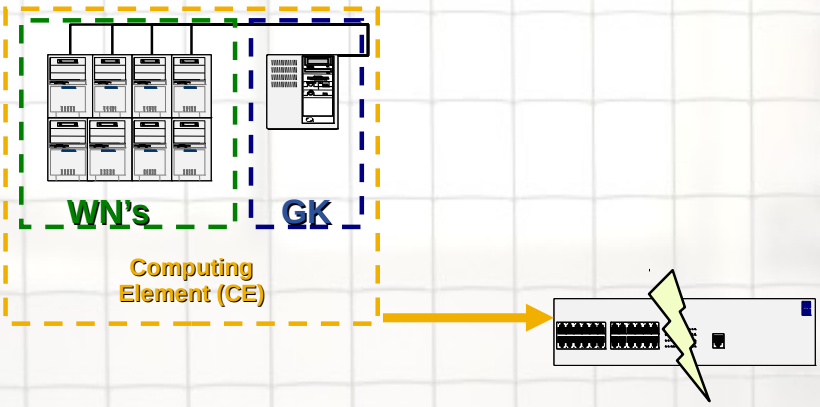
- For most of these blocks, they are associated with a set of services, which run on a given "*machine type*".
- There are two major sets of services:
  - **Local services: deployed and maintained by each site (also called Resource Center – RC).**
  - **Core services: central services installed only in some RC's, but used by all users to allow interaction with the global infrastructure.**

Grid Middleware: gLite

- Machine types for the local services:
  - Compute Element (CE).
    - Worker Nodes (WN).
  - Storage Element (SE).
  - Monitoring Box (APEL).
  - Site Berkeley-Database Information Index (Site BDII).
  - User Interface (UI).
- Machine types for the core services:
  - Virtual Organization Membership Service (VOMS).
  - Workload Management System (WMS).
  - Top-Berkeley-Database Information Index (BDII).
  - File Catalogues (FC).
  - File Transfer Services (FTS).
  - MyProxy server (PX)

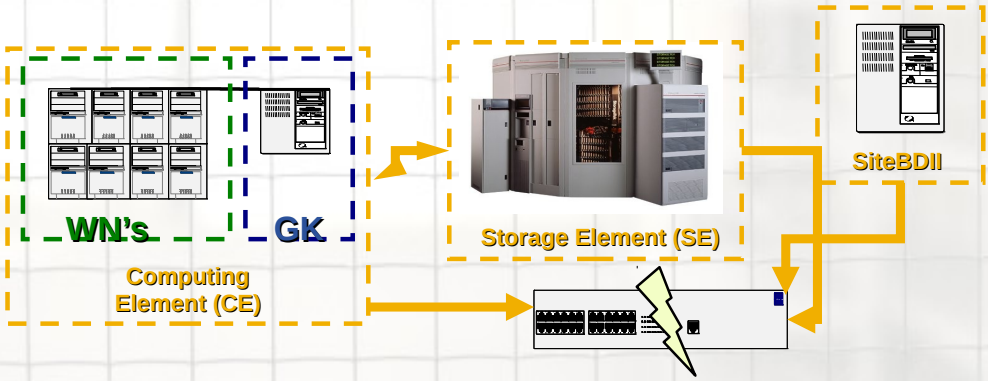
Grid Middleware: : gLite Local services

- Compute Element (CE):
  - Gatekeeper service (GK).
  - Authentication and authorization.
  - Interacts with the local batch system (PBS, TORQUE, LSF, Condor, SGE).
  - Runs a local information system publishing information regarding local resources.
- Worker Nodes (WN):
  - Where jobs are really executed.



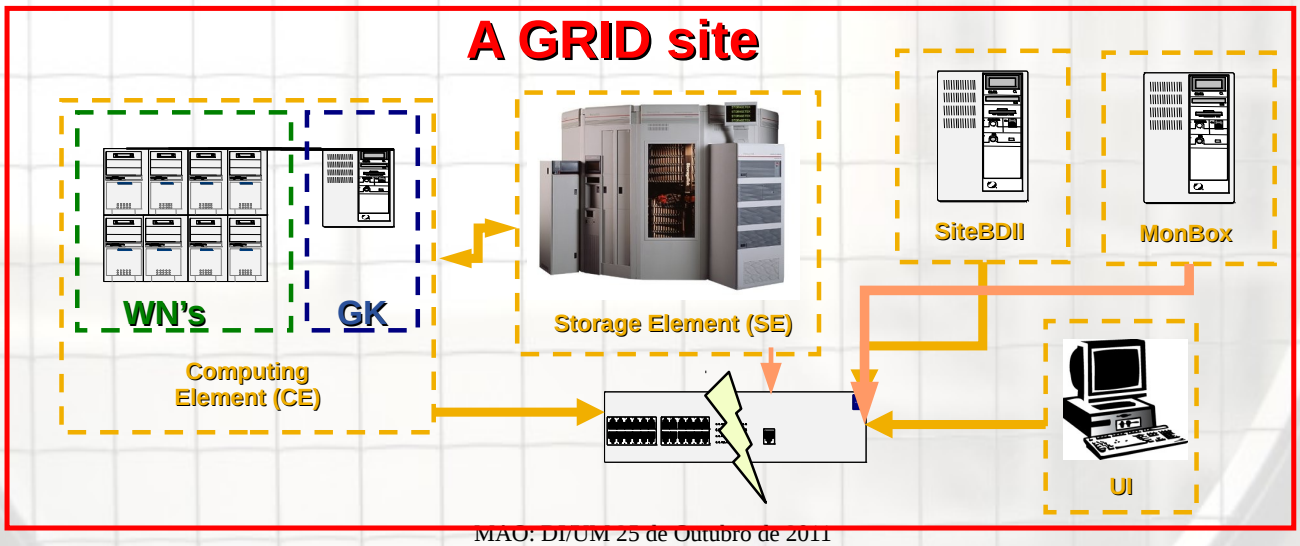
Grid Middleware: : gLite Local services

- The Storage Element (SE):
  - SRM - Storage Resource Management (StoRM, DPM, dCache and CASTOR).
  - Provides an interface for the grid user to access the local storage system (disk pools, HSM with tape backend, etc…).
  - Implements gridFTP (extension of the ftp protocol adding GSI security).
  - Runs a local information system publishing information regarding local resources.
- Site BDII:
  - Publishes information about all the local resources.



Grid Middleware: : gLite Local services

- The Monitoring Box (APEL) service:
  - Collects information given by sensors installed in the site machines, used for accounting purposes.
- User Interfaces (UI)
  - Contain client middleware tools and client libraries allowing the user to perform a large set of operations with grid resources (submission of jobs and storage and management of files).



Grid Middleware: gLite Core services

- The WMS:
  - **Implements the matchmaking process:**
    - **Match-Make the user requirements to the available resources. The status of site resources is obtained querying the top-BDII**
    - **Communicates with File Catalogues to determine in which sites files can be read or stored (if jobs require file manipulation)**
  - **Runs the Logging and Bookkeeping service (LB):**
    - **Stores the status of all the jobs submitted to the RB in a database which can be queried by the user through UI client command tools.**
- The top-BDII:
  - **Collects the information published by all the sites BDII's providing “on-line” information to other grid services.**
- PX:
  - **Service that allows the storage of users long lived proxies.**

Grid Middleware: gLite Core services

- The VOMS:
  - **Users are aggregated into Virtual Organizations (VOs).**
  - **Service containing information about wich users belong to a given VO, and more fine grained groups or roles.**
- Grid File Catalogues (FC’s):
  - **LCG File Catalog:**
    - **Is the service which maintains mappings between Logical File Names (LFNs), Global Unique Identifiers (GUID) and Physical Filenames (PFNs).**
    - **Allow the management and replication of files on the Grid.**
  - **AMGA File Catalog:**
    - **Service to store metadata information about files and groups/sets of files.**
- The FTS:
  - **Service for scheduling the transfer of large amounts of files between sites.**

But, how “big” is the GRID?

Metric		QR5 (increase from Apr 2010)
Resource Centres	EGI-InSPIRE providers	329
	Including integrated providers	346 (+6.8%)
Participating countries	EGI-InSPIRE providers	50
	With integrated providers	57 (+18.75%)
Installed computing capacity	CPU Cores EGI-InSPIRE providers	248,424
	CPU Cores also including integrated and peer providers	337,608
	Resource Centres supporting MPI	93
	Installed capacity (HEP-SPEC 061)	1.93 million
Installed storage capacity	Disk (PB)	106.7
	Tape (PB)	112.8
Usage 2010-2011	Jobs	949,000
	Wall clock hours	3.2 million hours/day (+86.5%)
	HEP-SPEC 06 CPU wall clock hours	25.7 million hours/day

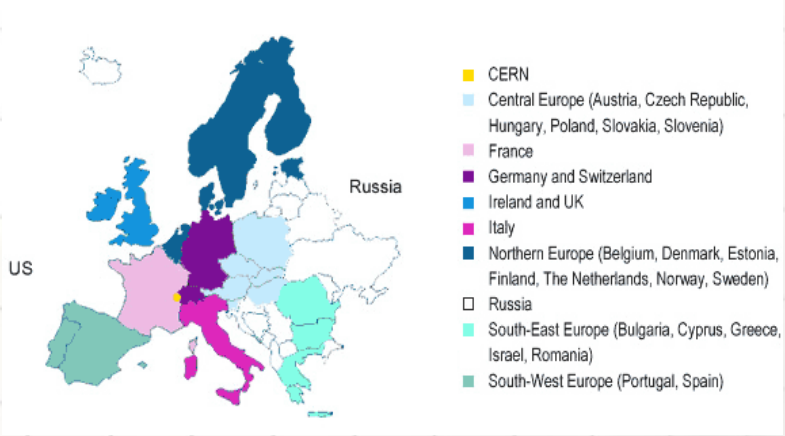
Grid Infrastructures in Portugal

Part II: Grid Infrastructures in Portugal

- **Sites integrated in the IBERGRID project.**
- **LIP Lisboa and LIP Coimbra.**
- **“Iniciativa Nacional Grid” (INGRID):**
  - **“Nó Central Grid” (NCG).**
- **“Iberian Grid” (IBERGRID).**
- **Certification Authority**

Sites integrated in the EGI project

- Several Grid infrastructures, **EGI**, EELA, Int-EU-Grid.
- Up to 2005 LIP was the only institution participating in Grid European projects mentioned before.
- Since 2005/06 several Portuguese institutes and Universities have shown interest in integrating their clusters into the EGI infrastructure.
- In the EGI context, Portuguese sites are part of the IBERGRID federation, together with the Spanish sites.



Sites integrated in the IBERGRID project

Normalised CPU time [units 1K.SI2K.Hours] by SITE and VO							
SITE	atlas	auger	biomed	cms	Other VOs	Total	%
CFP-IST	0	0	0	0	5,215	5,215	0.03%
IEETA	0	0	3,575	0	1,166	4,741	0.03%
LIP-Coimbra	2,001,466	0	74,437	0	40,484	2,116,387	11.42%
LIP-Lisbon	2,981,769	63,585	12,631	1,542,168	14,028	4,614,181	24.90%
NCG-INGRID-PT	4,677,237	3,461,672	367,013	2,553,706	719,695	11,779,323	63.56%
UMinho-CP	0	0	0	0	4,744	4,744	0.03%
UPorto	0	0	6,502	0	1,671	8,173	0.04%
Total	9,660,472	3,525,257	464,158	4,095,874	787,003	18,532,764	
Percentage	52.13%	19.02%	2.50%	22.10%	4.25%		

Dep. Informática - Univ. Minho
Univ. Porto
IEETA – Univ. Aveiro
Centro de Física de Plasmas – IST
LIP Lisboa
LIP Coimbra
NCG

LIP Sites

- PT Tier2 for the Atlas and CMS experiments @ LHC:
  - **3 sites:**
    - **LIP Lisbon.**
    - **LIP Coimbra.**
    - **Central Grid Node (NCG) - LNEC campus.**
  - **LIP sites:**
    - **Computing and storage resources are used by local and Grid users.**
    - **CE batch system:**
      - **Lisbon/NCG – Grid Engine.**
      - **Coimbra – Torque/Maui.**
    - **SE technology – StoRM SRM and Lustre, DPM for some VO's.**

LIP Lisboa and LIP Coimbra

- LIP Lisboa:
  - **WN's: 400 CPU cores**
  - **Storage: 150TB**
  - **Grid local services:**
    - **3 CE's**
    - **1 StoRMSRM**
    - **2 GridFTP servers**
    - **1 APEL**
    - **1 Site-BDII**
  - **Grid core services:**
    - **1 WMS**
    - **1 top-BDII**
    - **1 PX server**
    - **1 LFC**
- LIP Coimbra:
  - **WN's: 180 CPU cores**
  - **Storage: 75TB**
  - **Grid local services:**
    - **2 CE**
    - **1 StoRMSRM**
    - **2 GridFTP servers**
    - **1 APEL**
    - **1 Site-BDII**

NCG

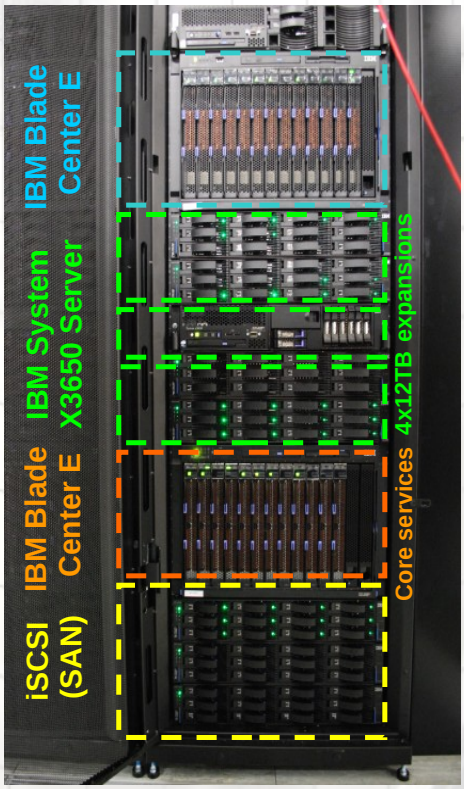
- NCG:
  - **WN's: 1000 CPU cores**
  - **Storage: 300TB**
  - **Grid local services:**
    - **3 CE's**
    - **1 StoRMSRM**
    - **2 GridFTP servers**
    - **1 APEL**
  - **Grid core services:**
    - **1 WMS**
    - **1 top-BDII**
    - **1 PX server**
    - **1 LFC**

INGRID

- INGRID: Iniciativa Nacional GRID
  - **The Portuguese NGI.**
  - **Push for resource sharing in Portugal.**
  - **Copes with the worldwide interest in Grid Computing.**
  - **Follows the path of European Grid Initiative (EGI) aiming to interoperate between different European National Grid Initiatives (NGIs).**
  - **Helps to fulfill the Portuguese responsibilities in the framework of present European Projects:**
    - **EGI, WLCG, IBERGRID...**
- INGRID Management Committee:
  - **FCT, UMIC and LIP.**
  - **LIP is in charge of the INGRID technical coordination.**
- "Nó Central Grid" (NCG)(**LNEC**/FCCN/LIP/UMIC).
  - **10 Gbps fibre link: Excellent connection to the national network backbone and to the GEANT PoP**

INGRID

- Nearline robotic storage:
  - **Grid accessible data repositories.**
  - **Hierarchical storage.**
- Core grid services:
  - **Server direct attached storage.**
- Grid cluster.
- Online grid storage.
- Internal networking:
  - **Non blocking layer2 / layer3 central switch**
  - **All ports can operate simultaneously at maximum speed**
  - **Aggregates the fibre uplink, the machines fibres, the copper Ethernet**



INGRID: Computing resources

- High Throughput Computing Servers:
  - **Several IBM bladecentres E:**
    - **2 quad-core AMD opteron 2356 processors @ 2.3 GHz**
- High Performance Computing Servers:
  - **IBM bladecentre H with infiniband switch.**
    - **2 quad-core Intel(R) Xeon(R) L5420 CPUs @ 2.50 GHz.**
    - **Blades with 20 Gbps Double Data Rate Host Channel Adaptors for data transmission between processors and I/O devices.**
    - **Blades, with 4 GB of RAM per core.**
    - **Running Scientific Linux 5 (SL5) x86_64 architecture.**
- Local Resource Management System: Grid Engine

INGRID: Storage resources

- Storage servers with expansion boxes:
  - **Several IBM X3650 servers running SL5 x86_64:**
    - **2 quad-core Intel(R) Xeon(R) L5420 CPUs @ 2.50 GHz.**
  - **Each server has associated:**
    - **2 LSI Mega Raid Controllers connected to the expansion boxes.**
    - **Expansion boxes in Raid 5 Volumes with 1 TB SATA-II disks.**
  - **Storage Network interfaces**
    - **Two 10/100/1000 BASE-T Broadcom Gigabit Ethernet.**
    - **One NetXen 10 Gigabit Ethernet PCIe controller.**
- Grid access enabled via the StoRM SRM interface:
  - **Lustre as underlying filesystem.**

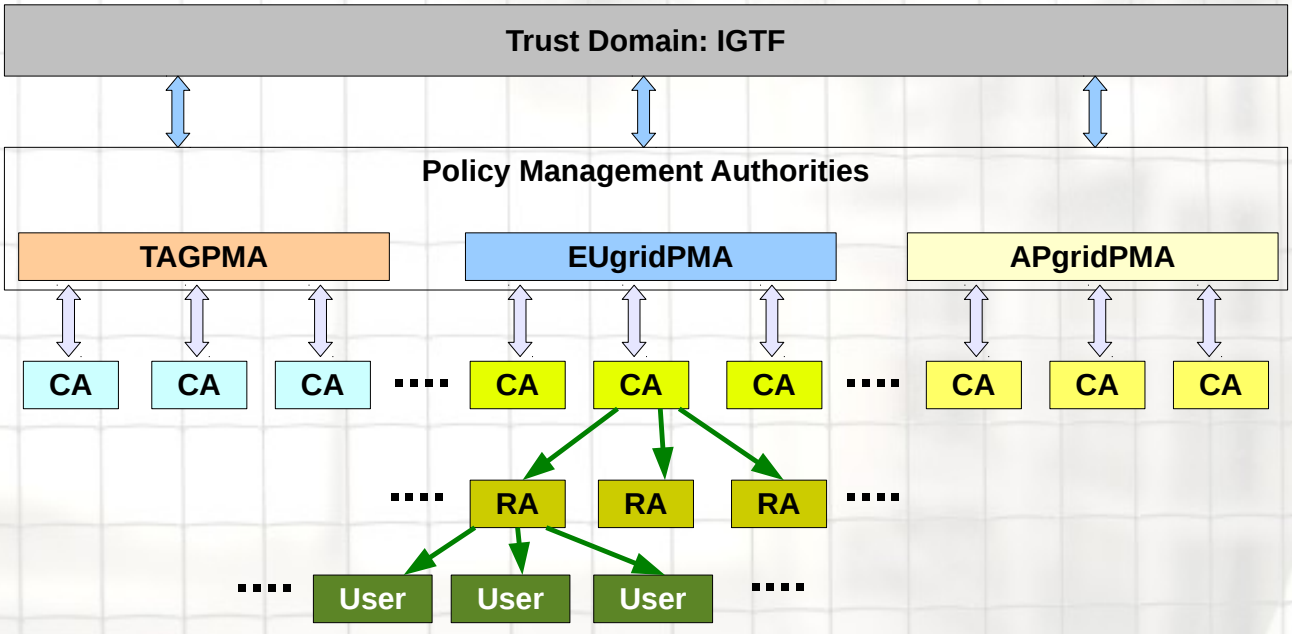
IBERGRID

- The Iberian Grid infrastructure joins the Portuguese and Spanish NGIs.
  - **The aim is to foster and simplify the scientific collaboration between researchers of both countries.**
  - **Enable common participation in EGI.**
  - **The NCG is the “home” for INGRID and also for Ibergrid.**
    - **A VOMS server is in place to hold INGRID and IBERGRID VO's.**
    - **Grid core services are operational.**
  - **Some of the base scientific areas where common research groups were identified:**
    - **Environment risk control.**
    - **Civil protection and emergency response.**
    - **Biomedical sciences.**
    - **etc,...**

Certification Authority

- There is usually one Grid CA per country or very large organization:
  - Each CA issues certificates for grid users and services within its geographical or administrative scope
  - To establish global grids a common trust domain had to be established
- The International Grid Trust Federation (IGTF) is the body that manages a global trust domain for grid computing serving the biggest grid infrastructures worldwide:
  - The IGTF is split in three regional Policy Management Authorities.
    - EugridPMA → Europe
    - ApgridPMA → Asia Pacific
    - TAGPMA → Americas
  - The three PMAs make use the Open Grid Forum to establish community policy and best practices for all PMAs.
- The Portuguese “Certification Authority” (CA):
  - Is operated by LIP.
  - Each Institute/University/Department/Laboratory should preferably, setup and operate a Registration Authority (RA).

Certification Authority



MAO: DI/UM 25 de Outubro de 2011

Certification Authority

http://ca.lip.pt

Certification Authority: sites

- For a new site:
  - **Contact: grid.support@lip.pt**
  - **Creation of a new Registration Authority (RA).**
  - **Deployment of the following “node types”, gLite 3.2.0:**
    - **Compute Element (CE)**
      - **Worker Nodes**
    - **Storage Element (SE)**
    - **APEL**
    - **Site-BDII**
    - **User Interface**
  - **Support “operations” VO (certification and monitoring).**

Certification Authority: users

- Users need:
  - **Single sign-on: the ability to logon to a machine and have the user's identity passed to other resources as required.**
  - **To trust owners of the resources they are using.**
- Providers of resources (computers, databases,...) need to:
  - **Trust users they do not know.**
  - **Minimize impact on security.**
  - **Have the ability to trace who did what.**
- The solution comes from:
  - **Digital Certificates.**
  - **Virtual Organizations.**

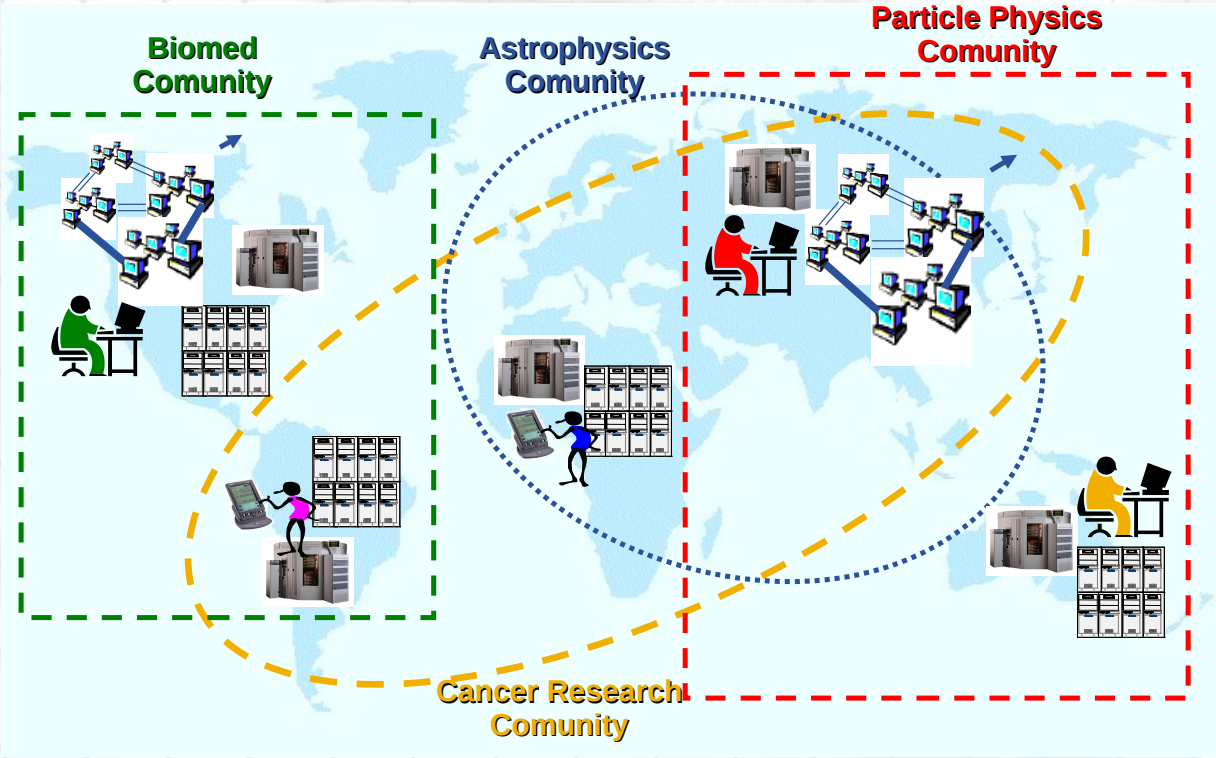
Certification Authority: users

- Resource providers are "opening themselves up" to itinerant users:
  - **A Secure Access to resources is provided through the X.509 Public Key Infrastructure.**
  - **Digital certificates identify uniquely its user/service identity.**
- Users/Services identity have to be certified by (mutually recognized) national CAs.
- Digital certificates allow a temporary delegation from users to processes executed "in user's name" (proxy certificates and myproxy certificates repositories).

Certification Authority: VOs

- Virtual Organizations (VO):
  - **People from different organizations but with common goals get together to solve their problems in a cooperative way.**
    - **Virtualized shared computing resources: VO members have access to computing resources outside their home institutions.**
    - **Virtualized shared data resources: VOs members can store and access data outside their home institutions.**
    - **Other resources may be shared and virtualized as well: Instruments, sensors, software and even people…**
- VO practical role:
  - **Set Common Agreed Policies for accessing resources**
  - **Administrates the VO membership list**
    - **Before joining a VO, a person must already have a valid certificate**
    - **VO administrators can reject persons which do not fulfill the VO policies.**

Certification Authority: VOs



MAO: DI/UM 25 de Outubro de 2011

Certification Authority:VOs

https://cic.egi.eu/

Certification Authority: VOs

voms admin for VO: **fusion** Current user: **Miguel Afonso Oliveira**

-->

Welcome to voms-admin registration for the **fusion** VO.

To access the VO resources, you must agree to the VO's Usage Rules. Please fill out all fields in the form below and click on the submit button at the bottom of the page.

After you submit this request, you will receive an email with instructions on how to proceed. Your request will not be forwarded to the VO managers until you confirm that you have a valid email address by following those instructions.

IMPORTANT:

By submitting this information you agree that it may be distributed to and stored by VO and site administrators. You also agree that action may be taken to confirm the information you provide is correct, that it may be used for the purpose of controlling access to VO resources and that it may be used to contact you in relation to this activity.

Your distinguished name (DN):

/C=PT/O=LIPCA/O=LIP/OU=Coimbra/CN=Miguel Afonso Oliveira

Your CA:

/C=PT/O=LIPCA/CN=LIP Certification Authority

Your email address:

Your institute:

Your phone number:

Comments for the VO admin:

☐ You agree on the VO's usage rules.

Register!

Voms-Admin version 2.0.15

Part III: Using GRID Computing

- **Authentication**
- **Resource discovery**
- **Job submission**
- **Job Description Language (JDL)**
- **Advanced job types**
- **Data management**

1. getting your certificate

a) create a certificate request: start at http://ca.lip.pt/

b) a responsible person at your institute attests that the certificate belongs to you.

2. request your VO-membership (VO-dependent)

3. export your certificate (using Firefox)

a) receive your certificate from GridKa-CA via your browser

b) go to the Preferences

c) under “Advanced” go to “Certificates” and click “Manage Certificates…”

d) select your certificate and click “Backup”

e) enter a name for the backup file, e.g. backup.p12

f) enter the passphrase for the System Security Device

g) enter a new password to protect the backup (very important!!!)

Important: use the same browser, on the same computer, for each of the previous steps!

Authentication

4.set-up your environment

- copy backup.p12 to your portal machine
- extract your keys from the backup

```
gks> mkdir $HOME/.globus/

gks> chmod 755 $HOME/.globus/

gks> openssl pkcs12 -nocerts -in backup.p12 -out $HOME/.globus/userkey.pem

gks> openssl pkcs12 -clcerts -nokeys -in backup.p12 -out \
        $HOME/.globus/usercert.pem

gks> chmod 400 $HOME/.globus/userkey.pem

gks> chmod 644 $HOME/.globus/usercert.pem

gks> ls -l $HOME/.globus/
total 8
-rw-r--r--  1 john cms 1846 Aug  6  2007 usercert.pem
-r--------  1 john cms 1743 Aug  6  2007 userkey.pem
```

public key

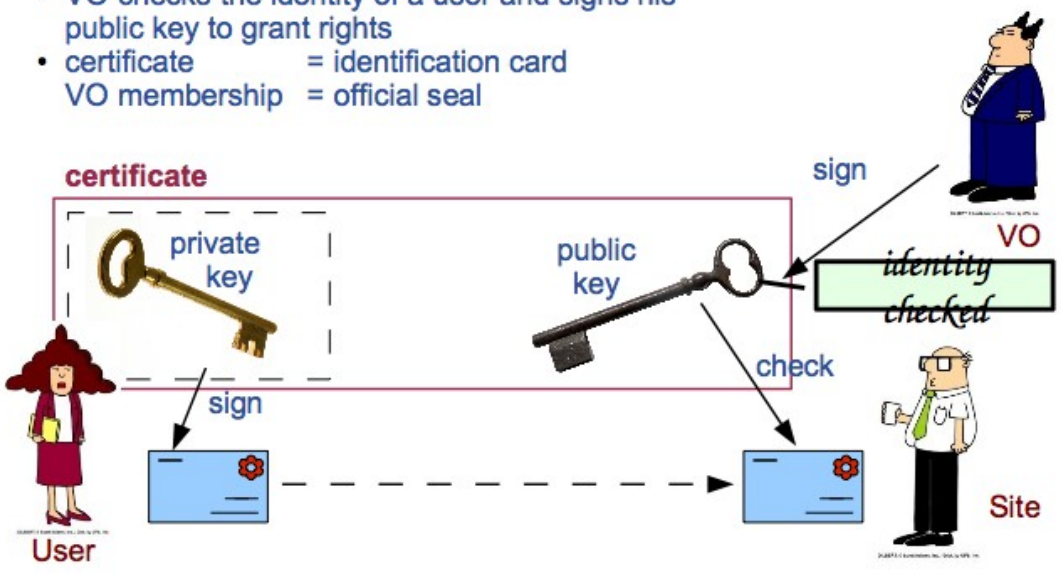
private key

- now you are ready

Authentication

How is the authentication done?

- authentication uses public key cryptography
- VO checks the identity of a user and signs his public key to grant rights
- certificate = identification card
  VO membership = official seal



MAO: DI//UM 25 de Outubro de 2011

Authentication

Remarks:

* keep your private key private.
* key sharing is absolutely forbidden.
* your tests certificates are connected to your identity.

But how does a job authenticate itself? Via a Proxy.

* A proxy is a temporary key pair with limited lifetime.
* It has the same rights as your persistent certificate.
* Proxies can be requested with a special role.

Authentication

- create a proxy with `voms-proxy-init [-voms <vo>[:/<vo>/Role=<role>]] [-valid hh:mm]`

```
gks> voms-proxy-init -voms dech -valid 10:00
Enter GRID pass phrase:
Your identity: /C=DE/O=GermanGrid/OU=Uni Karlsruhe/CN=John Doe
[...]
Your proxy is valid until Wed Jul  2 00:48:29 2008
```

- check your current proxy with `voms-proxy-info [-all]`

```
gks> voms-proxy-info
subject   : /C=DE/O=GermanGrid/OU=Uni Karlsruhe/CN=John Doe/CN=proxy
issuer    : /C=DE/O=GermanGrid/OU=Uni Karlsruhe/CN=John Doe
identity  : /C=DE/O=GermanGrid/OU=Uni Karlsruhe/CN=John Doe
type      : proxy
strength  : 512 bits
path      : /tmp/x509up_u12039
timeleft  : 9:58:33
```

- destroy your proxy with `voms-proxy-destroy`

```
gks> voms-proxy-destroy
```

MAO: DI//UM 25 de Outubro de 2011

Resource discovery

- Every site publishes information on the resources it provides to the GRID.
- A user can always query this subsystem to discover the resources it needs:

lcg-infosites --vo <vo> <option> -v <verbosity> -f <site> --is <bdii>

--vo <vo>: name of the VO to which the information to print is relevant.

<option>: specify what information to print (ce,se,all,closeSE,tag,etc).

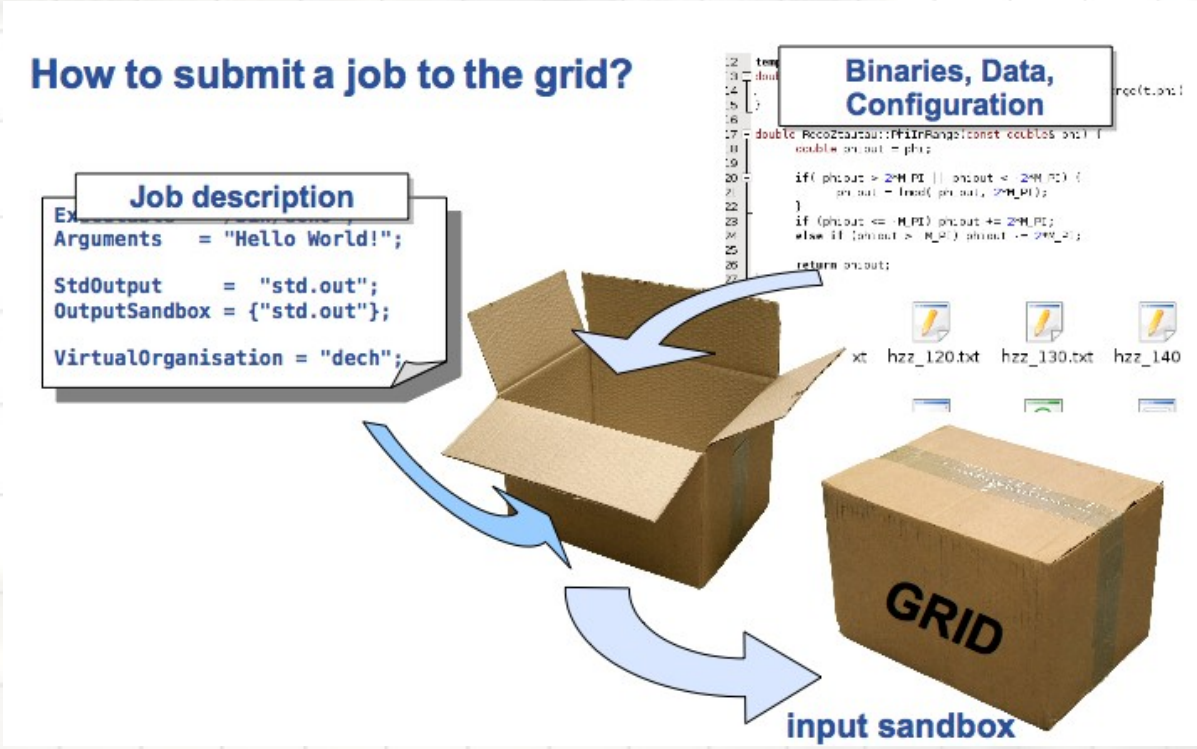
--is <bdii>: the BDII to query. If not defined, the BDII defined in the env.

Resource discovery

```
[miguel@ui1 ~]$ lcg-infosites --vo swetest ce
#CPU    Free    Total Jobs      Running Waiting ComputingElement
----------------------------------------------------------
1452      42        32             3      29    ce06.pic.es:2119/jobmanager-lcgpbs-gshort64
1452      44        32             3      29    ce07.pic.es:2119/jobmanager-lcgpbs-gshort64
1452      42       582           518      64    ce06.pic.es:2119/jobmanager-lcgpbs-gmedium64
  62       2         0             0       0    ce01.ific.uv.es:2119/jobmanager-pbs-short
1380      12        32             3      29    ce05.pic.es:2119/jobmanager-lcgpbs-gshort64
1452      44       588           517      71    ce07.pic.es:2119/jobmanager-lcgpbs-gmedium64
1380      12      1201           846     355    ce05.pic.es:2119/jobmanager-lcgpbs-glong64
1380      12       588           517      71    ce05.pic.es:2119/jobmanager-lcgpbs-gmedium64
   1       1         0             0       0    ce.egee.cesga.es:2119/jobmanager-lcgsge-swetest
1611    1172         2             2       0    egeece02.ifca.es:2119/jobmanager-lcgpbs-infiniband
1452      44      1202           846     356    ce07.pic.es:2119/jobmanager-lcgpbs-glong64
   8       6         0             0       0    lcg2ce.ific.uv.es:2119/jobmanager-pbs-short
1452      42      1189           847     342    ce06.pic.es:2119/jobmanager-lcgpbs-glong64
  34      34         0             0       0    ce01.up.pt:2119/jobmanager-lcgsge-swetest
1611    1172         0             0       0    egeece02.ifca.es:2119/jobmanager-lcgpbs-infinibandlarge
  32      21         0             0  444444    ce-egee.bifi.unizar.es:2119/jobmanager-lcgpbs-swetest
   8       4         0             0       0    lcg2ce.ific.uv.es:2119/jobmanager-pbs-swetest
1611    1172         0             0       0    egeece01.ifca.es:2119/jobmanager-lcgpbs-swetest
   8       4         0             0       0    lcg2ce.ific.uv.es:2119/jobmanager-pbs-swetestL
  22      11         0             0       0    grid001.fe.up.pt:2119/jobmanager-lcgsge-swetest
1611    1172         0             0       0    egeece01.ifca.es:2119/jobmanager-lcgpbs-infinibandlarge
1611    1172         2             2       0    egeece01.ifca.es:2119/jobmanager-lcgpbs-infiniband
  12      11         0             0       0    axon-g01.ieeta.pt:2119/jobmanager-lcgpbs-swetest
   9       9         0             0       0    golp-ce.ist.utl.pt:2119/jobmanager-lcgpbs-swetest
  62       0         0             0       0    ce01.ific.uv.es:2119/jobmanager-pbs-swetest
   4       4         0             0       0    lcg-ce.usc.cesga.es:2119/jobmanager-lcgpbs-swetest
  22      10         0             0       0    ramses.dsic.upv.es:2119/jobmanager-lcgpbs-swetest
  62       0         0             0       0    ce01.ific.uv.es:2119/jobmanager-pbs-swetestL
 160       0         0             0       0    grid006.gridc.lip.pt:2119/jobmanager-lcgpbs-swetest
  16      16         0             0       0    ce.cp.di.uminho.pt:2119/jobmanager-lcgpbs-swetest
1611    1172         0             0       0    egeece02.ifca.es:2119/jobmanager-lcgpbs-swetest
  12       3         0             0       0    ce02.lip.pt:2119/jobmanager-lcgsge-swetestgrid
 294     294         0             0       0    lcg-ce2.usc.cesga.es:2119/jobmanager-lcgpbs-swetest
1611    1172         0             0       0    egeece03.ifca.es:2119/jobmanager-lcgpbs-swetest
 337     337       209           181      28    ce2.egee.cesga.es:2119/jobmanager-lcgsge-GRID_swetest
  16      10         0             0       0    grid001.fc.up.pt:2119/jobmanager-lcgsge-swetest
 337     337       209           181      28    ce3.egee.cesga.es:2119/jobmanager-lcgsge-GRID_swetest
[miguel@ui1 ~]$
```

MAO: DI//UM 25 de Outubro de 2011

Job Submission



Job submission

- the JDL file specifies the needs of your job (e.g. input files), the output files to be copied back and the requirements concerning the computing resources on remote sites
- just a simple JDL file for the beginning (`hello_world.jdl`)

```
Executable    = "/bin/echo";
Arguments     = "Hello World!";

StdOutput     = "std.out";
StdError      = "std.err";
OutputSandbox = {"std.out","std.err"};
```

the file and the argument that is executed on the worker node

standard and error output of the executable are stored

the files with the output will be copied back in the output sandbox (everything else will be deleted)

Job submission

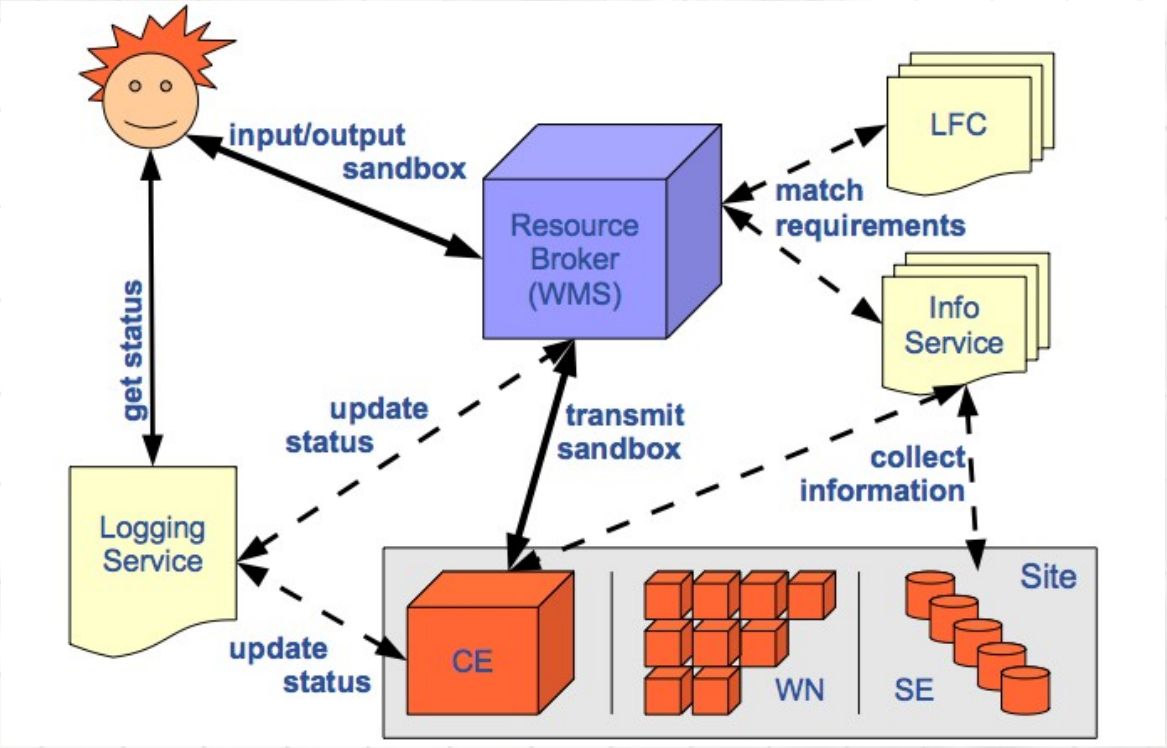
submit a job using glite-wms-job-submit

```
gks> glite-wms-job-submit -a hello_world.jdl

Connecting to the service https://grid-wms1.desy.de:7443/glite_wms_wmproxy_server


==================== glite-wms-job-submit Success ====================

The job has been successfully submitted to the WMProxy
Your job identifier is:

https://grid-lb0.desy.de:9000/cjajCLKfG3J8cgu_SX5phg

==========================================================
```

this is the unique id of this job

- you get a job identifier which you use for subsequent commands
- the option "-a" will be explained later

Job Submission



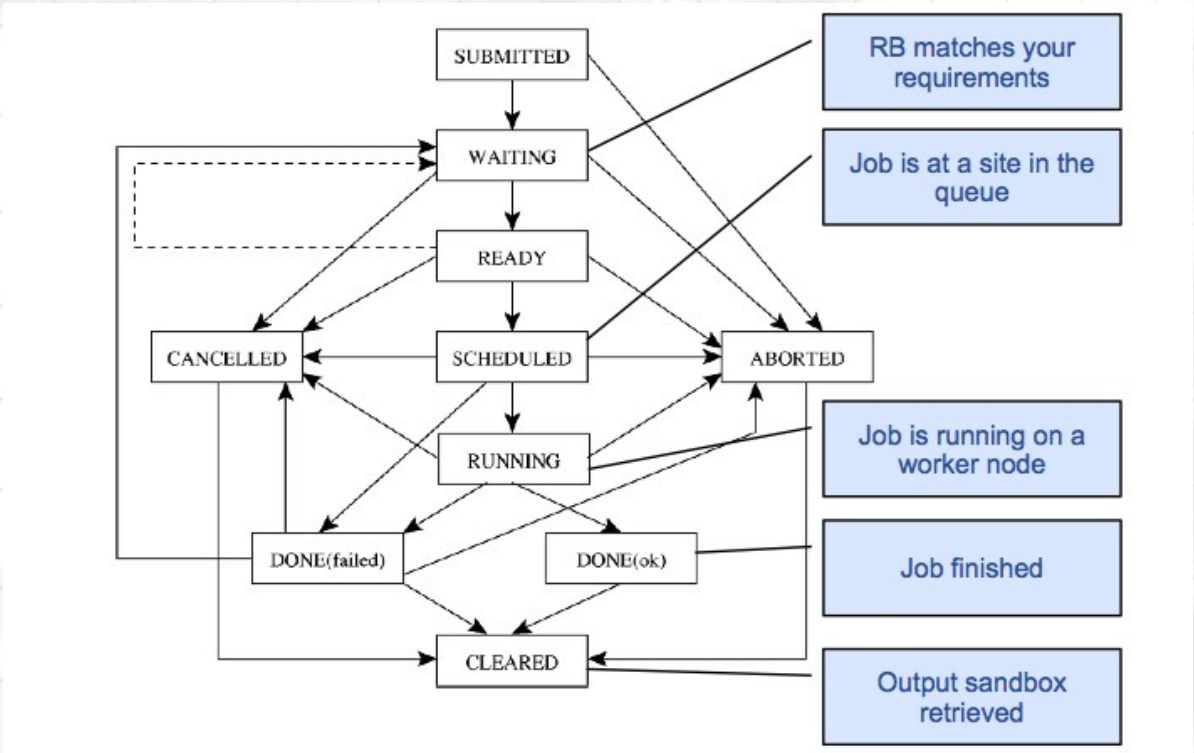
Job Submission

Check the status of your job with `glite-wms-job-status`

```
gks> glite-wms-job-status https://grid-lb0.desy.de:9000/cjajCLKfG3J8cgu_SX5phg

*************************************************************
BOOKKEEPING INFORMATION:

Status info for the Job : https://grid-lb0.desy.de:9000/cjajCLKfG3J8cgu_SX5phg
Current Status:     Scheduled
Status Reason:      Job successfully submitted to Globus
Destination:        udo-ce03.grid.tu-dortmund.de:2119/jobmanager-lcgpbs-dech
Submitted:          Thu Jul  3 12:01:36 2008 CEST
*************************************************************
```

- you can also pass several job identifiers

Job Submission



Job Submission

Retrieve the result of your job with `glite-wms-job-output`

```
gks> glite-wms-job-output --dir out \
        https://grid-lb0.desy.de:9000/cjajCLKfG3J8cgu_SX5phg

Connecting to the service https://grid-wms1.desy.de:7443/glite_wms_wmproxy_server


================================================================================

                        JOB GET OUTPUT OUTCOME

Output sandbox files for the job:
https://grid-lb0.desy.de:9000/cjajCLKfG3J8cgu_SX5phg
have been successfully retrieved and stored in the directory:
/home/johndoe/gks/out

================================================================================

gks> cat out/std.out
Hello World!
```


Job Submission

Overview

1) **`glite-wms-job-submit`** submits your job

```
glite-wms-job-submit -a file.jdl
```

2) **`glite-wms-job-status`** checks the status of your job

```
glite-wms-job-status https://<server>:9000/<id>
```

3) **`glite-wms-job-output`** retrieves the output of your job

```
glite-wms-job-output --dir <directory> https://<server>:9000/<id>
```

4) **`glite-wms-job-cancel`** cancels your job

```
glite-wms-job-cancel https://<server>:9000/<id>
```

Job Description Language

Use the input sandbox to send a program to the grid

get_info.jdl

```
Executable    = "job.sh";
Arguments     = "Hello World!";

StdOutput     = "std.out";
StdError      = "std.err";
InputSandbox  = {"job.sh"};
OutputSandbox = {"std.out", "std.err"};
```

job.sh

```
#!/bin/bash
echo "Arguments: $@"
echo "I am `whoami` on host `hostname`"
id
pwd
ls -la
cat /proc/cpuinfo
uname -a
echo $VO_DECH_SW_DIR
ls -lh  $VO_DECH_SW_DIR
```


Job Description Language

- use `glite-wms-job-list-match -a <JDL file>` to see all sites that match the requirements (for testing)
- see where the job file could run (for testing):

```
gks> glite-wms-job-list-match --rank -a hello_world.jdl

Connecting to the service https://grid-wms1.desy.de:7443/glite_wms_wmproxy_server

================================================================================

                    COMPUTING ELEMENT IDs LIST
 The following CE(s) matching your job requirements have been found:

        *CEId*                                                  *Rank*

 - ce-1-fzk.gridka.de:2119/jobmanager-pbspro-dech                  0
 - ce-2-fzk.gridka.de:2119/jobmanager-pbspro-dech                  0
 - ce-3-fzk.gridka.de:2119/jobmanager-pbspro-dech                  0
 - ce-4-fzk.gridka.de:2119/jobmanager-pbspro-dech                  0
 - ce-5-fzk.gridka.de:2119/jobmanager-pbspro-dech                  0
 - ce.bfg.uni-freiburg.de:2119/jobmanager-pbs-dech                 0
[...]
```

MAO: DI//UM 25 de Outubro de 2011

Job Description Language

More possibilities in the JDL file:

- specify environment variables:

```
Environment= { "VAR_A=23", "VAR_B=42" };
```

- define number of retries:

```
RetryCount = 0;            (after the job has started on the worker node)
ShallowRetryCount = 3;     (otherwise)
```

- define the "goodness" of a computing element

```
Rank = -other.GlueCEStateEstimatedResponseTime;      (default)
Rank = other.GlueCEStateFreeCPUs;
```

- specify requirements on the computing element:
  - you can access attributes of the Information System (see next slides)
  - use `glite-wms-job-list-match` to test your requirements!
  - boolean algebra and regular expression are possible:

```
Requirements = ( var1>=42
                 && (var2=="foo" || var2=="bar")
                 && !RegExp("bad",var3) );
```


Job Description Language

Tags and remotely installed software:

- special tags can be assigned to computing elements

```
Member("VO-dech-gks08",
       other.GlueHostApplicationSoftwareRunTimeEnvironment);
```

- each VO can centrally install software on sites (done by admins) and adds a specific tag to these sites
- you can list the available tags for a site:

```
gks> lcg-ManageVOTag -host ce-1-fzk.gridka.de -vo dech --list

VO-dech-fortune
VO-dech-gks08
```

- the label `VO-dech-fortune` indicates the existence of a script called `fortune.sh` under `$VO_DECH_SW_DIR` on the worker node
- use the label `VO-dech-gks08` to be sure to work on a working site

Job Description Language

Complete example:

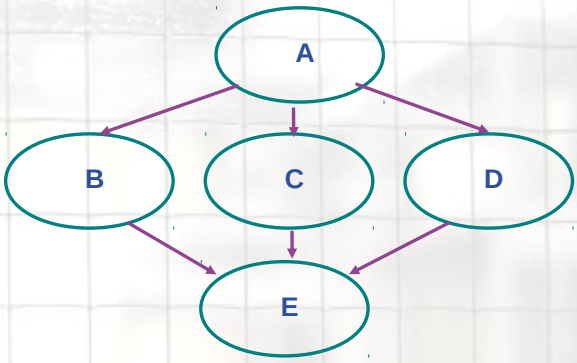
```
Executable    = "job.sh";
Arguments     = "Hello World!";

StdOutput     = "std.out";
StdError      = "std.err";
InputSandbox  = {"job.sh"};
OutputSandbox = {"std.out", "std.err"};

Requirements  = !RegExp("rwth", other.GlueCEUniqueID)
                 && other.GlueHostNetworkAdapterOutboundIP
                 && other.GlueCEPolicyMaxCPUTime >= 60
                 && other.GlueCEPolicyMaxWallClockTime >= 240
                 && other.GlueHostMainMemoryRAMSize >= 512
                 && Member("VO-dech-gks08",
                          other.GlueHostApplicationSoftwareRunTimeEnvironment);

VirtualOrganisation = "dech"
```


Advanced Job Types

Job Collection: a collection is a group of jobs with no dependencies:

- basically a collection of JDL's
- Can have common sandbox

Direct Acyclic Graph (DAG): a set of jobs where the input, output, or execution have interdependencies.



Parametric Job: is a job having one or more attributes in the JDL that vary their values according to parameters.

Why use advanced job types:

- One shot submission of a (possibly very large, up to thousands) group of jobs.
- Submission time reduction.
- Single call to WMProxy server.
- Single Authentication and Authorization process.
- Sharing of files between jobs.
- Availability of both a single Job ID and an ID for each single job in the group.

Introduction

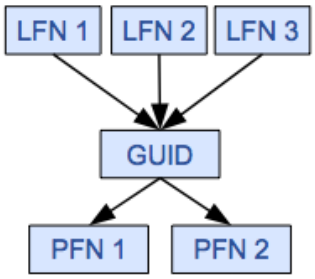
- store large files on storage elements (SE)
- book-keeping is done by the LCG File Catalog (LFC)
- several file name types for identification

Global unique identifier (GUID)
`guid:27d8a028-f8bc-41fd-a51c-4e6be3e8300a`

Logical file name (LFN)
given by the user, different LFN for one file possible
`lfn:/grid/dech/gks/john/romeojuliet.txt`

Physical file name (PFN)
the physical file name is different on each site

```
sfn://scaise-2.scai.fraunhofer.de/storage/dech/generated/200...
srm://grid-srm.rzg.mpg.de/pnfs/rzg.mpg.de/data/dech/generate...
```



- ***a modification is however impossible (may be stored on tapes)***

LCG File Catalog

- use `lcg-infosites` to find the LFC server for your VO
- environment variable LFC_HOST has to be set
  (may already be set on UIs, but not on WNs)
- commands are similar to the corresponding Linux commands

```
gks> lcg-infosites --vo dech lfc
rb.scai.fraunhofer.de

gks> export LFC_HOST=rb.scai.fraunhofer.de

gks> lfc-ls -l /grid/dech/gks
drwxrwxr-x   0 219      106                0 Jan 08 01:47 alice
drwxrwxr-x   0 219      106                0 Jun 15 12:17 bob

gks> lfc-mkdir /grid/dech/gks/john

gks> lfc-ls -l /grid/dech/gks
drwxrwxr-x   0 219      106                0 Jan 08 01:47 alice
drwxrwxr-x   0 219      106                0 Jun 15 12:17 bob
drwxrwxr-x   0 219      106                0 Jun 28 21:49 john
```

`-l` gives more information

Storage Elements

- files are physically saved on Storage Elements (SE) on the grid
- get a list of available storage elements

```
gks> lcg-infosites --vo dech se
Avail Space(Kb) Used Space(Kb)  Type    SEs
----------------------------------------------------------
203997005       85913288        n.a     grid-se3.desy.de
69924040        2008660         n.a     srm-dcache.desy.de
9342904161      15009863800     n.a     globe-door.ifh.de
10000000000     190000000000    n.a     gridka-dCache.fzk.de
66789956        4220592         n.a     lcg-se-std.gsi.de
9000000         n.a             n.a     grid-se.rzg.mpg.de
n.a             n.a             n.a     grid-srm.rzg.mpg.de
1               1               n.a     grid-srm.physik.rwth-aachen.de
1               1               n.a     grid-srm.physik.rwth-aachen.de
1927934760      2752777304      n.a     scaise-2.scai.fraunhofer.de
```


Data Management

File handling on the grid

- copy a file to the storage element `scaise-2.scai.fraunhofer.de`
  (use option -v to get more information)

```
gks> lcg-cr -d scaise-2.scai.fraunhofer.de \
        -l lfn:/grid/dech/gks/john/romeojuliet.txt file://$PWD/romeojuliet.txt
guid:bdc58dca-3bde-4617-9c37-acf4ab04ca7c
```

- check where the file has been stored

```
gks> lcg-lr lfn:/grid/dech/gks/john/romeojuliet.txt
sfn://scaise-2.scai.fraunhofer.de/storage/dech/generated/2008-06-28/file22370...

gks> lcg-lr guid:bdc58dca-3bde-4617-9c37-acf4ab04ca7c
sfn://scaise-2.scai.fraunhofer.de/storage/dech/generated/2008-06-28/file22370...
```

Data Management

– download the file

```
gks> lcg-cp lfn:/grid/dech/gks/john/romeojuliet.txt file:///tmp/romeojuliet.txt

gks> ls -lh /tmp/romeojuliet.txt
-rw-r--r--  1 john gks 139K Jun 28 23:42 /tmp/romeojuliet.txt
```

– replicate the file to the storage element `gridka-dCache.fzk.de`

```
gks> lcg-rep -d gridka-dCache.fzk.de lfn:/grid/dech/gks/john/romeojuliet.txt

gks> lcg-lr lfn:/grid/dech/gks/john/romeojuliet.txt
sfn://scaise-2.scai.fraunhofer.de/storage/dech/generated/2008-06-28/file22370...
srm://gridka-dCache.fzk.de/pnfs/gridka.de/dech/generated/2008-06-28/file4306b...
```


Data Management

– remove the file from a specific storage element

```
gks> lcg-del -s gridka-dCache.fzk.de lfn:/grid/dech/gks/john/romeojuliet.txt

gks> lcg-lr lfn:/grid/dech/gks/john/romeojuliet.txt
sfn://scaise-2.scai.fraunhofer.de/storage/dech/generated/2008-06-28/file22370...
```

– remove the file from all storage elements and the file catalog

```
gks> lcg-del -a lfn:/grid/dech/gks/john/romeojuliet.txt

gks> lcg-lr lfn:/grid/dech/gks/john/romeojuliet.txt
rb.scai.fraunhofer.de: /grid/dech/gks/john/romeojuliet.txt: No such file or
directory
lcg_lr: No such file or directory
```

Data Management

– overview:

LFC commands	file commands
lfc-ls	lcg-cr
lfc-mkdir	lcg-lr
lfc-rm	lcg-rep
lfc-ln	lcg-cp
lfc-rename	lcg-del

– it is also possible to assign comments as meta data to files

```
gks> lfc-setcomment /grid/dech/gks/john/romeojuliet.txt "William Shakespeare"

gks> lfc-ls --comment /grid/dech/gks/john/romeojuliet.txt
/grid/dech/gks/john/romeojuliet.txt William Shakespeare

gks> lfc-delcomment /grid/dech/gks/john/romeojuliet.txt
```


Data Management

– in the JDL file:

```
...

DataCatalog        = "http://rb.scai.fraunhofer.de:8085";
InputData          = {"lfn:/grid/dech/gks/john/romeojuliet.txt"};
DataAccessProtocol = {"rfio","gsiftp","gsidcap"};

...
```

– in your job file:

```
...

lcg-cp -v -V cms lfn:/grid/dech/gks/john/romeojuliet.txt \
        file://`pwd`/romeojuliet.tgz

...
```