



Enabling Grids for
E-science in Europe

www.eu-egEE.org

Tutorial Grille (LCG/EGEE)

LAPP, 25/09 au 26/09 2007

SERVICE de GRILLE

Système d'information

Présentation faite à partir des exemples du site NA3/EGEE:

<http://wiki.nesc.ac.uk/read/egEE-na3-etf?Etflnduction>

et des slides de Pierre Girard (French ROC deputy
CC-IN2P3 site administrator)

Christine Leroy CEA/DAPNIA

c.leroy@cea.fr

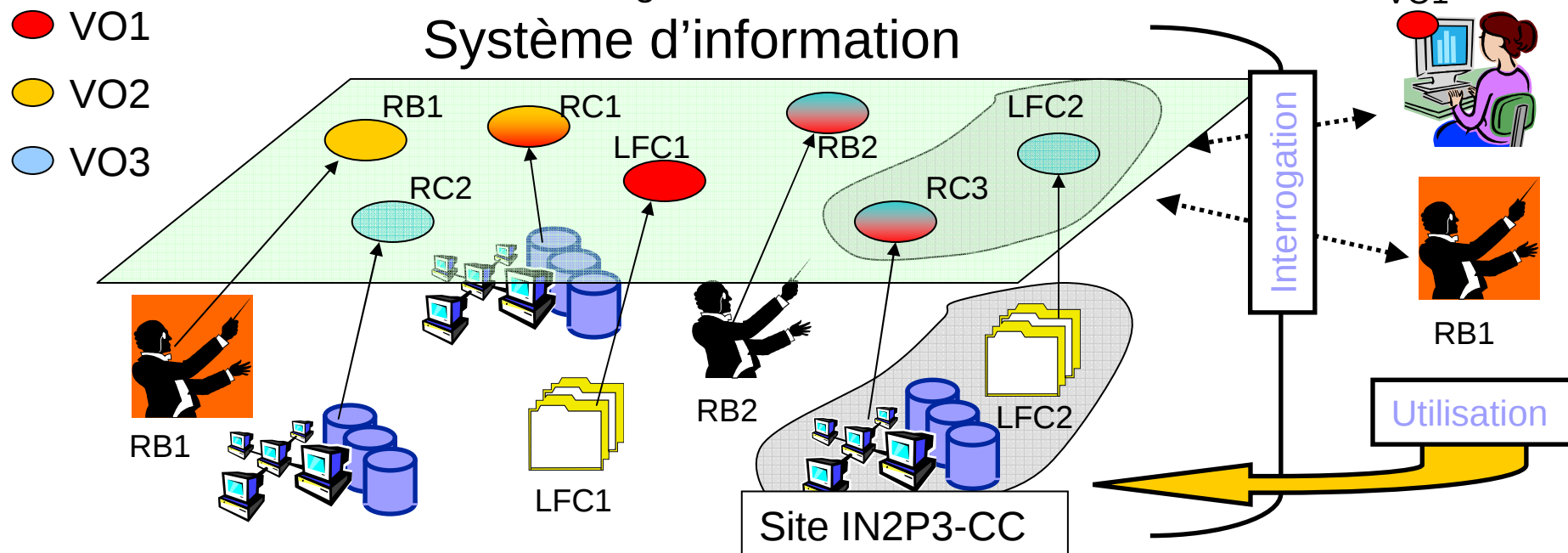
- Le système d'information dans EGEE/LCG
 - origine,
 - architecture,
 - protocole,
 - format des données
- Les outils du système d'information
- Un autre système : R-GMA
- Exploitation et Système d'information
- Conclusion

- Quoi ?
 - Système chargé de collecter des informations sur l'état des ressources/services mis à disposition sur la Grille.
- Pourquoi ?
 - Découvrir les ressources/services de la grille et leur nature
 - Disposer des données pertinentes pour utiliser les ressources et les services offerts par des sites hétéroclites.
 - Vérifier l'état de santé des ressources et services de la grille.
- Comment ?
 - En monitorant localement l'état et la description des ressources/services, et en publiant les données "fraîchement" collectées sur le système d'information.
 - En adoptant un modèle de données "commun à/connu de" tous les composants/acteurs de la grille qui ont besoin d'interagir avec les ressources/services de la grille.
 - En offrant les outils qui permettent d'alimenter et d'interroger le système d'information.

Système d'Information de LCG/EGEE

Principes

- Principes
 - Chaque site publie
 - Une description des ressources/services qu'il fournit par VO
 - L'état actuel de ses ressources (CPUs libres, Espace de stockage, etc.)
 - Chaque VO publie
 - Ce qu'elle a installé sur chaque site ("Tags" des Software Managers)
 - Les acteurs interrogent le SI pour savoir comment utiliser les services/ressources de la grille



Système d'Information de LCG/EGEE

Origine et Architecture



MDS: Monitoring and Discovery Service

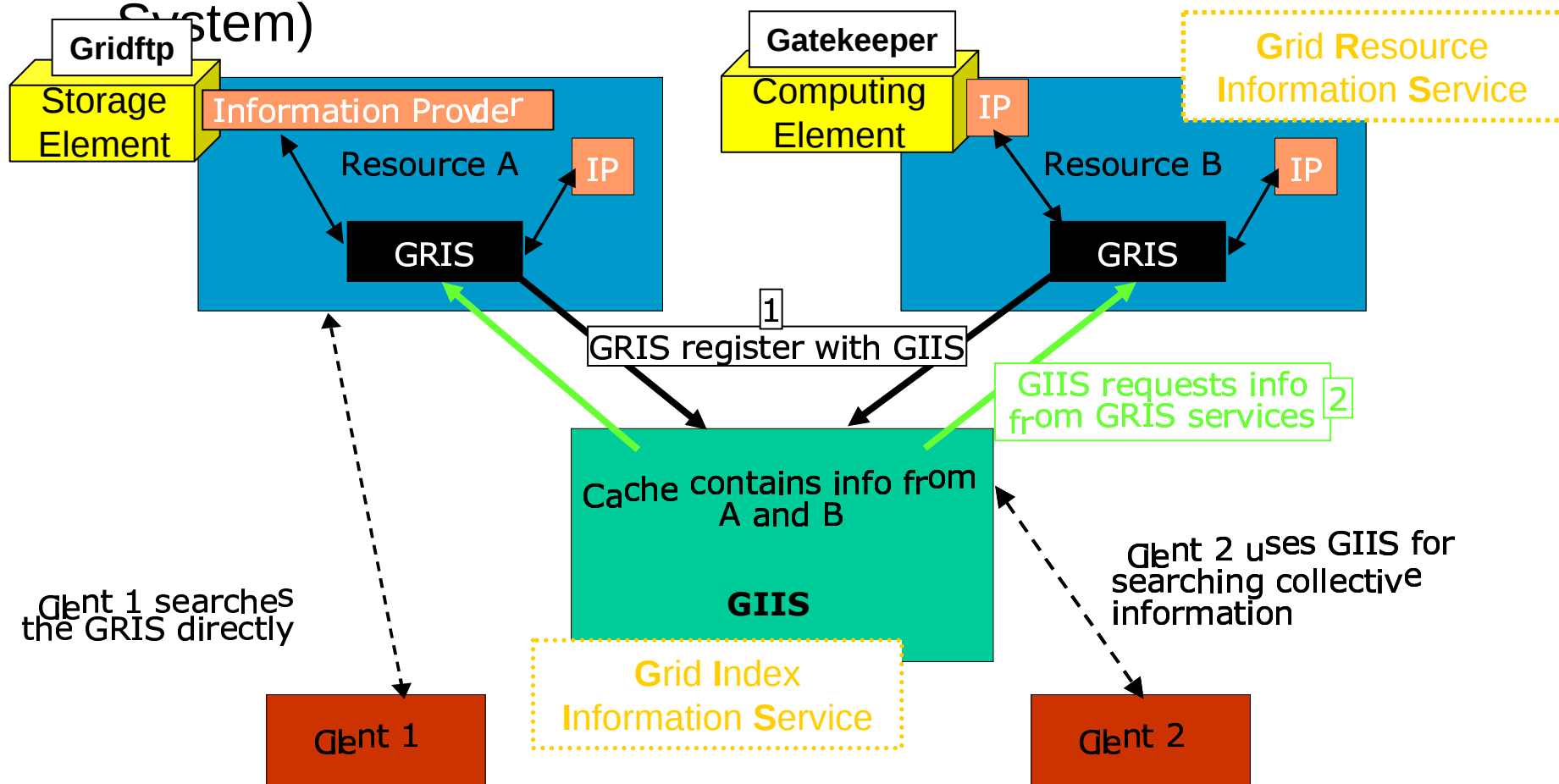
- ▶ Provient de la boîte à outils Globus
- ▶ Technologie utilisée pour le système d'information LCG/EGEE

Comment?:

- 1st. Sur chaque site l'état des services (informations statiques et dynamiques) est rapporté à des **serveurs**
- 2nd. Un **système centrale** interroge ces serveurs et stocke ces informations dans une base de données
- 3rd. Ces informations seront accessibles à travers le **protocole d'accès: LDAP**
- 4th. Le système centrale fournit l'information dans un **schema prédéfini: GlueSchema**

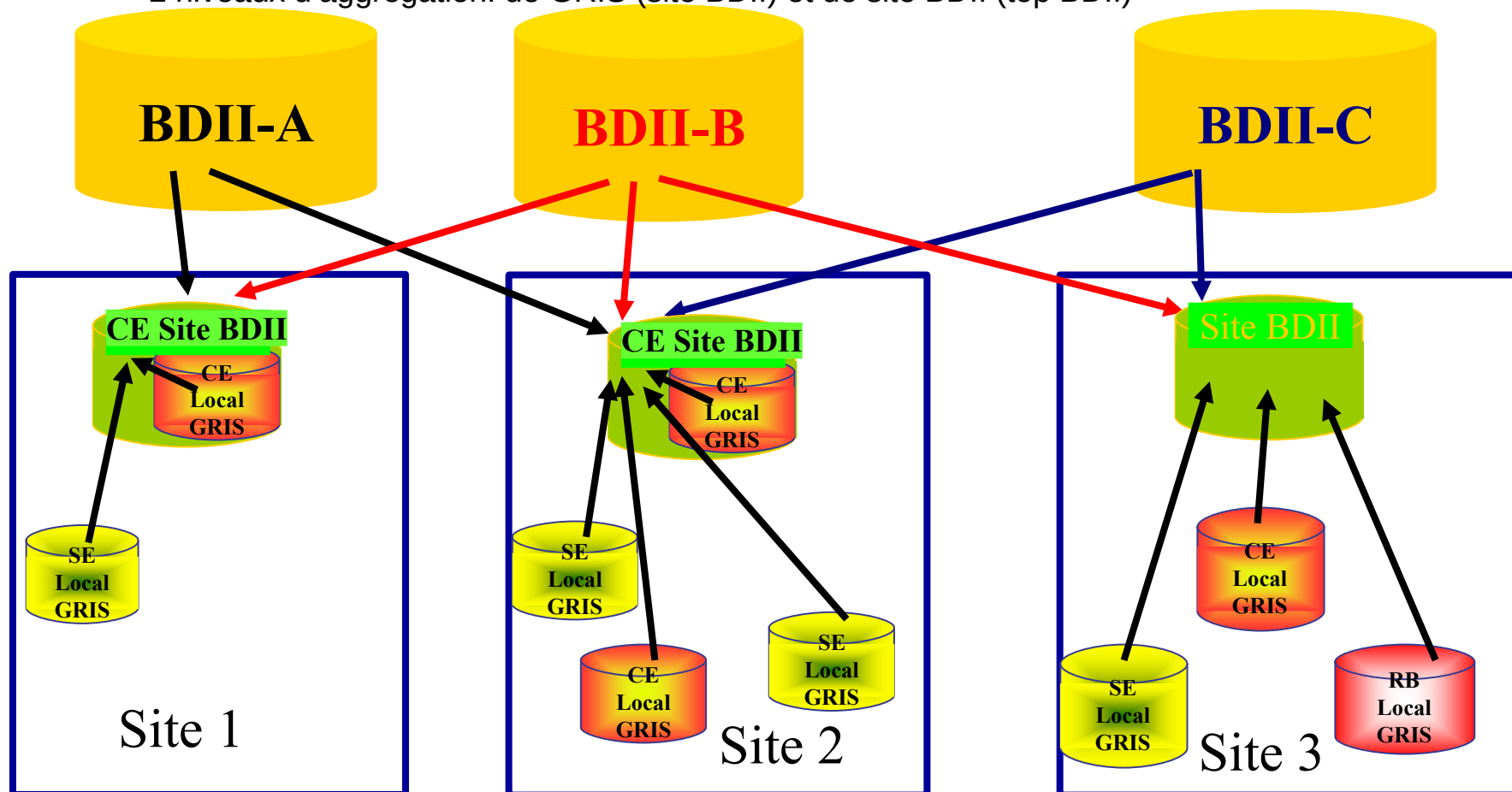
Architecture globale : Origine du SI actuel

- Globus: architecture MDS-2 (Monitoring and Discovery System)



Architecture globale : SI actuel

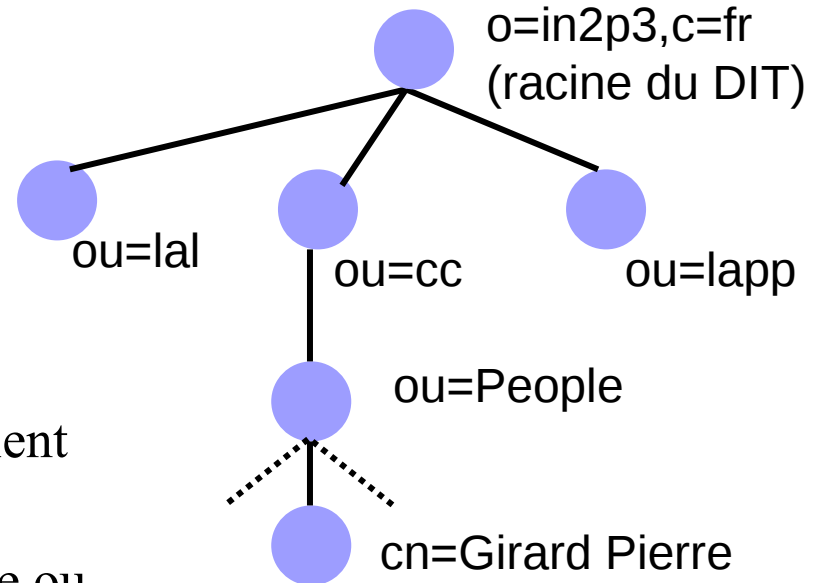
- LCG-2/GLITE : Aggrégations des systèmes d'information des sites
 - BDII (Berkeley Database Information Index): même fonctionnalité que le GIIS, avec une implémentation différente (meilleure gestion de cache)
 - 2 niveaux d'aggrégation: de GRIS (site BDII) et de site BDII (top BDII)



Mise en oeuvre Annuaire LDAP

Directory Information Tree

- GRIS, GIIS et BDII sont des serveurs d'annuaire LDAP (openldap)
 - LDAP : Lightweight Directory Access Protocol
- Modèle de données
 - Arborescence de nœuds
 - Chaque nœud (entrée de l'annuaire) contient des **attributs**
 - La structure d'un nœud est définie par une ou des **classes** au schéma prédéfini.
- Modèle de nommage
 - Distinguished Name (DN):
cn=Girard Pierre,ou=People,ou=cc,o=in2p3,c=fr
- L'information est importé/exporté depuis/vers les serveurs LDAP par des fichiers au format **LDIF** (LDAP Data Interchange Format)



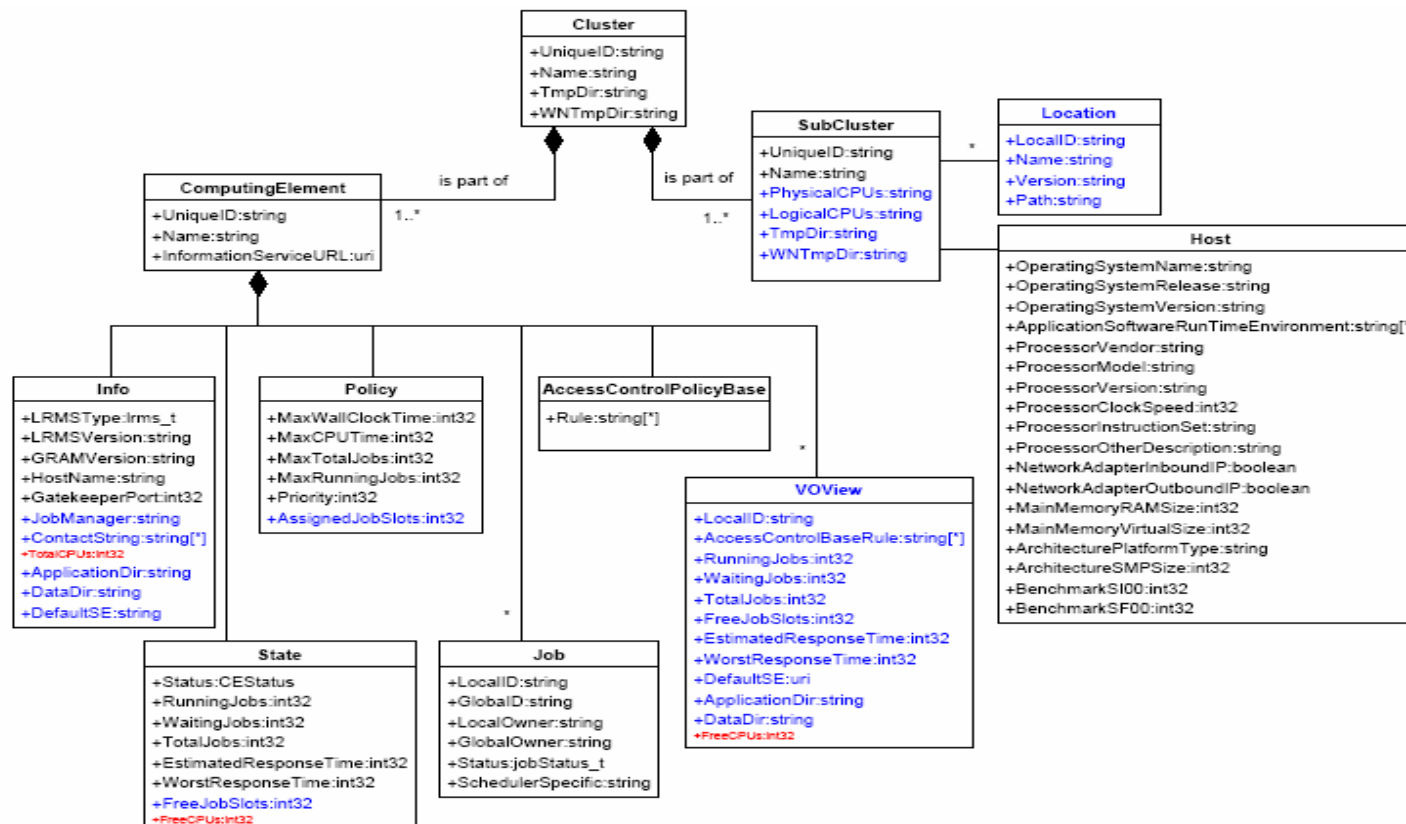
cn: Girard Pierre
phone: +33 9999999999
uid: girardpi
mail: pierre.girard@in2p3.fr

objectClass: top
objectClass: person
objectClass: organizationalPerson
objectClass: inetOrgPerson

Glue Schema in EGEE/LCG: Design

GLUE Schema (v1.3):

- Définition de schémas LDAP décrivant chaque composant des ressources/services de la grille
- <http://glueschema.forge.cnaf.infn.it/>



Quelques Exemples du Glue Schema (I)

1. Some General Attributes:

- Base class (**objectclass: GlueTop**): No attributes
- Schema Version Number (**objectclass: GlueSchemaVersion**)
 - **GlueSchemaVersionMajor**: Major Schema Version Number
 - **GlueSchemaVersionMinor**: Minor Schema Version Number

2. Attributes for the CE

- Base Class for the CE information (**objectclass: GlueCETop**) : No attributes
- CE (**objectclass: GlueCE**)
 - **GlueCEUniqueID**: unique identifier for the CE
 - **GlueCEName**: human-readable name of the service
- CE Status (**objectclass: GlueCEState**)
 - **GlueCEStateRunningJobs**: number of running jobs
 - **GlueCEStateWaitingJobs**: number of jobs not running
 - **GlueCEStateTotalJobs**: total number of jobs (running + waiting)
 - **GlueCEStateStatus**: queue status: queueing (jobs accepted but not running), production (jobs accepted and run), closed (neither accepted nor run), draining (jobs not accepted but those already queued are running)
 - **GlueCEStateWorstResponseTime**: worst possible time between the submission of the job and the start of its execution

Quelques Exemples du Glue Schema (II)

3. Attributes for the SE

- Base Class (**objectclass: GlueSETop**) : No attributes
- Architecture (**objectclass: GlueSLArchitecture**)
 - **GlueSLArchitectureType**: type of storage hardware (disk, tape, etc)
- Storage Service Access Protocol (**objectclass: GlueSEAccessProtocol**)
 - **GlueSEAccessProtocolType**: protocol type to access or transfer files
 - **GlueSEAccessProtocolPort**: port number for the protocol
 - **GlueSEAccessProtocolVersion**: protocol version
 - **GlueSEAccessProtocolAccessTime**: time to access a file using this protocol

4. Mixed Attributes

- Association between one CE and one or more SEs (**objectclass: GlueCESEBindGroup**)
 - **GlueCESEBindGroupCEUniqueID**: unique ID for the CE
 - **GlueCESEBindGroupSEUniqueID**: unique ID for the SE

Utilisation du Système d'Information

- Par les Ressources/services
- Par les utilisateurs

Utilisation du Système d'Information

Comment interroger les différents composants

Abbreviations:

BDII: Berkeley DataBase
Information Index

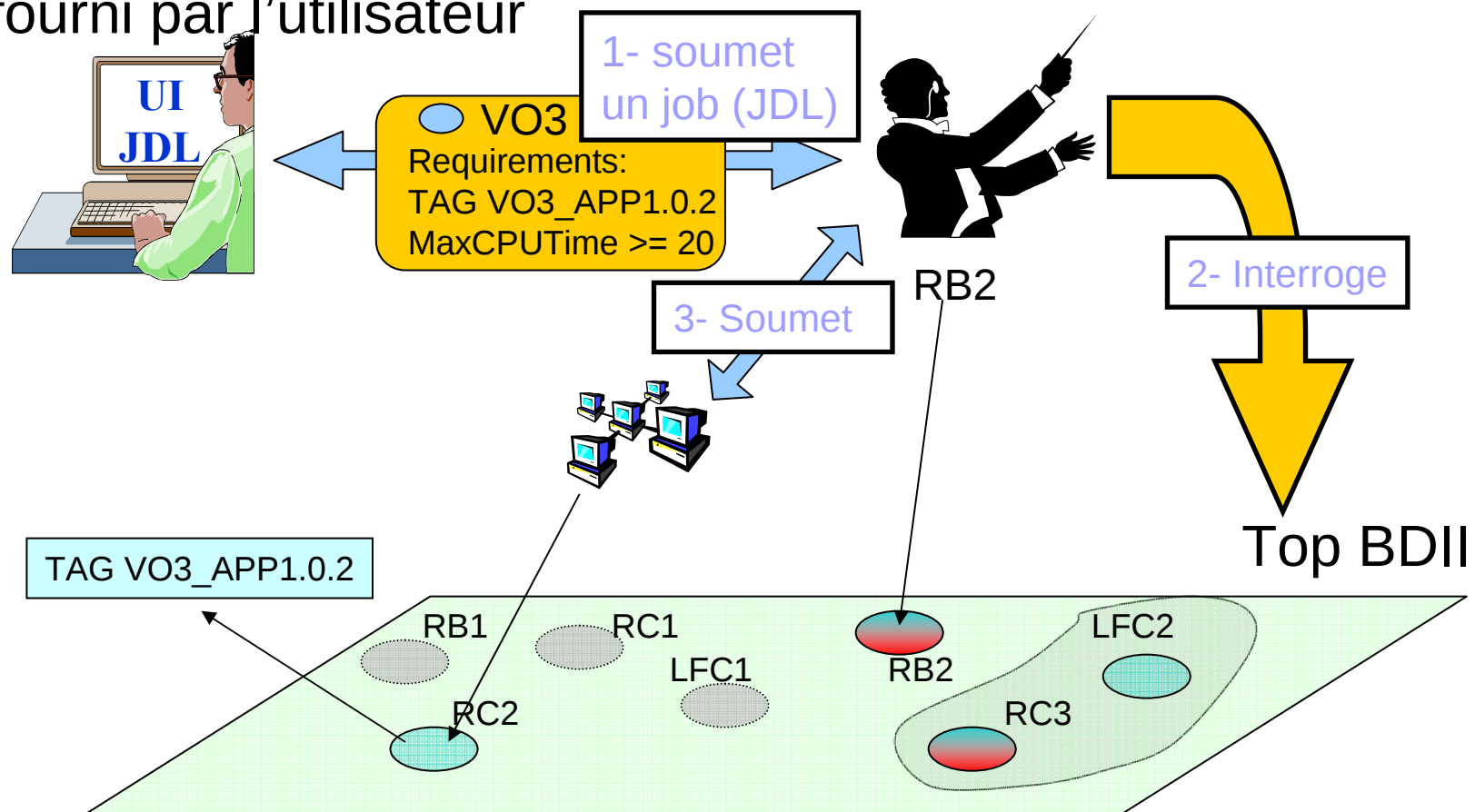
GRIS: Grid Resource
Information Server

Each site
can run
a **top BDII**. It
collects the information
coming from the site-BDII
`% ldapsearch -x -h <hostname>
-p 2170 -b "o=grid"`

At each site, a **site BDII** collects the
information
given by the GRISs
`% ldapsearch -x -h <hostname> -p 2170
-b "mds-vo-name=<name>,o=grid"`

Local GRISes run on CEs and SEs at each site and report
dynamic and static information
`% ldapsearch -x -h <hostname> -p 2135
-b "mds-vo-name=local,o=grid"`

- Le SI est interrogé par le RB pour choisir une queue de soumission en fonction du JDL (Job Description Language) fourni par l'utilisateur



Utilisation du Système d'Information niveau utilisateur - Idapsearch

♠ Idapsearch

```
% Idapsearch \
```

```
-x \
```

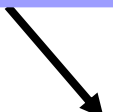
```
-H ldap://grid017.ct.infn.it:2170
```

```
-b 'mds-vo-name=local,o=grid' \
```

```
'(objectclass=GlueSE)' \
```

```
GlueSEUniqueID \
```

Read port of the BDII



Simple authentication

\ Uniform resource identifier

Base DN for search

Filter

Attributes to be returned

(Make “man Idapsearch” to retrieve the whole set of options)

The Idapsearch Implementation in EGEE/LCG

Some wrappers of Idapsearch exist in LCG middleware

Utilisation du Système d'Information niveau utilisateur - Idapsearch

```
$ Idapsearch -LLL -x -h topbdii.grif.fr:2170 -b mds-vo-name=in2p3-lapp,mds-vo-name=local,o=grid  
'(&(objectclass=GlueSA)(GlueSAType=permanent)(GlueSALocalID=atlas))'  
GlueSAStateAvailableSpace GlueSAStateUsedSpace
```

```
dn: GlueSALocalID=atlas,GlueSEUniqueID=lapp-se01.in2p3.fr,mds-vo-name=IN2P3-LA  
PP,mds-vo-name=local,o=grid  
GlueSAStateAvailableSpace: 198160000  
GlueSAStateUsedSpace: 2041965656
```

Sortie sous format LDIF

Utilisation du Système d'Information niveau utilisateur - lcg-infosites

> **lcg-infosites --vo <your_vo> feature --is <your_bdii>**

- It's mandatory to include the **vo** and the **feature**
- The **-is** option means the BDII you want to query. If not supplied, the BDII defined into the **LCG_GFAL_INFOSYS** will be interrogated

Features and descriptions:

closeSE	Names of the CEs where the user's VO is allowed to run together with their corresponding closest SEs
ce	Number of CPUs, running and waiting jobs and names of the CEs
se	SEs names together with the available and used space
lfc	Name of the lfc for the user's VO
all	It groups all the features just described
help	Description of the script

lcg-infosites: Examples (I)

```
> lcg-infosites --vo alice se --is lxb2006.cern.ch
```

```
*****
```

```
These are the data for alice: (in terms of SE)
```

```
*****
```

Avail Space (Kb)	Used Space (Kb)	SEs
-----	-----	-----
33948480	2024792	se.prd.hp.com
506234244	62466684	teras.sara.nl
1576747008	3439903232	gridkap02.fzk.de
1000000000000	500000000000	castorgrid.cern.ch
304813432	133280412	gw38.hep.ph.ic.ac.uk
651617160	205343480	mu2.matrix.sara.nl
1000000000000	1000000000	lcgads01.gridpp.rl.ac.uk
415789676	242584960	cclcgse01.in2p3.fr
264925500	271929024	se-a.ccc.ucl.ac.uk
668247380	5573396	seitep.itep.ru
766258312	681359036	t2-se-02.lnl.infn.it
660325800	1162928716	tbn17.nikhef.nl
1000000000000	1000000000000	castorftp.cnaf.infn.it
14031532	58352476	lcgse01.gridpp.rl.ac.uk
1113085032	1034242456	zeus03.cyf-kr.edu.pl
	[... ..]	

lcg-infosites: Examples (II)

```
> lcg-infosites --vo alice ce --is lxb2006.cern.ch
```

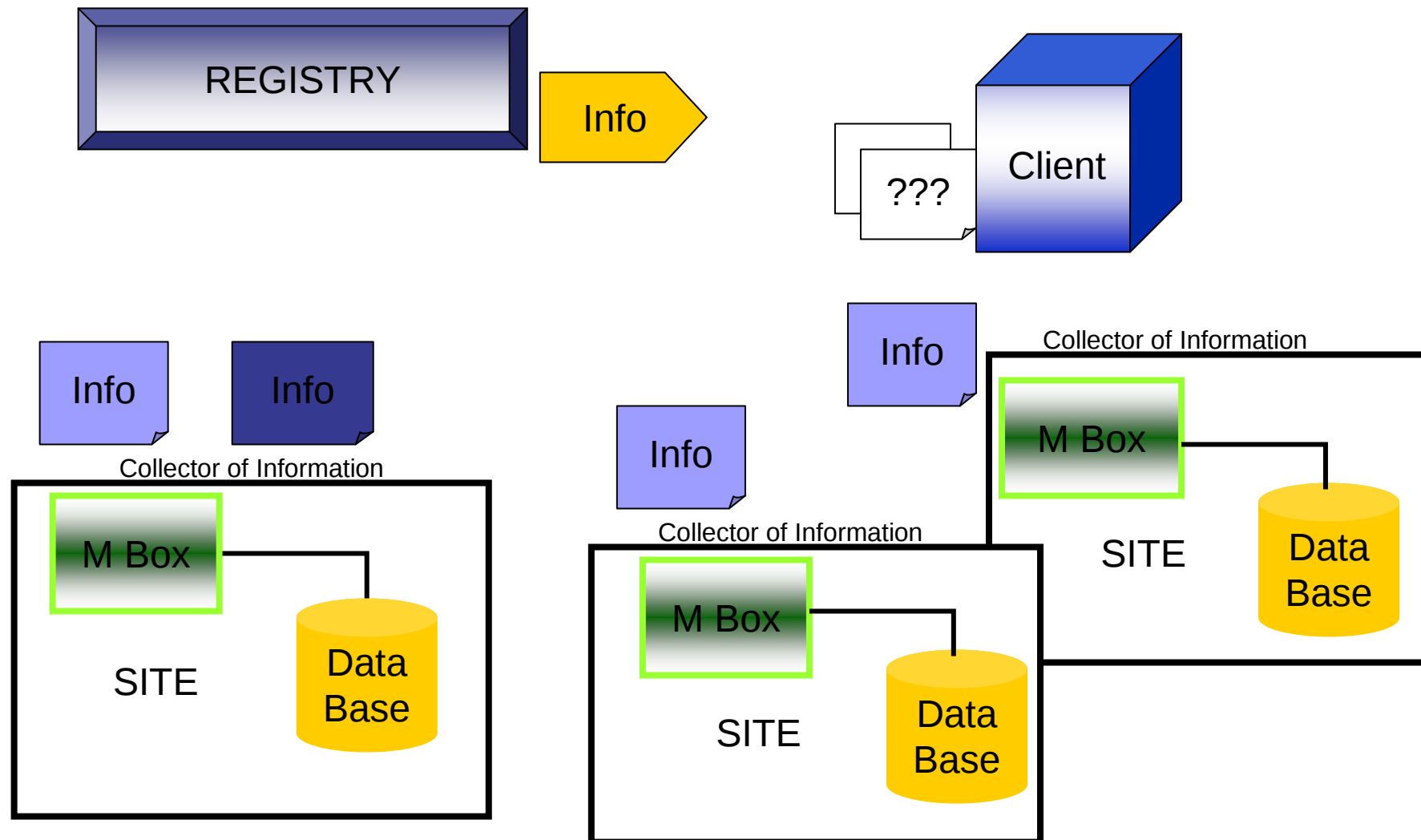
```
*****
These are the data for alice: (in terms of CPUs)
*****
#CPU   Free   Total Jobs   Running   Waiting   Computing Element
-----
52      51      0           0         0      ce.prd.hp.com:2119/jobmanager-lcgpbs-long
16      14      3           2         1      lcg06.sinp.msu.ru:2119/jobmanager-lcgpbs-long
[.....]
The total values are:
-----
10347  5565      2717      924      1793
```

R-GMA: Characteristics

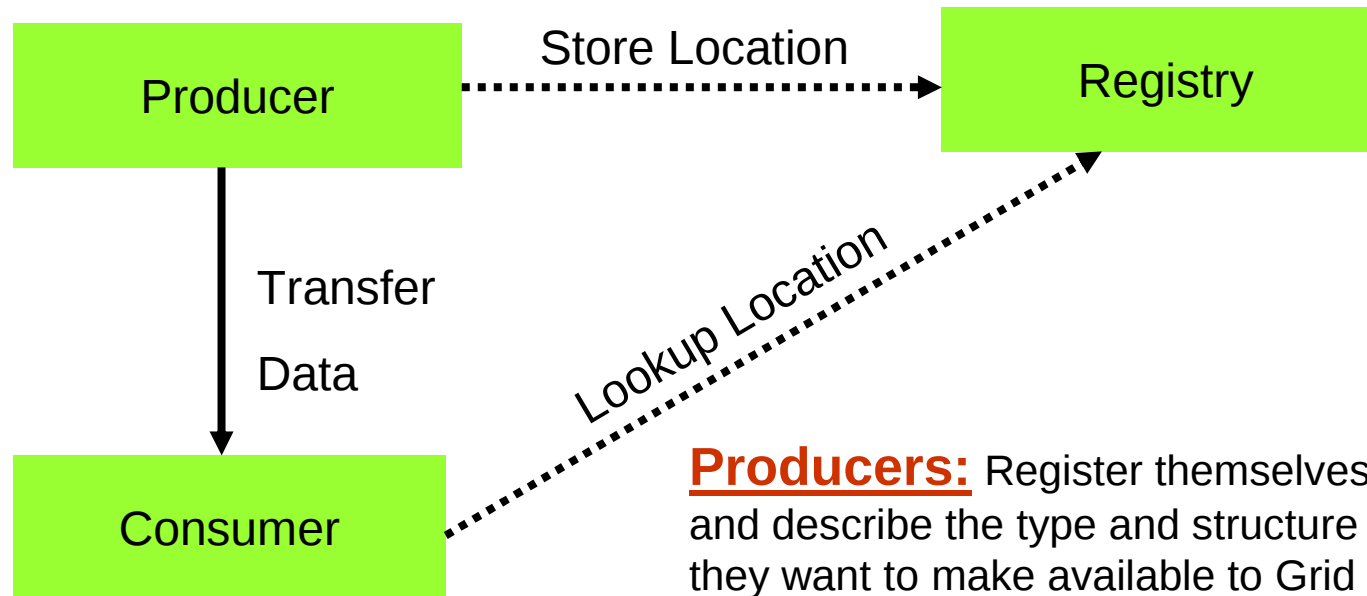
R-GMA (Relational Grid Monitoring Architecture):

- All data modeled as tables
- SQL as query language. It can express most queries in one expression
- You have a Relational DB for each VO
- allows cross queries between different entries
- Anyone can introduce new information in the system in a very easy way
- It is quite dynamic with new Producers of information being notified by existing Consumers

R-GMA: Design



R-GMA Architecture



Producers: Register themselves with the Registry and describe the type and structure of the information they want to make available to Grid

Consumers: Query the Registry to find out the information available and locate Producers which provide such information. They can connect directly the Producers

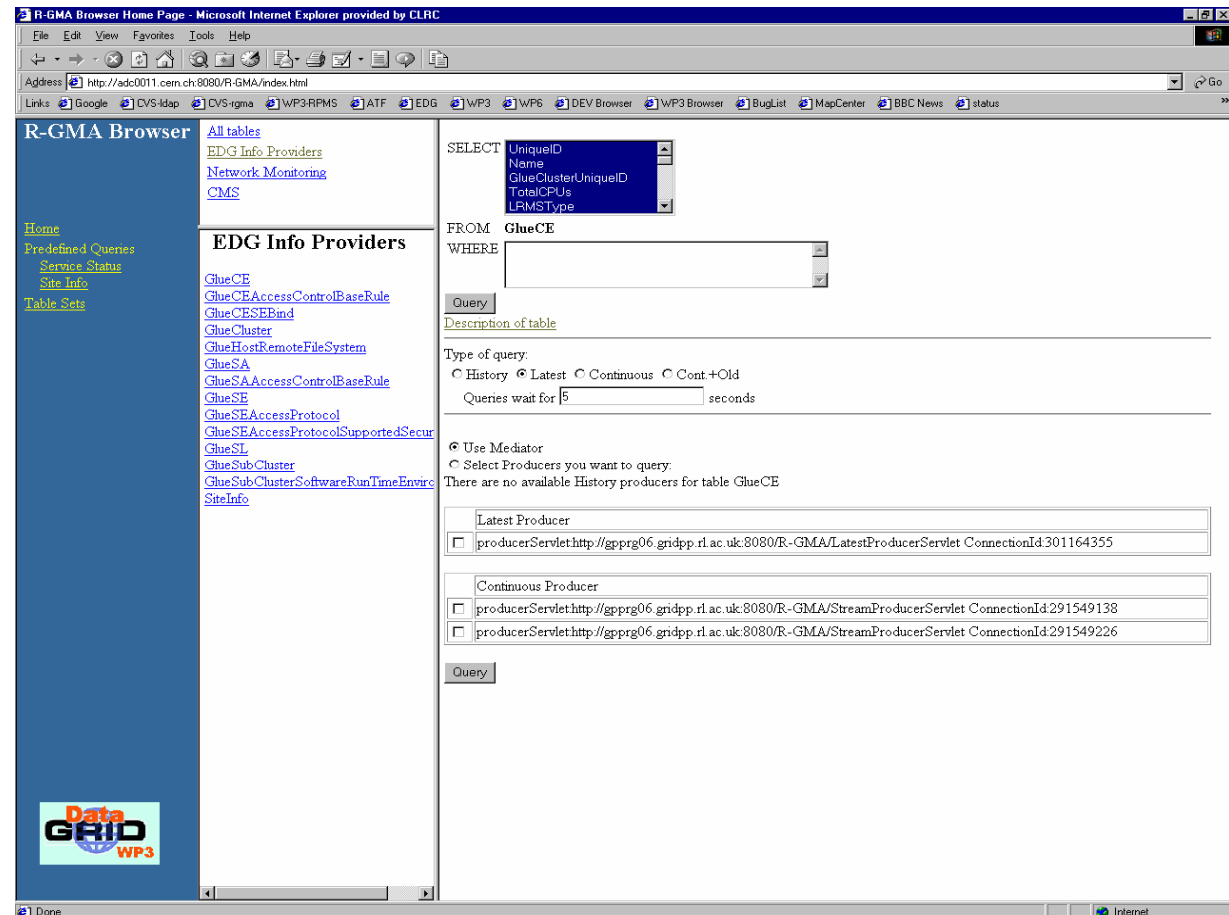
Registry: General collector, its arrow line represents the main flow of data

R-GMA tools: Browser

The user can retrieve the R-GMA information via the browser servlet

<http://lcgic01.gridpp.rl.ac.uk:8080/R-GMA/index.html>

It shows the schema, the registered producers and allows simple queries



edg-rgma: Virtual Database

- You can use it to produce or retrieve information
- Make **edg-rgma -c help** to retrieve more information

- \$ rgma -c "select Uniqueld, Type, Endpoint from GlueService where GlueSite_Uniqueld='IN2P3-LAPP' AND Type='ResourceBroker'"

- +-----+-----+-----+
- | Uniqueld | Type | Endpoint |
- +-----+-----+-----+
- | lapp-rb01.in2p3.fr:7772 | ResourceBroker | lapp-rb01.in2p3.fr:7772 |
- +-----+-----+-----+
- 1 rows

Exploitation et Système d'information GOC DB

- Référenciel de tous les sites de LCG/EGEE
 - l'URL LDAP du Site BDII de chaque site
 - Le statut/type (Certified/Production) de chaque site
 - La déclaration éventuelle de « Scheduled Downtime »
 - ...
- ➔ Utilisable pour générer automatiquement la configuration des Top BDIIs avec la liste des sites déclarés en « Production »
- ➔ <http://grid-deployment.web.cern.ch/grid-deployment/gis/lcg2-bdii/dteam/lcg2-all-sites.conf>

The screenshot shows the GOCDB website interface in a Windows Internet Explorer browser. The page title is "GOCDB :: Site info for GRIF". The URL in the address bar is <https://goc.gridops.org/site/list?id=239>. The page displays site information for GRIF, including its official name, domain, home URL, contact email, contact telephone, emergency telephone, operating hours, time zone, GIIIS URL, tier, and primary site status.

Overlaid on the screenshot is a diagram illustrating the data flow. At the top, a cylinder labeled "GOC DB" is connected by a double-headed arrow labeled "HTTP" to a box labeled "Config". The "Config" box is connected by a double-headed arrow labeled "Config" to a green box containing the text "URL SiteBDII1" and "URL SiteBDII2". This green box is connected by a double-headed arrow to a red cylinder labeled "BDII".

Exploitation et Système d'information

Tests fonctionnels et filtrage du SI

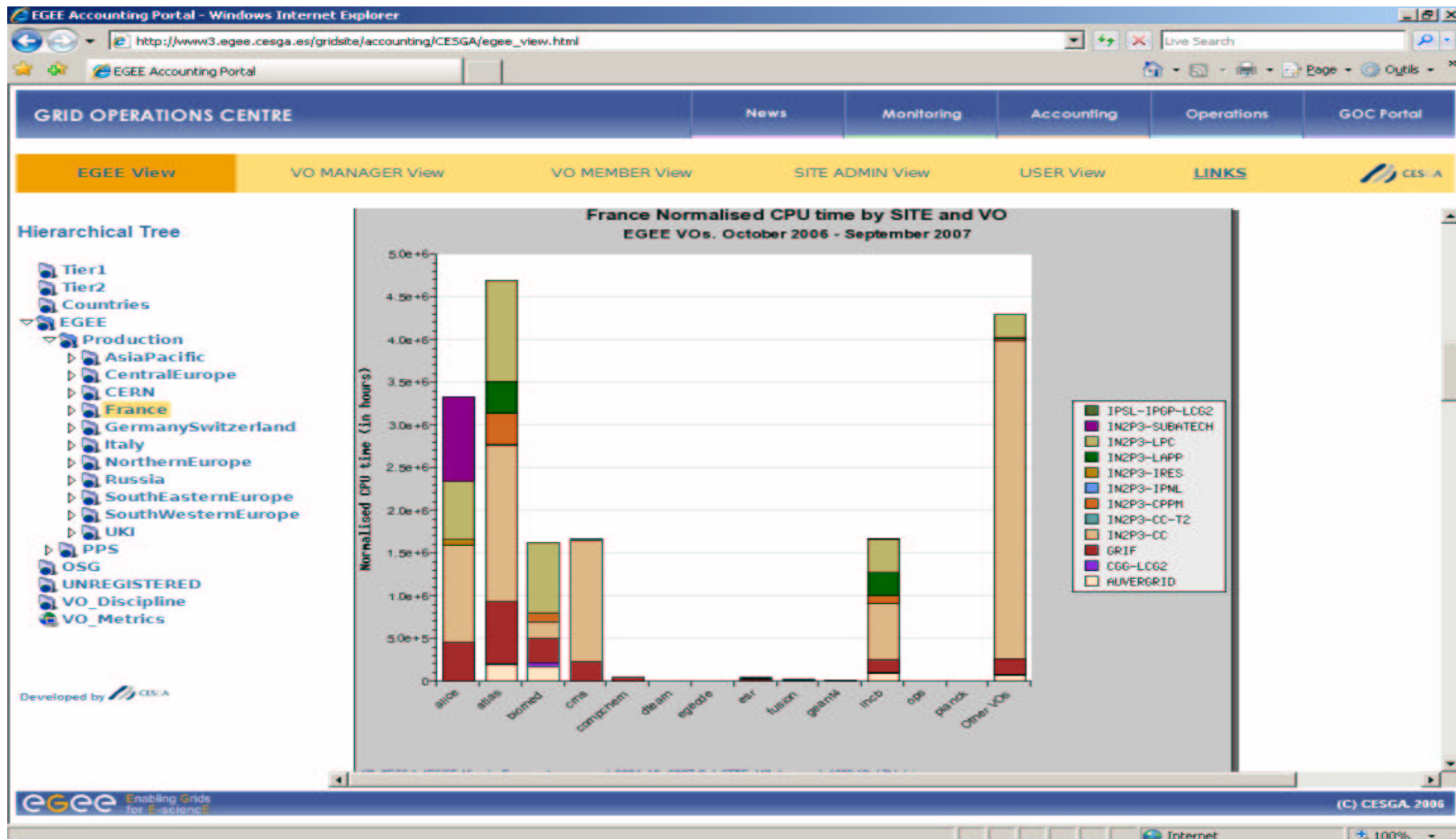
- SAM (Service Availability Monitoring)
 - <https://lcg-sam.cern.ch:8443/sam/sam.py>
 - système soumettant régulièrement des tests sur les sites
 - Le résultat des tests est utilisable pour filtrer les Top BDIIs en supprimant les sites en échec.

SAM -- 2007/01/09 - 15:15:05 - Mozilla Firefox

https://lcg-sam.cern.ch:8443/sam/sam.py?sensors=CE®ions=France&vo=ops&ord

No	RegionName	SiteName	NodeName	Status
1	France	AUVERGRID	int15auvergridce01.univ-bpclermont.fr	OK
2	France	AUVERGRID	int43auvergridce01.univ-bpclermont.fr	MAINT
3	France	AUVERGRID	obsauvergridce01.univ-bpclermont.fr	OK
4	France	CGG-LCG2	ce1.egee.fr.cgg.com	ERROR
5	France	GRIF	grid10.lal.in2p3.fr	OK
6	France	GRIF	ipnl2001.in2p3.fr	OK
7	France	GRIF	lpnce.in2p3.fr	OK
8	France	GRIF	polgrid1.in2p3.fr	OK
9	France	IN2P3-CC	cclegceli02.in2p3.fr	OK
10	France	IN2P3-CC	cclegceli03.in2p3.fr	OK
11	France	IN2P3-CPPM	marseillece01.mrs.grid.cnrs.fr	OK
12	France	IN2P3-IRES	sbgat417.in2p3.fr	OK
13	France	IN2P3-LAPP	lapp-ce01.in2p3.fr	OK
14	France	IN2P3-LPC	chlcgce01.in2p3.fr	OK
15	France	IN2P3-LPC	chlcgce02.in2p3.fr	OK
16	France	IN2P3-LPC	chlcgce03.in2p3.fr	OK
17	France	IN2P3-SUBATECH	nanlcg01.in2p3.fr	OK
18	France	IPSL-IPGP-LCG2	hudson.datagrid.jussieu.fr	MAINT

Exploitation et Système d'information l'Accounting



Conclusions

- Le SI LCG/EGEE fournit les informations nécessaires à l'utilisation des ressources/services de la grille
 - Pour les utilisateurs
 - Pour les services de grille
 - ~200 sites sont ainsi publiés
- Il est organisé autour d'une architecture à 3 niveaux de serveurs LDAP
 - GRIS $\Rightarrow$ Site BDII $\Rightarrow$ Top BDII
 - Utilise un modèle de données défini par le GLUE Schema
 - Attention aux temps de latence liés à l'accumulation de cache
- Criticité du SI
 - Stabilité en progression : répartition des Top BDIIs par ROC (fédérations)
- R-GMA (Relational Grid Monitoring Architecture), l'alternative ?
 - Approche « BD Relationnelle distribuée »
 - Possibilité d'extension du modèle de données du SI
 - Annoncé comme le remplaçant de l'architecture basé sur LDAP
 - $\Rightarrow$ Ne semble pas encore avoir fait ses preuves: utilisé seulement pour l'accounting

Exercices: Système d'Information

A) Il existe 2 systèmes d'information sur la Grille de Calcul EGEE/LCG.

- 1) Le plus ancien et celui qui est toujours utilisé est basé sur LDAP, il s'agit du MDS avec ses BDII et GRIS.

=> exercices avec ldapsearch et outils LCG/EGEE (lcg-infosites ; lcg-info)

- 2) Le plus récent est utilisé pour la comptabilité, il s'agit de R-GMA. Il est basé sur TOMCAT et MySQL.

B) Ou trouver des informations relatives aux sites, aux VO.

Les sites de monitoring et d'accounting

Exercices: A) Système basé sur LDAP

a) LDAPSEARCH

- 1. En utilisant la commande `ldapsearch` interrogez le GRIS présent sur le CE *du LAPP*:

`ldapsearch -x -h lapp-ce01.in2p3.fr -p 2135 -b 'mds-vo-name=local, o=grid'`

- 2. Interrogez le BDII du site (site-BDII):

`ldapsearch -x -h lapp-site01.in2p3.fr -p 2170 -b 'mds-vo-name=IN2P3-LAPP, o=grid'`

Exercices: A) Système basé sur LDAP

a) LDAPSEARCH

- 3. Recherchez toutes les VO acceptées par le RB du site

```
ldapsearch -x -h lapp-site01.in2p3.fr -p 2170 -b 'mds-vo-name=IN2P3-LAPP, o=grid' \  
'objectclass=GlueService' GlueServiceOwner
```

- 4. Recherchez les tags logiciels présents sur ce site

```
ldapsearch -x -h lapp-site01.in2p3.fr -p 2170 -b 'mds-vo-name=IN2P3-LAPP, o=grid' \  
objectclass=GlueSubCluster GlueHostApplicationSoftwareRunTimeEnvironment
```

Exercices: A) Système basé sur LDAP b) LCG-INFOSITES

- The **lcg-infosites** command can be used as an easy way to retrieve information on Grid resources for most use cases.

USAGE: `lcg-infosites --vo <vo name> options -v <verbose level> --is <BDII to query>`

The "lcg-infosites" command is actually just a perl script wrapping a series of LDAP commands and was developed to allow the user to retrieve information on Grid resources for the most common cases.

Before beginning it is worth observing that "lcg-infosites" does not use your VOMS proxy certificate and hence all commands need to include the option "--vo " (note that you can actually query the information for any VO).

Exercices: A) Système basé sur LDAP b) LCG-INFOSITES

lcg-infosites options

ce	The information related to number of CPUs, running jobs, waiting jobs and names of the CEs are provided. All these data group all VOs together. With "-v 1" only the names of the queues will be printed while with "-v 2" The RAM Memory together with the operating system and its version and the processor included in each CE are printed.
se	The names of the SEs supported by the user's VO together with the kind of Storage System, the used and available space will be printed. With "-v 1" only the names of the SEs will be printed.
closeSE	The names of the CEs where the user's VO is allowed to run together with their corresponding closest SEs are provided.
lfc	Name of the lfc Catalog for the user's VO.
tag	The names of the tags relative to the software installed in site is printed together with the corresponding CE.
all	It groups together the information provided by ce, se, lfc and rmc.
is	If not specified the BDII defined in default by the variable LCG GFAL INFOSYS will be queried. However the user may want to query any other BDII without redefining this environment variable. This is possible specifying this argument followed by the name of the BDII which the user wants to query. All options admits this argument.

Exercices: A) Système basé sur LDAP b) LCG-INFOSITES

- 5. Recherchez tous les CE pour une VO

lcg-infosites --vo vo.lapp.in2p3.fr ce

- 6. Recherchez les SE pour une VO

lcg-infosites --vo vo.lapp.in2p3.fr se

- 7. The closest SE to a CE is defined by the manager of the CE. To see the closest SE to each CE use the command:

lcg-infosites --vo vo.lapp.in2p3.fr closeSE

Exercices: A) Système basé sur LDAP c) LCG-INFO

- **lcg-info examples**

The "lcg-info" command is similar to the "lcg-infosites" except that it is used to list either CE's or SEs satisfying a given set of conditions on their attributes and to print, for each of them, the values of a given set of attributes.

- 8. The first step is to view the list of attributes available for both the matching and printing aspects of 'lcg-info' :

lcg-info --list-attrs

,

Exercices: A) Système basé sur LDAP c) LCG-INFO

lcg-info options

--list-attrs	Prints a list of the attributes that can be queried.
--list-ce	Lists the CEs which satisfy a query, or all the CEs if no query is given.
--list-se	Lists the SEs which satisfy a query, or all the SEs if no query is given.
--query	Restricts the output to the CEs (SEs) which satisfy the given query.
--bdii	Allows to specify a BDII in the form :. If not given, the value of the environmental variable LCG_GFAL_INFOSYS is used. If that is not defined, the command returns an error.
--sed	Print the output in a "sed-friendly" format.
--attrs	Specifies the attributes whose values should be printed.
--vo	Restricts the output to CEs or SEs where the given VO is authorized. Mandatory when VO-dependent attributes are queried upon.

Exercices: A) Système basé sur LDAP c) LCG-INFO

- **9. "lcg-info"** is very similar to the usage of the "Requirements" tag in a JDL file along with the command "glite-job-list-match". The **"lcg-info"** command can therefore be **useful when constructing the "Requirements" tag in a JDL file**. Note that "lcg-info" does not in most (but not all) cases need a VO to be specified.

lcg-info --vo lapp --list-ce --query 'Tag=MPICH' --attrs 'FreeCPUs,Tag'

- 10. "lcg-info" can also query the attributes of CE's and searches for the closest CE to the SE "lapp-se01.in2p3.fr".

lcg-info --list-se --query 'SE=lapp-se01.in2p3.fr' --attrs CloseCE

Exercices: Système basé sur R-GMA

- Les informations contenues dans R-GMA sont visualisables via un navigateur ou via ligne de commande.
- 1. Visualisez les CE de la Grille de Calcul. Avec votre navigateur allez sur le R-GMA du site :

<https://lapp-mon01.in2p3.fr:8443/R-GMA/>

Cliquez sur le lien en haut « Tables Sets » puis « [GLUE Info Providers](#) » puis « [BlueCE](#) » et sur la droite cliquez sur le lien « [Query this table](#) », with type of query '[continuous](#)'

- 2. Utilisez la ligne de commande. Sur l'UI, tapez la commande *rgma*

Exécutez l'aide et visualisez des exemples :

[help examples](#)

Exercices: Système basé sur R-GMA

- 3. Visualisez les tables présentes

show tables

Visualisez la description d'une table

describe GlueService

- 4. Faites quelques commandes SQL.

set query continuous

set timeout 120 seconds

La liste des CE :

*select * from GlueService*

Les services d'un site :

*select * from GlueService where GlueSite_UniqueId='IN2P3-LAPP'*

Exercices: Système basé sur R-GMA

- 5. Visualisez les CE avec une autre commande :

*select * from GlueCE*

Les CE avec des CPU de libre :

select UniqueId, FreeCpus from GlueCE where FreeCpus > 10

- 6. Pour sortir :

quit

Où trouver des informations relatives aux sites, aux VOs.

Les sites de monitoring et d'accounting



- 1. Info sur les sites: avec votre navigateur, connectez-vous sur le site du GOC page du LAPP :

<https://goc.gridops.org/site/list?id=128>

Observez les informations présentées.

- 2. Etat du site: Avec votre navigateur, connectez-vous sur le site des tests SAM:

<https://lcg-sam.cern.ch:8443/sam/sam.py>

Sélectionner les tests « CE »

Ou trouver des informations relatives aux sites, aux VOs. Les sites de monitoring et d'accounting



- 3. informations sur les VOs: avec votre navigateur, connectez-vous sur le site CIC:

<http://cic.gridops.org/index.php?section=home&page=homepage>

Selectionner le menu « vo users »

Selectionner une VO de votre choix et observer son IDCard

- 4. accounting: avec votre navigateur, connectez-vous sur:

http://www3.egee.cesga.es/gridsite/accounting/CESGA/egee_view.html

Dans l'arborescence de gauche, choisir France

Observez les différents Graphes