




EGEE Enabling Grids for E-science

Grid: architecture and components (on the example of the gLite middleware)

Slides contributed by Dr. Ariel Garcia
Forschungszentrum Karlsruhe

www.eu-egee.org






EGEE Enabling Grids for E-science

Outline

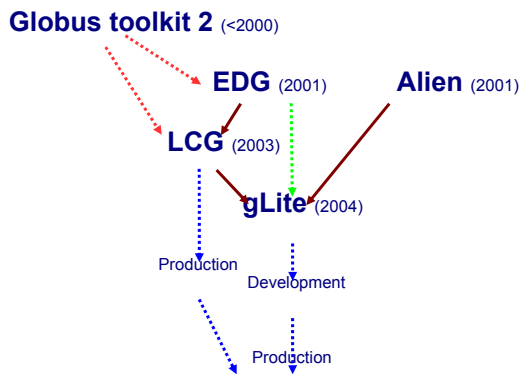
- Some history
- Grid and the middleware
- gLite components, functionality and architecture
 - security
 - information
 - job management
 - data management
- Conclusions

Grid: Architecture and Components , 11.2.2006



EGEE Enabling Grids for E-science


Some history



```


graph TD
    GT2["Globus toolkit 2 (<2000)"]
    EDG["EDG (2001)"]
    LCG["LCG (2003)"]
    Alien["Alien (2001)"]
    gLite["gLite (2004)"]
    
    GT2 -.-> EDG
    GT2 -.-> LCG
    EDG -.-> Alien
    EDG -.-> gLite
    LCG -.-> gLite
    
    gLite --> Dev["Development"]
    gLite --> Prod["Production"]
  
```

Grid: Architecture and Components , 11.2.2006



EGEE Enabling Grids for E-science


gLite – small is beautiful !

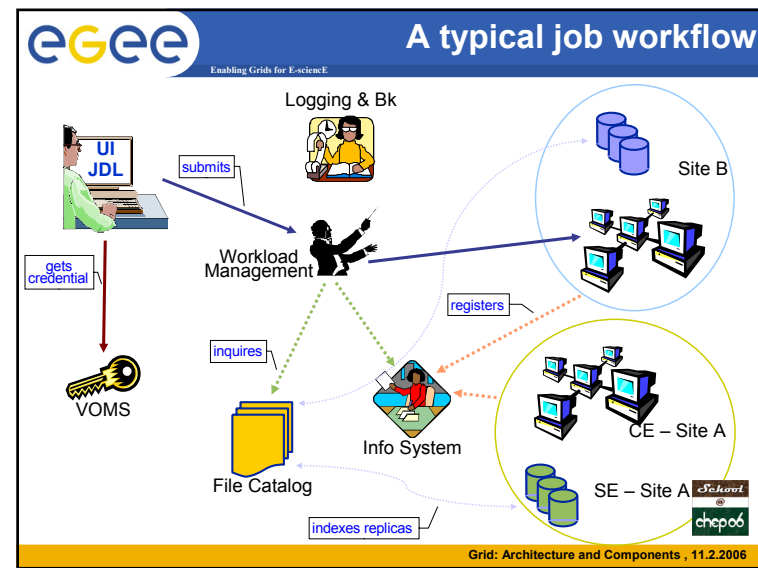
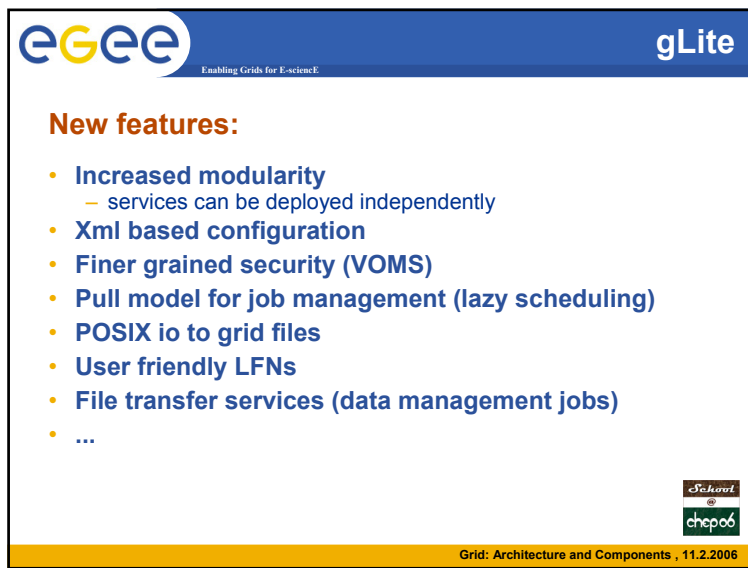
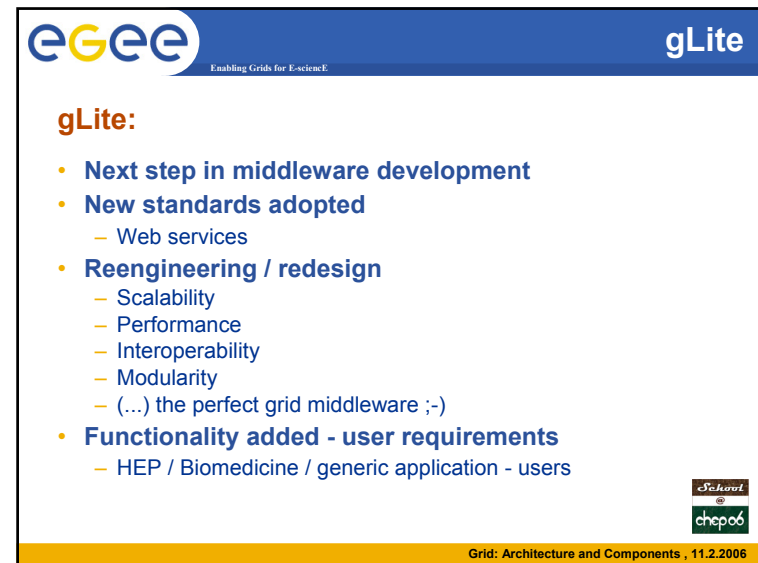
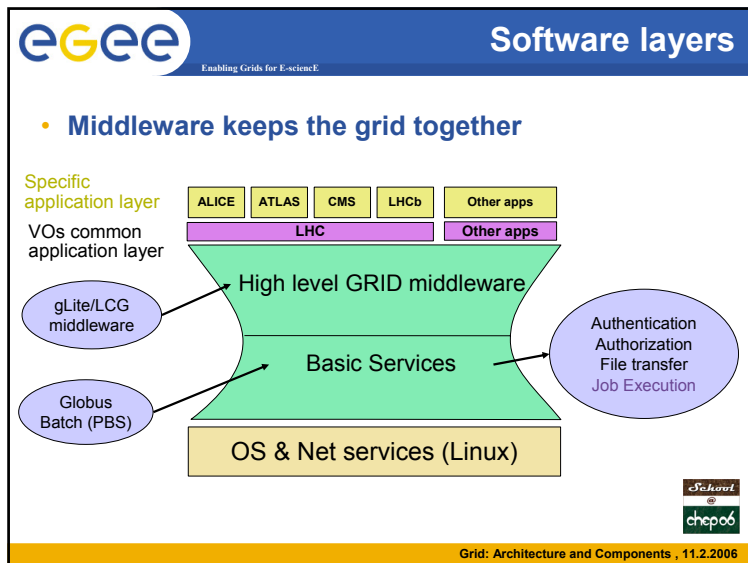


- **Lightweight (existing) services**
 - Easily and quickly deployable
 - Use existing services where possible as basis for re-engineering
- **Interoperability**
 - Allow for multiple implementations
- **Resilience and Fault Tolerance**
- **Co-existence with deployed infrastructure**
 - Reduce requirements on site components
 - Co-existence (and **convergence**) with LCG-2 and Grid3 are essential for the EGEE Grid service
- **Service oriented approach**
 - Follow WSRF standardization
 - No mature WSRF implementations exist to date so start with plain WS (WS-I)
 - Provide framework to others so higher-level services can be developed quickly

Architecture: <https://edms.cern.ch/document/476451>

Grid: Architecture and Components , 11.2.2006






EGEE Enabling Grids for E-science

Components

- **@site**
 - Computing Element (CE)
 - gateway to local computing resources (cluster of worker nodes)
 - Worker Nodes (WN)
 - Storage Element (SE)
 - gateway to local storage (disk, tape)
 - a gridftp server, an SRM interface, IO server
 - User Interfaces (UI)
 - user's access point to the grid
 - client programs using some/all grid services

CE, SE: layer of abstraction, local peculiarities irrelevant




Grid: Architecture and Components , 11.2.2006

EGEE Enabling Grids for E-science

Components

- **Grid- or VO-wide**
 - Security
 - Virtual Organization Server (VOMS)
 - MyProxy server (Proxy)
 - Information System (IS)
 - Job handling
 - Workload Management System (WMS)
 - Logging & Bookkeeping (LB)
 - Data management
 - File catalog (e.g. FiReMan)
 - File Transfer Service (FTS)
 - File Placement Service (FPS)




Grid: Architecture and Components , 11.2.2006

EGEE Enabling Grids for E-science


VOMS

Virtual Organization Membership Service

- **Multiple VOs**
- **Multiple roles in VO**
 - compatible X509 extensions
 - signed by VOMS server
- **Web admin interface**
- **Supports MyProxy**
- **Resource providers grant access to VOs or roles**
- **Sites map VO members/roles to local auth mechanism (unix user accounts)**
 - allows for local policy



Layer of abstraction: individual members irrelevant



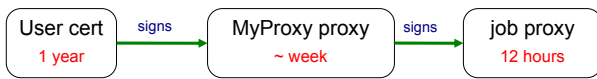

Grid: Architecture and Components , 11.2.2006

EGEE Enabling Grids for E-science

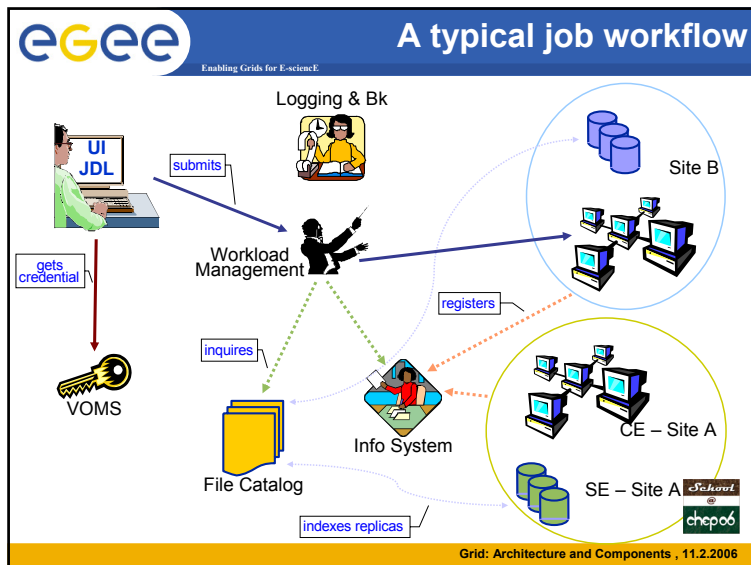
MyProxy

- **MyProxy**
 - allows longer lived jobs / increases security
 - WMS renews proxy
 - users should not produce long lived proxies :-)
 - allows for secure user mobility
 - user does not need to copy globus-keys around

Stores medium-lived proxy (days ~ weeks)

Grid: Architecture and Components , 11.2.2006

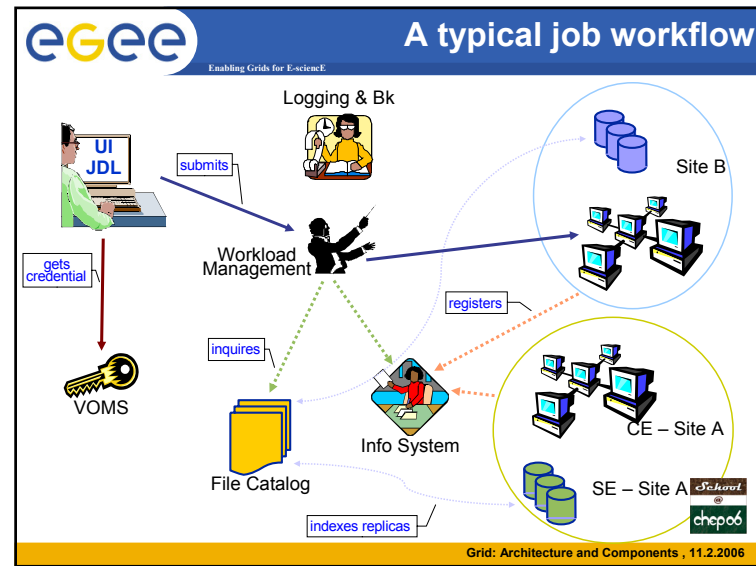
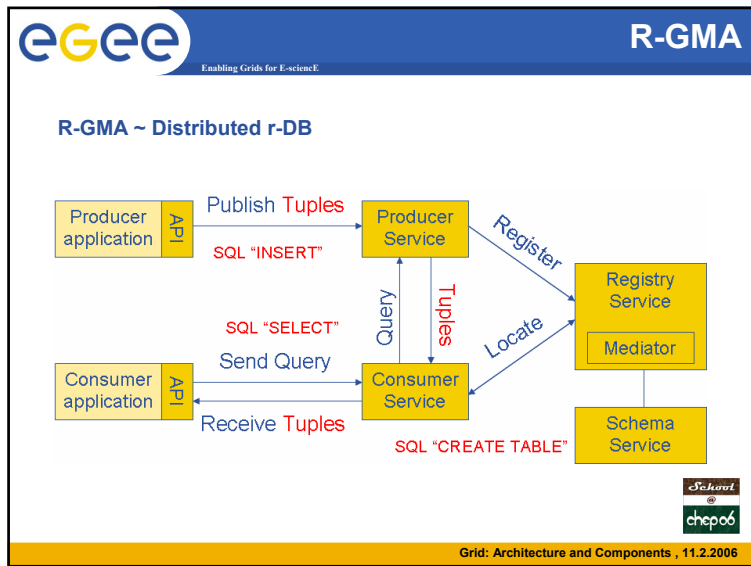


EGEE Enabling Grids for E-science

Information system

- **Based on R-GMA**
 - relational (database-like) implementation of the GGF Grid Monitoring Architecture
 - distributed
- **Aggregates service information from multiple grid sites**
 - hosts, resources (CPU, storage)
 - accepted VOs
 - based on Glue schema
- **Used by WMS (= RB's) to collect information on sites**
 - defines WMS's view of the Grid!
- **Generic Service Discovery API**
 - used by replica management tools to locate SEs, Catalogs
- **R-GMA system also used for monitoring :-)**

Grid: Architecture and Components, 11.2.2006




EGEE Enabling Grids for E-science

Job Submission

- **WMS finds best location for job**
 - considering job requirements and available resources (CPUs, files)
 - Push model: WMS pushes job to CE
 - Pull model: CE asks the WMS for jobs
 - gets resource information from IS and File Catalogs
- **JSS (Condor) provides reliable submission system**
- **LB keeps track of job's status**

• **WMS is primary job execution interface for users**
 • **each server allows only certain VOs / groups**

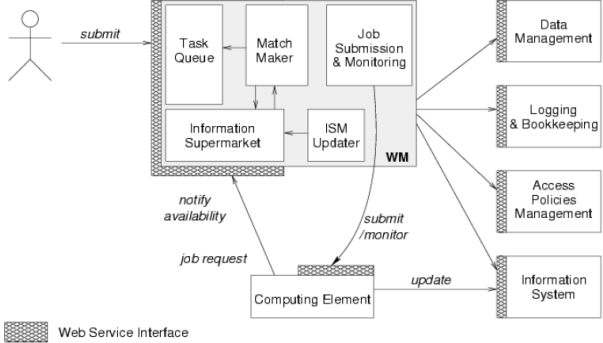
Layer of abstraction: sites irrelevant



Grid: Architecture and Components, 11.2.2006

EGEE Enabling Grids for E-science

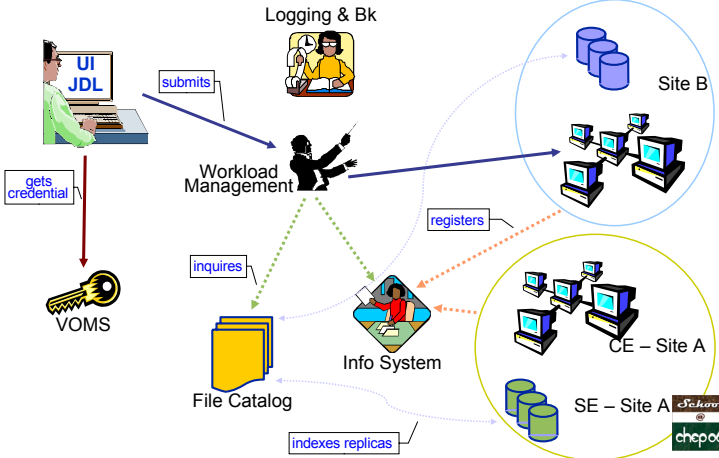
Job Management Services



Grid: Architecture and Components, 11.2.2006

EGEE Enabling Grids for E-science

A typical job workflow




Grid: Architecture and Components, 11.2.2006

EGEE Enabling Grids for E-science

Data Management Services

- **Storage Element**
 - Storage Resource Manager not provided by gLite
 - POSIX-I/O gLite-I/O
 - Access protocols gsiftp, https, rfiio, ...
- **Catalogs**
 - File catalog
 - Replica catalog
 - File authorization service
 - Metadata catalog

gLite FiReMan catalog (MySQL and Oracle)
gLite standalone metadata catalog
- **File Transfer**
 - File Transfer Service
 - File Placement Service



Grid: Architecture and Components, 11.2.2006

EGEE Enabling Grids for E-science

Data management

- Catalog remembers locations of files**
 - only deals with their locations (not data, not transfers!)
 - data transfer handled separately: PFNs point to actual storage location and access protocol
- Files can be replicated on multiple SEs**
- Each file registered has a unique ID**
 - same file gets different IDs when registered multiple times
- LFNs are names that make sense to you**

Layer of abstraction: file location irrelevant

School
chep06

Grid: Architecture and Components, 11.2.2006

EGEE Enabling Grids for E-science

Example: FiReMan Catalog

The diagram illustrates the FiReMan Catalog architecture. At the top is a red oval labeled "Metadata". Below it are two blue boxes labeled "SymLink" and a green box labeled "Logical File Name". The "Logical File Name" is connected to a yellow box labeled "Global Unique Identifier". The "Global Unique Identifier" is connected to three yellow boxes labeled "Storage URL".

FileCatalog (green text) is associated with the Logical File Name and Global Unique Identifier. **ReplicaCatalog** (red text) is associated with the Global Unique Identifier and Storage URLs.

- FileCatalog** properties:
 - Unique
 - User-defined
 - Mutable
 Example: `/grid/me/test.txt`
- Global Unique Identifier** properties:
 - Unique
 - System-defined
 - Immutable GUID
 Example: `000-000-001-002`
- Storage URL** property:
 - Allows file retrieval
 Example: `srm://host.net:8443/srm?SFN=srm/`
`my.site/myvo/grid/me/test.txt`

School
chep06

Grid: Architecture and Components, 11.2.2006

EGEE Enabling Grids for E-science

DM Interactions

The diagram shows the interactions between various components in the Data Management system. A central figure represents a user. Arrows indicate the following interactions:

- VOMS** (Virtual Organization Monitor) provides "Get credential" to the user and "Store credential" to **MyProxy**.
- MyProxy** provides "Proxy renewal" to the user.
- The user interacts with **gLite I/O**, **SRM**, and **gridFTP** for "File I/O".
- The user interacts with **Fireman** for "File namespace and Metadata mgmt" and "File replication".
- Fireman** interacts with **Storage Element** (containing "Storage") via "WSDL" and "API".
- Fireman** interacts with **StorageIndex** (containing "Database") for "Replica Location".
- StorageElement** interacts with **StorageIndex** for "Replica Location".
- StorageElement** interacts with **FPS** (File Transfer and Placement Service) and **Transfer Agent** for "File replication".
- FPS** and **Transfer Agent** interact with **WMS** (Workflow Management System).
- WMS** interacts with **StorageIndex**.

School
chep06

Grid: Architecture and Components, 11.2.2006

EGEE Enabling Grids for E-science

gLite-I/O

Provides POSIX-like access to grid files

- both CLI & API
- GUID or LFN can be used, i.e. `open("/grid/myFile");`


The diagram illustrates the gLite-I/O architecture. A **Client Application** (part of a "Worker Node UI") interacts with a **Grid I/O Server** through a "secure Grid I/O" interface. The **Grid I/O Server** interacts with a **File Authorization Service** for "Access Authorization" and with **Catalogs** for "Mapping resolution (LFN/GUID to SURJ)". The **Grid I/O Server** also interacts with a **Storage Element** for "Local user Authentication" and "SRM resolution (SURJ to URL)". The **Storage Element** contains a "Native I/O Interface" (with sub-interfaces: `dcap`, `rto`, `chrp`, `hio`, `nfs`), an "SRM Interface", and a "File Transfer Interface" (with sub-interfaces: `GridFTP`, `...`). The **Storage Element** is connected to a "Storage Back-End" (containing `diachs`, `CPM`, `CASTOR`, `NeST`, `SRB`, `disk`, `...`).

School
chep06

Grid: Architecture and Components, 11.2.2006

eGee Enabling Grids for E-science **File Transfer Service**


- **Handles data management jobs**
 - “RB” for data jobs
- **Responsible for reliable file transfers between grid sites**
 - transfers (sets of) files between 2 SE's
 - endpoints with same protocol (gsiftp, ...)
- **Can be shared among VOs**



Grid: Architecture and Components , 11.2.2006

eGee Enabling Grids for E-science **File Transfer Service**


- **Transfer jobs**
 - identifier
 - state
 - files (source/destination PFN pairs)
 - support MyProxy
 - glite-transfer-submit
 - glite-transfer-status
- **Channels**
 - point to point (cern.ch – fzk.de) queues
 - state
 - bandwidth
 - concurrent transfers
 - can be managed
 - production channels
 - default channel (free internet)



Grid: Architecture and Components , 11.2.2006

eGee Enabling Grids for E-science **File Placement Service**

- **Understands logical source files**
 - copy lfn:///grid/myvo/mytest.txt
- **Understands logical destination**
 - transfer to cern.ch
- **Updates the File Catalogs**
 - registers new replica SURL in Fireman
- **Builds on FTS**



Grid: Architecture and Components , 11.2.2006

eGee Enabling Grids for E-science **Practical Demonstration ?**

- voms-proxy-init --voms gilda
- glite-job-list-match hello.jdl
- glite-job-submit -o myjobid hello.jdl
- glite-job-status -i myjobid
- glite-job-output -i myjobid --dir .
- glite-put myfile.txt lfn:///myfile.txt

glite-catalog-mkdir


Try it out for yourself on
<https://glite-tutor.ct.infn.it>



Grid: Architecture and Components , 11.2.2006

EGEE Enabling Grids for E-science **voms-proxy-init**


```
[ruediger@grid019:~]$ voms-proxy-init --voms gilda
Your identity: /C=IT/O=GILDA/OU=Personal
Certificate/L=FZK/CN=Ruediger
Berlich/Email=ruediger.berlich@iwr.fzk.de
Enter GRID pass phrase:
Creating temporary proxy ..... Done
Contacting voms.ct.infn.it:15001
[/C=IT/O=GILDA/OU=Host/L=INFN
Catania/CN=voms.ct.infn.it/Email=emidio.giorgio@ct.infn.it]
"gilda" Done
Creating proxy ..... Done
Your proxy is valid until Fri Jan 20 16:19:50 2006
```



Grid: Architecture and Components , 11.2.2006

EGEE Enabling Grids for E-science **hello.jdl**

```
[ruediger@grid019:~]$ cat hello.jdl
Type = "Job";
JobType = "Normal";
Executable = "/bin/echo";
StdOutput = "hello.out";
StdError = "hello.err";
OutputSandbox =
{"hello.err", "hello.out"};
Arguments = "Hello World";
RetryCount = 7;
```



Grid: Architecture and Components , 11.2.2006


EGEE Enabling Grids for E-science **glite-job-list-match hello.jdl**

```
[ruediger@glite-tutor:~]$ glite-job-list-match hello.jdl

Selected Virtual Organisation name (from proxy certificate extension):
gilda
Connecting to host glite-rb.ct.infn.it, port 7772

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

*CEId*
dgt01.ui.savba.sk:2119/jobmanager-lcgpbs-infinite
dgt01.ui.savba.sk:2119/jobmanager-lcgpbs-long
dgt01.ui.savba.sk:2119/jobmanager-lcgpbs-short
egee008.cnaf.infn.it:2119/blah-pbs-infinite
egee008.cnaf.infn.it:2119/blah-pbs-long
egee008.cnaf.infn.it:2119/blah-pbs-short
[...]
```



Grid: Architecture and Components , 11.2.2006

EGEE Enabling Grids for E-science **glite-job-submit hello.jdl**


```
[ruediger@glite-tutor:~]$ glite-job-submit hello.jdl

Selected Virtual Organisation name (from proxy certificate extension): gilda
Connecting to host glite-rb.ct.infn.it, port 7772
Logging to host glite-rb.ct.infn.it, port 9002

*****
JOB SUBMIT OUTCOME
The job has been successfully submitted to the Network Server.
Use glite-job-status command to check job current status. Your job
identifier is:

- https://glite-rb.ct.infn.it:9000/dXyAlwgZruIywRsBMfjyFw

*****
```




Grid: Architecture and Components , 11.2.2006

EGEE Enabling Grids for E-science **glite-job-status**

```
[ruediger@glite-tutor:~]$ glite-job-status https://glite-
rb.ct.infn.it:9000/dXyAlwgZruIywRsBMfjyFw
```

BOOKKEEPING INFORMATION:

Status info for the Job : https://glite-
rb.ct.infn.it:9000/dXyAlwgZruIywRsBMfjyFw
Current Status: Scheduled
Status Reason: Job successfully submitted to Globus
Destination: dgt01.ui.savba.sk:2119/jobmanager-lcgpbs-short
Submitted: Fri Jan 20 04:28:27 2006 CET




Grid: Architecture and Components , 11.2.2006

EGEE Enabling Grids for E-science **glite-job-status**

```
[ruediger@glite-tutor:~]$ glite-job-status https://glite-
rb.ct.infn.it:9000/dXyAlwgZruIywRsBMfjyFw
```

BOOKKEEPING INFORMATION:

Status info for the Job : https://glite-
rb.ct.infn.it:9000/dXyAlwgZruIywRsBMfjyFw
Current Status: Running
Status Reason: Job successfully submitted to Globus
Destination: dgt01.ui.savba.sk:2119/jobmanager-lcgpbs-short
Submitted: Fri Jan 20 04:28:27 2006 CET




Grid: Architecture and Components , 11.2.2006

EGEE Enabling Grids for E-science **glite-job-status**

```
[ruediger@glite-tutor:~]$ glite-job-status https://glite-
rb.ct.infn.it:9000/dXyAlwgZruIywRsBMfjyFw
```

BOOKKEEPING INFORMATION:

Status info for the Job : https://glite-
rb.ct.infn.it:9000/dXyAlwgZruIywRsBMfjyFw
Current Status: Done (Success)
Exit code: 0
Status Reason: Job terminated successfully
Destination: dgt01.ui.savba.sk:2119/jobmanager-lcgpbs-short
Submitted: Fri Jan 20 04:28:27 2006 CET



Grid: Architecture and Components , 11.2.2006


EGEE Enabling Grids for E-science **glite-job-output**

```
[ruediger@glite-tutor:~]$ glite-job-output https://glite-
rb.ct.infn.it:9000/dXyAlwgZruIywRsBMfjyFw
```

Retrieving files from host: glite-rb.ct.infn.it (for https://glite-
rb.ct.infn.it:9000/dXyAlwgZruIywRsBMfjyFw)

JOB GET OUTPUT OUTCOME


Output sandbox files for the job:
- https://glite-rb.ct.infn.it:9000/dXyAlwgZruIywRsBMfjyFw
have been successfully retrieved and stored in the directory:
/tmp/glite/glite-ui/ruediger_dXyAlwgZruIywRsBMfjyFw



Grid: Architecture and Components , 11.2.2006

EGEE Enabling Grids for E-science **result**


```
[ruediger@glite-tutor:~]$ ls -l /tmp/glite/glite-  
ui/ruediger_dXyAlwgZruIywRsBMfjyFw  
total 4  
-rw-r--r-- 1 ruediger users    0 Jan 20 04:32 hello.err  
-rw-r--r-- 1 ruediger users   12 Jan 20 04:32 hello.out  
[ruediger@glite-tutor:~]$ cat /tmp/glite/glite-  
ui/ruediger_dXyAlwgZruIywRsBMfjyFw/hello.out  
Hello World
```



Grid: Architecture and Components , 11.2.2006

EGEE Enabling Grids for E-science **result**


running this example
took just 10 minutes ...



Grid: Architecture and Components , 11.2.2006

EGEE Enabling Grids for E-science **Conclusions**

- More standards compliant (WS)
- More security, virtualization of resources
- Some components evolving keeping compatibility
- Commands renamed, same functionality
- New / rearchitected components
- Several required features implemented
- Some requirements still pending
- New features expected
- Still moving target
- Current: **gLite 1.5**
- Expected: **merger of LCG and gLite**



Grid: Architecture and Components , 11.2.2006