



Enabling Grids for E-science

Additional tools for EGEE users

Three days were not enough for everything...

Gergely Sipos
MTA SZTAKI
Budapest

2nd Rio Grid School
3-5. July, 2007
Rio de Janeiro, Brazil

www.eu-egee.org

www.glite.org



- **Programming APIs for gLite services**
 - WMPProxy (~WMS API)
 - SEE-GRID File Management API (~LFC&LCG API)
 - GFAL API → you already know...
- **Additional services in gLite**
 - R-GMA
- **RESPECT – Initiative to collect useful tools that work:**
 - GridWay
 - GANGA
 - P-GRADE → you already know...



Enabling Grids for E-scienceE

WMPProxy

www.eu-egee.org

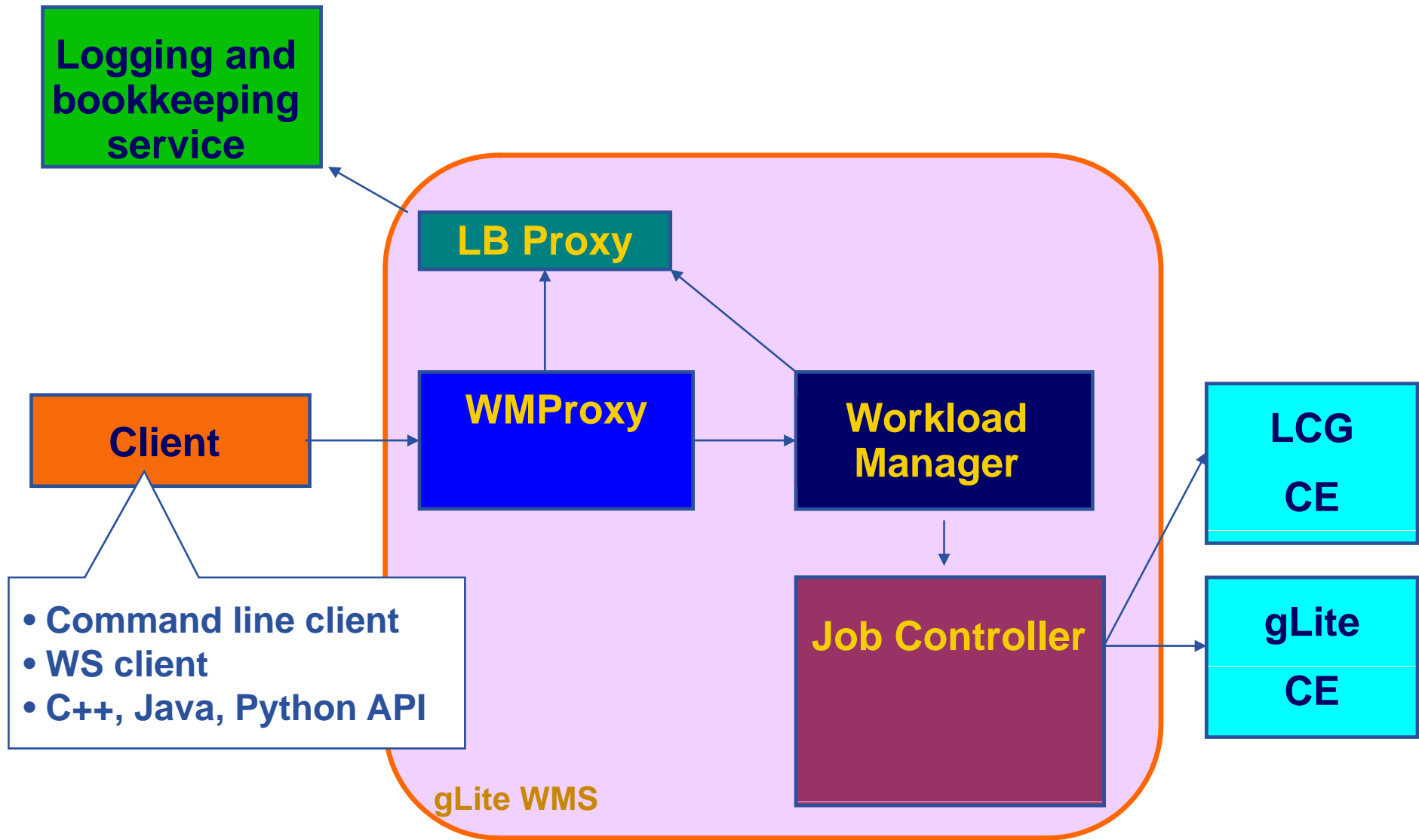
www.glite.org



- The WMPProxy is the service responsible to provide access to the WMS functionality through a Web Service Interface
- The gLite WMPProxy Server can be either accessed directly through the published WSDL, the C++ command line interface, or the API
- In these two links you can find a guide about how to interact with the main services exposed by the WMPProxy through the provided API Java and C++ API

<https://grid.ct.infn.it/twiki/bin/view/GILDA/ApiJavaWMPProxy>

<https://grid.ct.infn.it/twiki/bin/view/GILDA/WMPProxyCPPAPI>



```
[larocca@glite-tutor:~/API]$ java WMPProxyGetProxyAndSubmit
```

```
WMPProxyGetProxyAndSubmit.java
"A simple client to interact with the WMPProxy Server."
```

```
Author: Giuseppe La Rocca (giuseppe.larocca@ct.infn.it)
I.N.F.N. - Sez. of Catania - ITALY
Via S.Sofia, 64 - 95123 Catania
Phone: +39.095.378.53.74
```

Usage :

```
java WMPProxyGetProxyAndSubmit -h[elp]
java WMPProxyGetProxyAndSubmit <user_proxy> <delegation_id> <wmproxy_server> <InputSandboxFiles>
                                     <jdl_file> <CAcertsPath> [CAs paths (optional)]
```

where:

```
<user_proxy>           ... the file containing the user's credentials
<delegation_id>       ... the string used to save the user's delegation

<wmproxy_server>     ... the entry point of the WMPProxy Server to contact
                       (e.g. : https://glite-rb3.ct.infn.it:7443/glite_wms_wmproxy_server)

<InputSandboxFiles> ... The list of file(s) to transfer to the WMPProxy Server
<jdl_file>           ... the jdl file to submit to the grid

<CAcertsPath>        ... the path location of the directory containing all the Certificate
                       Authorities files
```

```
Contacting... https://glite-rb2.ct.infn.it:7443/glite_wms_wmproxy_server with the proxy..
/tmp/x509up_u512
```

Your job has been successfully submitted.

```
jobID = [ https://glite-rb2.ct.infn.it:9000/XAoY7FZg LJjgCp4U9grsBw ]
```

```
for (int index = 0; index < InputSandboxFiles.length; index++)
{
  String toURL = front + "2811" + rear;
  toURL = toURL + "/" + InputSandboxFiles[index];
  fromURL = "file:/// " + InputSandboxFiles[index];

  try {
    GlobusURL from = new GlobusURL(fromURL);
    GlobusURL to = new GlobusURL(toURL);

    UrlCopy uCopy = new UrlCopy();
    uCopy.setDestinationUrl(to);
    uCopy.setSourceUrl(from);
    uCopy.setUseThirdPartyCopy(true);

    uCopy.copy();
  } catch (Exception e) {System.err.println(e.getMessage());}
}
```

**Specify the Destination
and Source URL(s)**

**Copy file(s) from the UI to
the Resource Broker**

The script, thanks to the **UrlCopy** Class, performs the copy of the InputSandbox files to the reduced path of the WMS as you can see:

```
[root@glite-rb2 root]# cd /var/glite/SandboxDir/XA/
```

```
[root@glite-rb2 root]# ll https_3a_2f_2fglite-  
rb2.ct.infn.it_3a9000_2fXAoY7FZgLGjC4U9grsBw/input/
```

```
-rwxrwxr-x 1 gilda001 glite 30 Jan 11 09:05 start_hostname.sh
```

With the job finishes you can retrieve the output file(s) as follow:

```
[larocca@glite-tutor:~/API]$ java WMPProxyGetOutputAndPurge
Usage :
  java WMPProxyGetOutputAndPurge -h[elp]
  java WMPProxyGetOutputAndPurge <user_proxy> <wmproxy_server> <jobId>
                                     <dirPath> <CAcertsPath> [CAs paths (optional)]
+-----+
WMPProxy URL = [https://glite-rb2.ct.infn.it:7443/glite_wms_wmproxy_server]
proxyFile    = [/tmp/x509up_u512]
JobID        = [https://glite-rb2.ct.infn.it:9000/XAoY7FZgLJjgCp4U9grsBw]
dirPath      = [/home/larocca/API/]
CA certs     = [/etc/grid-security/certificates/]
+-----+
```

**List of file(s) retrieved from to the
Resource Broker to the user's
account**

```
Start downloading output file(s)..
file n. 1
-----
name = [gsiftp://glite-rb2.ct.infn.it:2811/var/glite/SandboxDir/XA/https_3a_2f_2fglite-
rb2.ct.infn.it_3a9000_2fXAoY7FZgLJjgCp4U9grsBw/output/hostname.err]
size = [0]

file n. 2
-----
name = [gsiftp://glite-rb2.ct.infn.it:2811/var/glite/SandboxDir/XA/https_3a_2f_2fglite-
rb2.ct.infn.it_3a9000_2fXAoY7FZgLJjgCp4U9grsBw/output/hostname.out]
size = [28]
```




API Documentation

<http://trinity.datamat.it/projects/EGEE/wiki/apidoc/3.1/htmljava/index.html>



Datamat – WMPProxy quick start

<http://trinity.datamat.it/projects/EGEE/wiki/wiki.php?n=WMPProxyClient.QuickStart>



JDL Attributes guide for WMPProxy

<https://edms.cern.ch/document/590869/1>



WMPProxy user guide

<https://edms.cern.ch/document/674643/1>



Enabling Grids for E-scienceE

SEE-GRID File management API

www.eu-egee.org

www.glite.org



- **SEE-GRID File Management Java API supports most of the data management operations offered by LFC and LCG_UTILS.**
- **These Java API are compatible with LCG 2.7.x and gLite grid middleware.**

Method Summary

java.lang.String	getCatalogType () Returns type of used grid file catalogue.
Item	getItem (java.lang.String pathName) Returns the Item of the corresponding type for a given pathname.
DirectoryItem	getRoot () Returns root directory of grid file catalogue.
SEList	getSEList () Returns list of available storage elements.
java.lang.String	getSeparator () Returns default pathname-separator character for used grid file catalogue.
java.lang.String	getVO () Returns name of Virtual Organisation.

Directory management: DirectoryItem class

Method Summary	
boolean	<u>canExecute()</u> Test for execute permission.
boolean	<u>canRead()</u> Test for read permission.
boolean	<u>canWrite()</u> Test for write permission.
boolean	<u>copyAndRegister</u> (java.lang.String sourceFilePath, java.lang.String destinationSE) Copies and registers file in grid catalogue directory.
boolean	<u>copyAndRegister</u> (java.lang.String sourceFilePath, java.lang.String destinationFileName, java.lang.String destinationSE) Copies and registers file in grid catalogue directory.
boolean	<u>createNewAlias</u> (java.lang.String newAliasPathname) Creates the Item's alias with a given pathname.
boolean	<u>exists()</u> Test if the item denoted by pathname exists.
java.lang.String[]	<u>getAliases()</u> Returns the list of Item's aliases.
java.lang.String	<u>getComment()</u> Returns associated comment.
int	<u>getFileMode()</u> Returns the filemode value describing item's type and permissions.
int	<u>getGID()</u> Returns the Group ID (GID) of the group owning the Item.
java.lang.String	<u>getGroup()</u> Returns the name of the group owning the Item.

<u>Item</u>	<u>getParent()</u> Returns parent item.
java.lang.String	<u>getParentPathName()</u>
long	<u>getSize()</u> Returns size in bytes.
int	<u>getUID()</u> Returns the User ID (UID) of the user owning the Item.
java.lang.String	<u>getUser()</u> Returns the name of the user owning the Item.
boolean	<u>mkdir</u> (java.lang.String name) Creates subdirectory with the given name.
boolean	<u>mkdir</u> (java.lang.String name, <u>LFCFileMode</u> lfcFileMode) Creates subdirectory with the given name and permissions.
protected void	<u>populateChildren()</u> Fetches the items contained by the directory.
void	<u>refresh()</u> Refreshes the cached information about the directory.
boolean	<u>renameTo</u> (java.lang.String newPathName) Renames/moves the item to a given pathname.
void	<u>setComment</u> (java.lang.String comment) Assigns a new comment to the item.

Method Summary

File management: FileItem class

boolean	<u>canExecute()</u> Test for execute permission.
boolean	<u>canRead()</u> Test for read permission.
boolean	<u>canWrite()</u> Test for write permission.
boolean	<u>createNewAlias</u> (java.lang.String newAliasPathname) Creates the Item's alias with a given pathname.
boolean	<u>delete()</u> Deletes file.
boolean	<u>deleteReplicaFromSE</u> (java.lang.String se) Deletes replica of a file from specified Storage element
boolean	<u>deleteReplicaFromSurl</u> (java.lang.String surl) Deletes replica specified by surl.
boolean	<u>download</u> (java.lang.String destinationFile) Downloads file to local filesystem.
boolean	<u>download</u> (java.lang.String surl, java.lang.String destinationFile) Downloads file to local filesystem.
boolean	<u>exists()</u> Test if the item denoted by pathname exists.
java.lang.String[]	<u>getAliases()</u> Returns the list of Item's aliases.
java.lang.String	<u>getComment()</u> Returns comment associated with file.



int	<code>getFileMode ()</code> Returns the filemode value describing item's type and permissions.
int	<code>getGID ()</code> Returns the Group ID (GID) of the group owning the Item.
java.lang.String	<code>getGroup ()</code> Returns the name of the group owning the Item.
java.lang.String	<code>getGUID ()</code> Returns guid of a file.
<code>Item</code>	<code>getParent ()</code> Returns parent item.
java.lang.String	<code>getParentPathName ()</code>
java.lang.String[]	<code>getReplicas ()</code> Returns list of file's replicas.
long	<code>getSize ()</code> Returns size in bytes.
int	<code>getUID ()</code> Returns the User ID (UID) of the user owning the Item.
java.lang.String	<code>getUser ()</code> Returns the name of the user owning the Item.
void	<code>refresh ()</code> Refreshes the cached information about the file.
boolean	<code>renameTo (java.lang.String newPathName)</code> Renames/moves the item to a given pathname.
boolean	<code>replicate (java.lang.String se)</code> Replicates file.

File management: FileItem class



SEE-GRID File Management Java API Documentation

<http://grid02.rcub.bg.ac.yu/LFCJavaAPI/files/docs/javadoc/version1.2/index.html>



Source code (version 1.2)

<http://grid02.rcub.bg.ac.yu/LFCJavaAPI/files/downloads/SEE-GRIDFileManagementAPIv1.2.zip>



Source code (version 1.1)

<http://grid02.rcub.bg.ac.yu/LFCJavaAPI/files/downloads/SEE-GRIDFileManagementAPIv1.1.zip>



Enabling Grids for E-scienceE

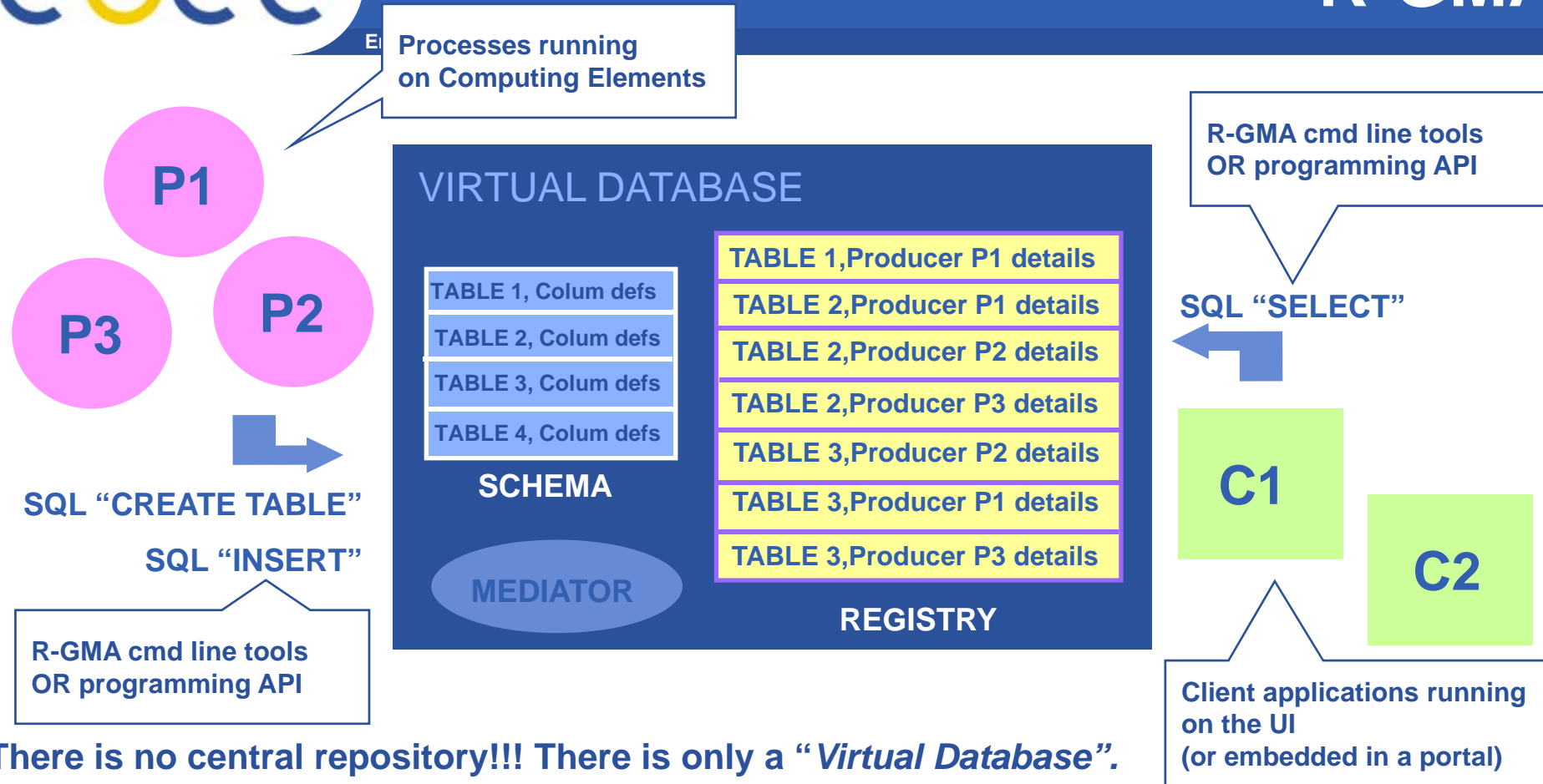
R-GMA

www.eu-egee.org

www.glite.org



- **Uniform method to access and publish both information and monitoring data.**
- **From a user's perspective, an R-GMA installation currently appears similar to a single relational database.**
- **GMA (Grid Monitoring Architecture) was developed by the Global Grid Forum (Predecessor of Open Grid Forum – www.ogf.org)**
- **R-GMA (Relational GMA) was created:**
 - To simplify use of GMA
 - To give a relational view



There is no central repository!!! There is only a "Virtual Database".

Schema is a list of table definitions: additional tables/schema can be defined by applications

Registry is a list of data producers with all its details.

Producers publish data.

Consumers read data published.

- **R-GMA overview page.**
 - <http://www.r-gma.org/>
- **R-GMA in EGEE**
 - <http://hepunix.rl.ac.uk/egee/jra1-uk/>
- **R-GMA command line tool**
 - <http://hepunix.rl.ac.uk/egee/jra1-uk/glite-r1/command-line.pdf>
- **R-GMA Browser Home Page**
 - <https://rgmasrv.ct.infn.it:8443/R-GMA/>



Enabling Grids for E-science

EGEE NA4 RESPECT initiative

*Recommended External Software Packages for
Egee CommuniTies*

www.eu-egee.org

www.glite.org

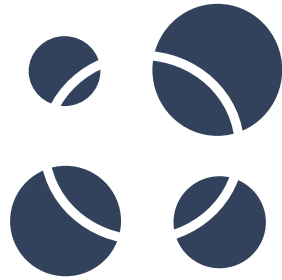


- **EGEE gLite middleware: implementations of base grid services that serve as a platform for high-level services**
- **Not aim to provide a comprehensive solution for any particular grid application**
- **Software from external providers must be used in conjunction with gLite to satisfy fully the needs of the user community**

- **The array of available grid software is vast!**
- **RESPECT (Recommended External Software Packages for EGEE Communities) program aims identifying useful, well-supported software for EGEE users**

- Having a set of external software packages that **enhances the functionality of the gLite middleware**
- Reduces the amount of application development, and **generally accelerates the adoption of grid technologies**
- **Reduce the pressure on the EGEE middleware activity** to provide solutions for services outside of the core functionality
- Integration and testing activities in **EGEE can concentrate on core gLite issues**
- Increasing the number of users via a **more attractive platform** and having more varied services

- RESPECT provides **list** of software that focuses on those packages that are genuinely useful for an EGEE application and that are generic enough to be useful to other applications
- The RESPECT program is **not a general repository** of grid software
- **Current RESPECT tools:**
 - GANGA
 - GridWay
 - P-GRADE Portal
- **Further information:**
<http://egeena4.lal.in2p3.fr/> → “Grid software” menu



The GridWay Metascheduler



GridWay

GridWay

one of the tools recognised by EGEE's RESPECT program

Alternative to WMS

Higher level command line UI

Examples of use:

Alternative broker – no need for close CE-SE

Many similar jobs

Resources outside EGEE also to be used

User-site-specific policies are required (priorities of users' jobs)

...



What is GridWay?

GridWay is a meta-scheduler that works on top of Globus-based services (e.g. GRAM, MDS & GridFTP, CE)

For the user

A Local resource management system-like (LRMS) environment for submitting, controlling & monitor jobs

A way to execute your applications on the Grid, without having to worry about resource brokering, file staging or failures

For the Grid Application Developer

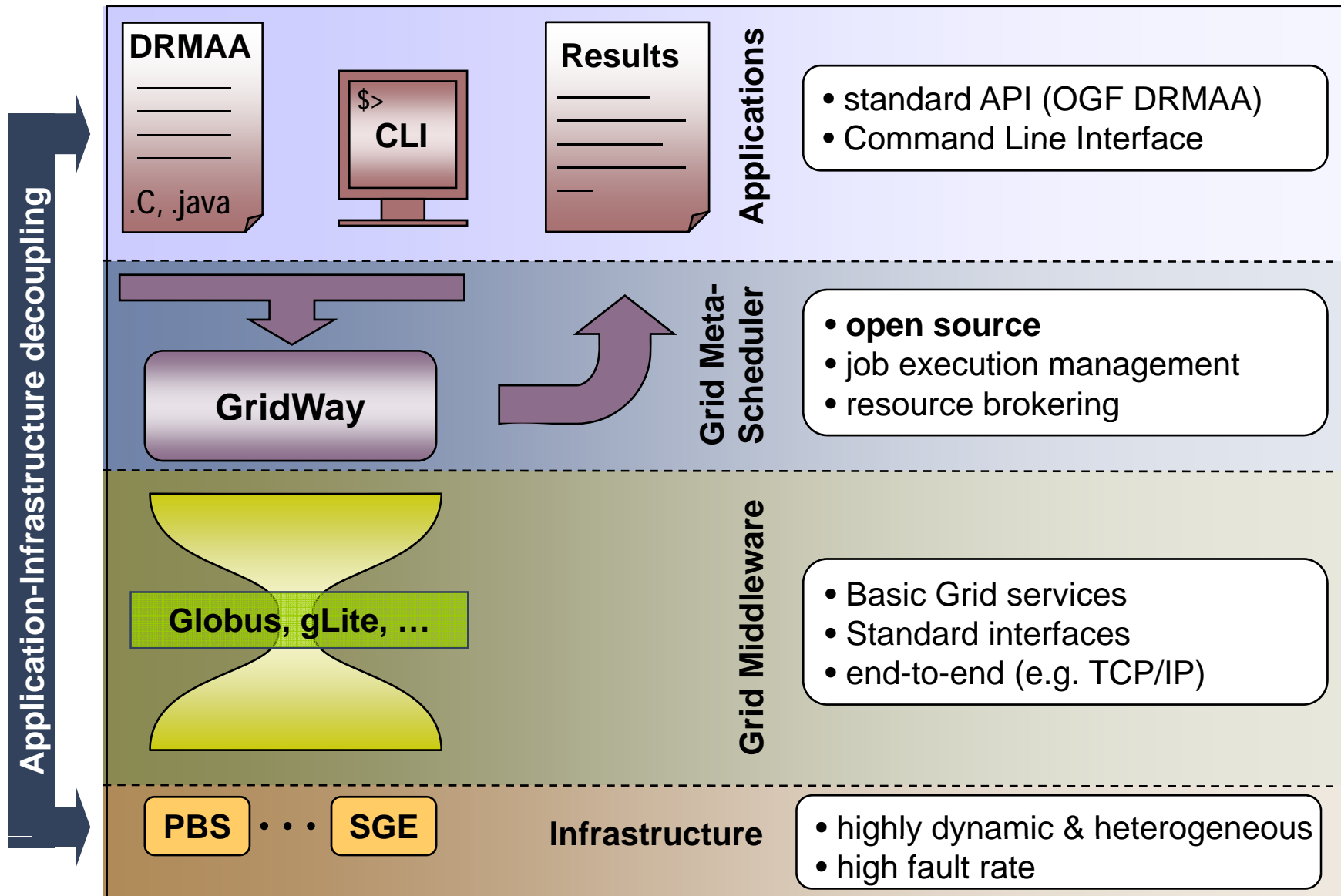
- A standard-base development framework for Grid Applications
- JAVA and C bindings of DRMAA API – Programming API to manage jobs

For the System Administrator

- A policy-driven job scheduler, implementing a wide range of access and Grid-aware policies.

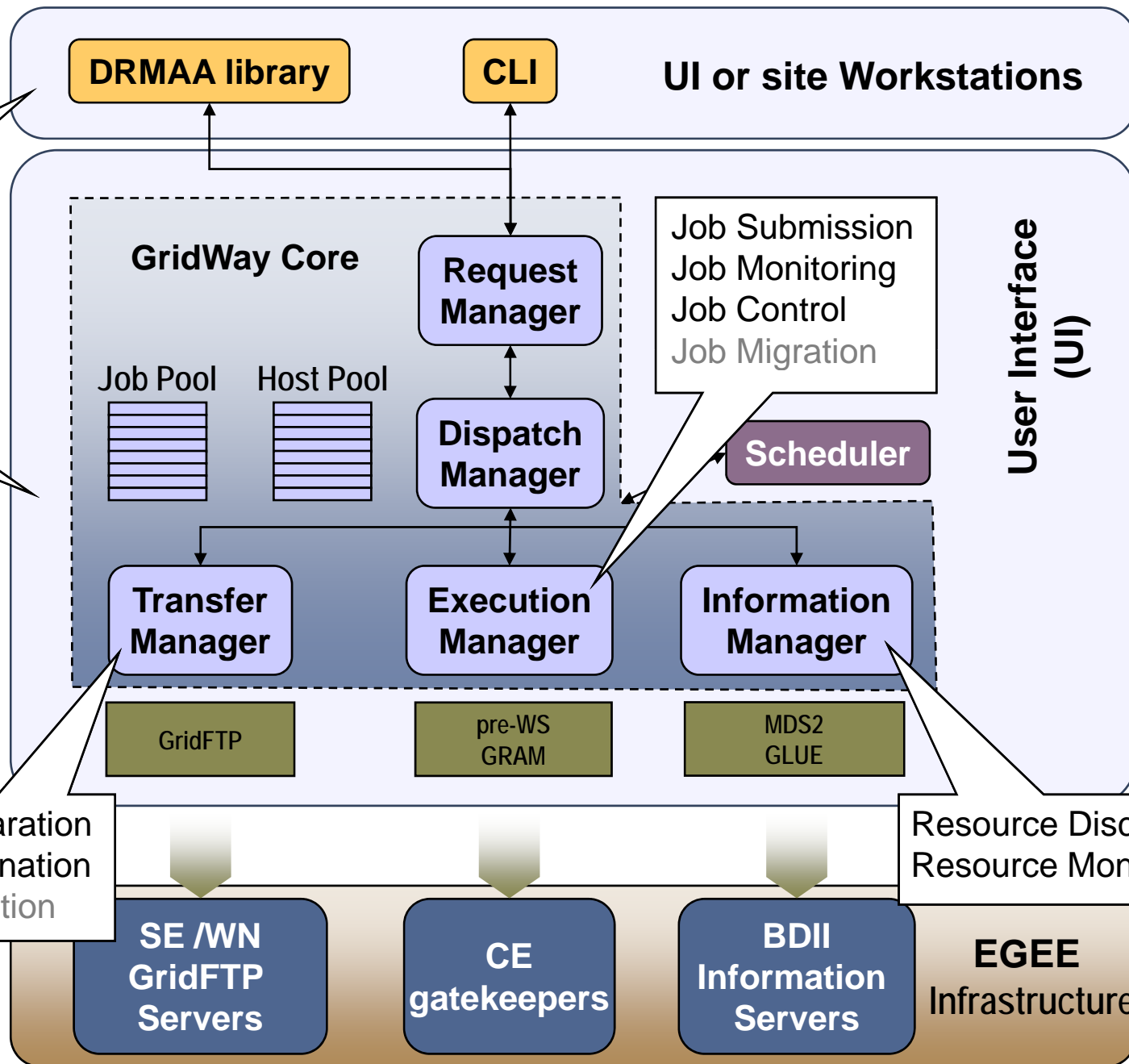


What is GridWay?



How do I use GridWay in EGEE?

A broker and a user or programming interface integrated





How do I use GridWay in EGEE?

Job Template – similar to, but not the same as JDL!

```
# Execution variables
EXECUTABLE = job
ARGUMENTS = ${TASK_ID} ${TOTAL_TASKS} 100000
ENVIRONMENT = LD_LIBRARY_PATH=/usr/local/lib

# Resource selection parameters
REQUIREMENTS = HOSTNAME= "*.dacya.ucm.es"
RANK = CPU_MHZ

# I/O files
INPUT_FILES = my_inputfile
OUTPUT_FILES = my_outputfile

# Standard streams
STDOUT_FILE = stdout_file.${TASK_ID}
STDERR_FILE = stderr_file.${TASK_ID}
```

Parameter
study jobs

Parameter
study jobs



How do I use GridWay in EGEE?

gwps: display job information and status

USER	JID	AID	TID	DM	EM	START	END	EXEC	XFER	EXIT	NAME	HOST
ruben	0	--	--	done	----	15:31:57	15:44:08	0:10:01	0:01:26	0	job1.jt	cluster.pnpi.nw.ru
rg	1	--	--	done	----	15:31:58	15:44:11	0:09:59	0:01:26	0	MPI.jt	e1.egee.fr.cgg.com
rg	2	--	--	done	----	17:07:44	17:21:09	0:11:27	0:01:28	0	maratra.jt	aquila.dacya.ucm.es
nacho	3	--	--	prol	----	17:07:47	----	0:11:19	0:01:43	--	maratra.jt	e1.egee.fr.cgg.com
rg	4	--	--	done	----	17:41:29	17:55:07	0:11:29	0:01:27	0	maratra.jt	heplnx201.pp.ac.uk
rg	5	--	--	done	----	17:41:32	17:54:05	0:10:24	0:01:28	0	test.jt	e1.egee.fr.cgg.com
jlvezq	6	--	--	pend	----	10:58:38	----	0:54:06	0:58:37	--	test.jt	gridgate.cs.tcd.ie

gwhost: display resources information and status

HID	OS	ARCH	MHZ	%CPU	MEM(F/T)	DISK(F/T)	N(U/F/T)	LRMS	HOSTNAME
0	Scientific	i686	1001	0	513/513	0/0	0/169/224	jobmanager-lcgpbs	cg02.ciemat.es
1	Scientific	i686	1000	0	1536/1536	0/0	0/2/30	jobmanager-lcgpbs	lcgce01.jin.ru
2	Scientific	i686	2800	0	2048/2048	0/0	0/1/98	jobmanager-lcgpbs	lcg6.smsu.ru
3	Scientific	i686	1266	0	2048/2048	0/0	0/0/6	jobmanager-pbs	ce1.cgg.com
4	Scientific	i686	3000	0	2048/2048	0/0	0/0/56	jobmanager-pbs	cluster.nw.ru
5	Linux2.6.16	x86	3216	73	862/2027	114643/118812	0/1/1	Fork	cygnus.ucm.es
6	Linux2.6.16	x86	2211	0	671/1003	76882/77844	0/2/2	SGE	aquila.ucm.es
7	Linux2.6.16	x86	3215	0	133/2027	109735/118812	0/1/1	Fork	draco.ucm.es
8	Linux2.6.16	x86	3200	0	513/513	0/0	0/1/2	SGE	ursa.ucm.es
9	Linux2.6.16	x86	2211	100	673/1003	76876/77844	0/2/2	PBS	hydrus.ucm.es



How do I use GridWay in EGEE?

Other Commands

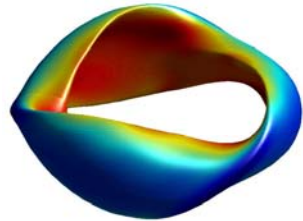
- **gwhistory**: display job execution history

HID	START	END	PROLOG	WRAPPER	EPILOG	MIGR	REASON	QUEUE	HOST
2	15:40:22	15:44:11	0:00:15	0:03:15	0:00:19	0:00:00	----	fusion	e1.egee.fr.cgg.com
1	15:36:22	15:40:09	0:00:09	0:03:21	0:00:17	0:00:00	err	fusion	e2.egee.cesga.es
0	15:32:22	15:36:11	0:00:07	0:03:23	0:00:19	0:00:00	err	fusion	ce-egee.bifi.unizar

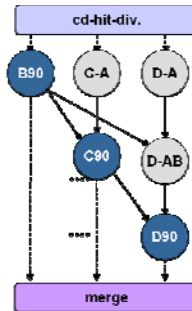
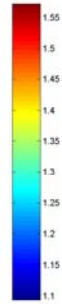
- **gckill**: signals a job (kill, stop, resume, reschedule)
- **gws submit**: submits a job, or an array job
- **gwwait**: waits for zombie state of a job (any, all, set)
- **gwuser**: displays information about users
- **gwacct**: prints accounting information



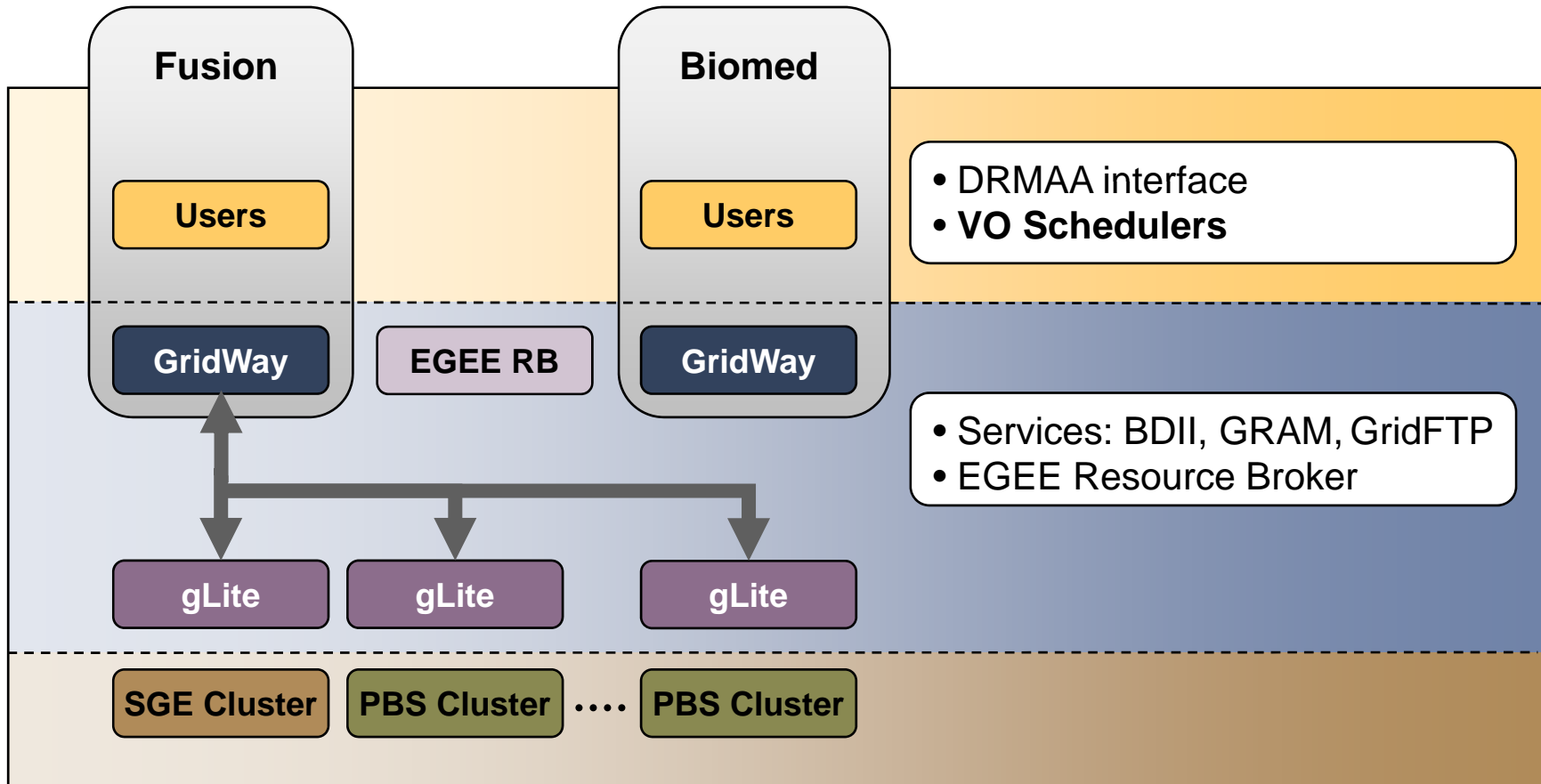
Who is using GridWay in EGEE?



Massive Ray Tracing



CD-HIT workflow





Where can I get GridWay?

Download the software

- From the Gridway webpage: www.gridway.org
- From the ETICS repository
- From the Globus CVS repository (cvs.globus.org)

Install the software

- Install it on your desktop computer OR
- Ask your institute to make a central installation OR
- Ask your VO to make a central installation

More Information

- Gridway webpage: www.gridway.org
- Application porting with GridWay

<http://www.gridway.org/successstories/applicationporting.php>

- Infrastructures using GridWay

<http://www.gridway.org/successstories/projectsinfrastructures.php>



Enabling Grids for E-scienceE

GANGA

www.eu-egee.org

www.glite.org



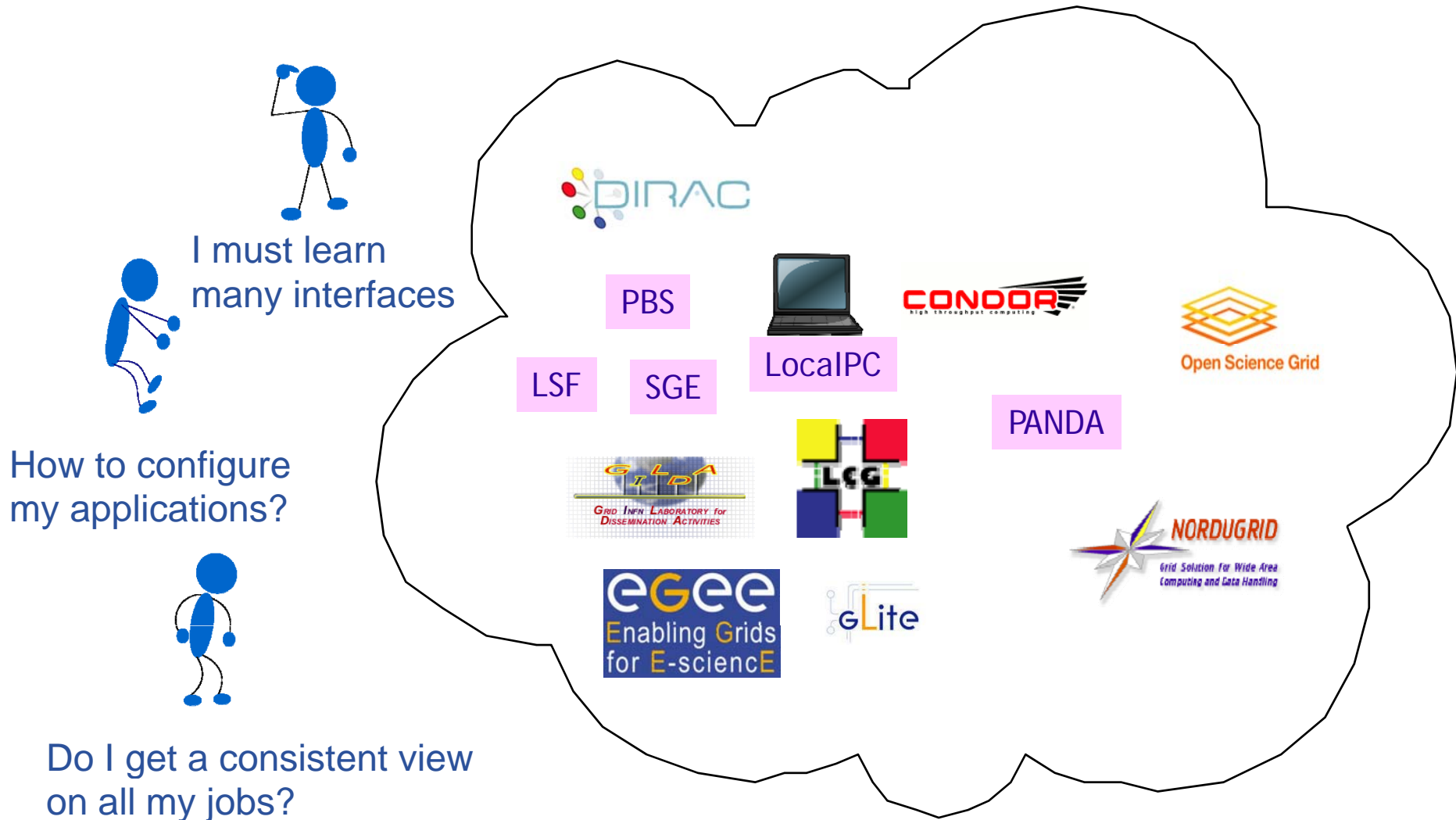
Goals:

- provide a simple and consistent way of preparing, organising and executing jobs on different computing infrastructures
- provide a clean interface which can be used:
 - interactively (CLI / python interpreter)
 - as a Python API for scripting
 - through a GUI
- Make it easy and integrated with application environment
- Allow quick transition between local PC, cluster, Grid...
- Organize work, keep history of jobs,....

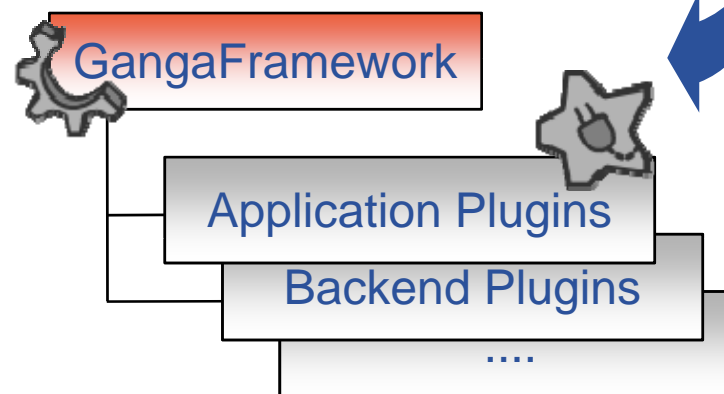
- In practice users deal with multiple computing backends



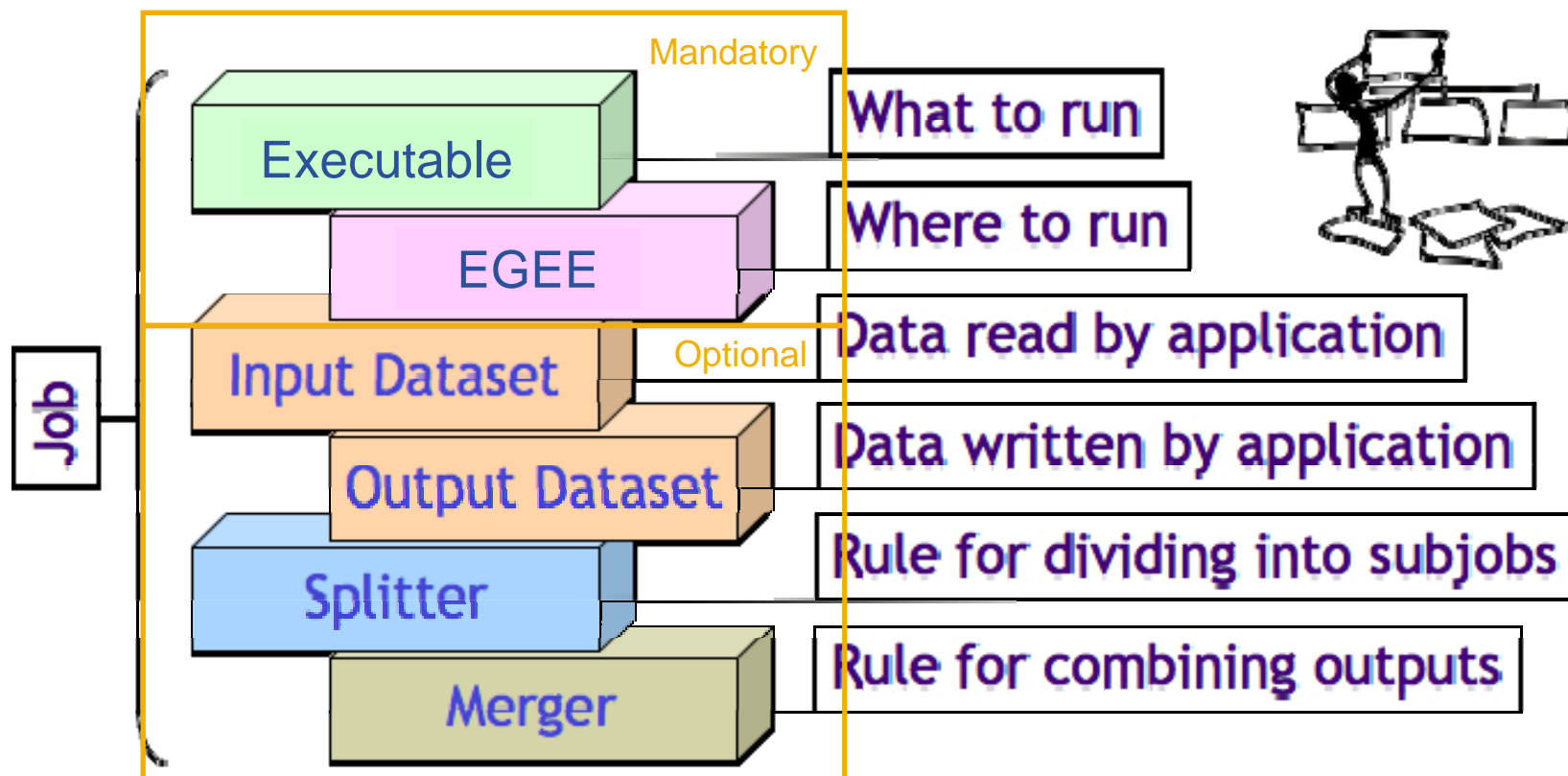
- FAQ: running applications on multiple computing backends



- **Ganga: Job Management Interface**
 - a utility which you download to your computer
 - or it is already installed in your institute in a shared area
 - *for example: /nfs/sw/ganga/install/4.3.2*
 - it is an **add-on** to installed software
 - comes with a set of plugins for some applications
 - **open** - other applications and backend may be easily added
 - *even by users*



Where the Ganga journey starts ...



```
$ ganga athena \
--inDS myInputDataset.txt\
--outputdata myOutput.root \
--split 3 \
--maxevt 100 \
--lsf \
jobOptions.py
```

Scripting mode

quick

```
j = Job()
j.application=Athena()
j.application.prepare()
j.application.option_file='jobOptions.py'
```

CLIP mode
application

```
j.inputdata=DQ2Dataset()
j.inputdata.type='DQ2_LOCAL'
j.inputdata.dataset="myInputDataset.txt"
```

inputdata

```
j.outputdata=DQ2OutputDataset()
j.outputdata.outputdata=['myOutput.root']
```

outputdata

```
j.splitter = AthenaSplitterJob(numsubjobs=3)
j.merger = AthenaOutputMerger()
```

Splitter & Merger

```
j.backend = LSF()
j.submit()
```

```
j2 = j.copy()
j2.backend=LCG( CE='ce102.cern.ch:2119/jobmanager-lcglsf-grid_2nh_atlas' )
j2.submit()
```



flexible

- **Ganga Home:**
<http://cern.ch/ganga>
- **Official Ganga User's Guide:**
<http://cern.ch/ganga/user/html/GangaIntroduction/>
- **GangaTutorial GPI Reference Manual :**
<http://ganga.web.cern.ch/ganga/release/4.3.2/reports/html/Manuals/GangaTutorialManual.html>
- **Looking for help:**
project-ganga-developers@cern.ch

- **gLite services can be accessed through programming APIs too**
 - GFAL API (yesterday), WMPProxy, SEE-GRID
- **R-GMA**
 - Relational database that mediates between your job and you
- **RESPECT program: EGEE NA4 initiative to identify useful tools that work and has user support**
 - List of software is at <http://egeena4.lal.in2p3.fr/>
 - GridWay: Broker and higher level command line client
 - Alternative to WMS; parametric jobs
 - GANGA
 - Object oriented cmd line interface for WMS; parametric jobs, splitter, merger components
 - P-GRADE:
 - Web portal, workflow and parameter study support



Enabling Grids for E-scienceE

Questions?

www.eu-egee.org

www.glite.org

