



Soumission de jobs de calcul

David Bouvet, David Weissenbach
Observatoire de Meudon, 01-02/02/2011



Plan

- Rappel noeuds de la grille
- Soumission de job : *proxy* et scénario
- JDL
- Commandes de soumission
- *Job perusal*
- *VO software area*



Rappel nœuds de la grille

- **UI (*User Interface*) : point d'accès à la grille WLCG/EGI**
 - n'importe quelle machine sur laquelle l'utilisateur a un compte personnel
 - fournit CLI pour soumission/gestion des jobs, lister les ressources, gérer les données sur la grille
- **CE (*Computing Element*) : interface entre la grille et le système de batch du site**
- **WN (*Worker Node*) : noeuds sur lesquels tournent les jobs**
- **SE (*Storage Element*) : point d'accès aux ressources de stockage de données (serveurs de disques, système de stockage de masse)**
 - supporte différents types de protocole/interface d'accès aux données



Soumission de jobs

- L'utilisateur soumet un job via le WMS (*Workload Management System*) de la grille
- Le WMS essaie d'optimiser l'utilisation des ressources et d'exécuter les jobs des utilisateurs le plus rapidement possible
- Le WMS interagit avec les noeuds suivants :
 - UI (*User Interface*) : point d'accès pour les utilisateurs
 - LB (*Logging and Bookeeping*) : stocke les informations concernant le job pour des requêtes utilisateurs.
 - BDII (*Information Index*) : un serveur LDAP qui collecte les informations concernant les ressources grille. Il est utilisé par le RB pour sélectionner les ressources
 - catalogue de fichiers

➤ Fin de transition entre 2 générations de RB/WMS

- LCG RB : en fin de vie, peu de fonctionnalités avancées, performances limitées
 - Commandes : `edg-job-xxx`
- gLite WMS :
 - Commandes : `glite-wms-xxx`
 - délégation de proxy (WMS) : nécessaire pour interagir avec le WMS (WMPProxy)
 - automatique : option `-a`, effectuée lors de la soumission
 - explicite : `glite-wms-job-delegate-proxy + -d` à la soumission
 - VOMS *proxy renewal* (y compris les attributs VOMS)
 - soumission de jobs par lot



Plan : vous êtes ici

- Rappel nœuds de grille
- Soumission de job : *proxy* et scénario
- JDL
- Commandes de soumission
- *Job perusal*
- *VO software area*



Soumission de jobs : création d'un proxy

- **voms-proxy-init -voms vo.formation.idgrilles.fr**
Cannot find file or dir: /afs/in2p3.fr/home/d/dbouvet/.glite/vomses
Your identity: /O=GRID-FR/C=FR/O=CNRS/OU=CC-LYON/CN=David Bouvet
Enter GRID pass phrase:
Creating temporary proxy Done
Contacting cclcgvomsli01.in2p3.fr:15001 [/O=GRID-FR/C=FR/O=CNRS/OU=CC-LYON/CN=cclcgvomsli01.in2p3.fr] "vo.formation.idgrilles.fr" Done
Creating proxy Done
Your proxy is valid until Sat Mar 13 02:56:14 2010
- **voms-proxy-info -all**
subject : /O=GRID-FR/C=FR/O=CNRS/OU=CC-LYON/CN=David Bouvet/CN=proxy
issuer : /O=GRID-FR/C=FR/O=CNRS/OU=CC-LYON/CN=David Bouvet
identity : /O=GRID-FR/C=FR/O=CNRS/OU=CC-LYON/CN=David Bouvet
type : proxy
strength : 512 bits
path : /tmp/x509up_u2028
timeleft : 11:58:25
=== VO vo.formation.idgrilles.fr extension information ===
VO : vo.formation.idgrilles.fr
subject : /O=GRID-FR/C=FR/O=CNRS/OU=CC-LYON/CN=David Bouvet
issuer : /O=GRID-FR/C=FR/O=CNRS/OU=CC-LYON/CN=cclcgvomsli01.in2p3.fr
attribute : /vo.formation.idgrilles.fr/Role=NULL/Capability=NULL
timeleft : 11:58:25



Soumission de job : proxy

- `voms-proxy-init -voms cms -valid 24:00`
- `openssl x509 -in /tmp/x509up_u`id -u` -text`

Certificate:

Data:

Version: 3 (0x2)

Serial Number: 2239 (0x8bf)

Signature Algorithm: md5WithRSAEncryption

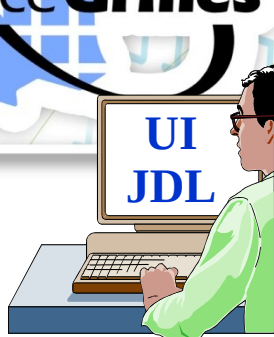
Issuer: C=IT, O=GILDA, OU=Personal

Certificate, L=CLERMONT-FERRAND, CN=CLERMONT-FERRAND01/Email=emmanuel

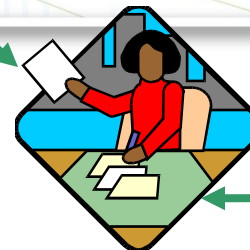
[.medernach@clermont.in2p3.fr](mailto:medernach@clermont.in2p3.fr)

Validity...

Soumission de jobs : scénario



Information System (IS)



Resource Broker (RB)



Logging & Bookkeeping (LB)

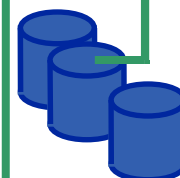


Job Submission Service (JSS)

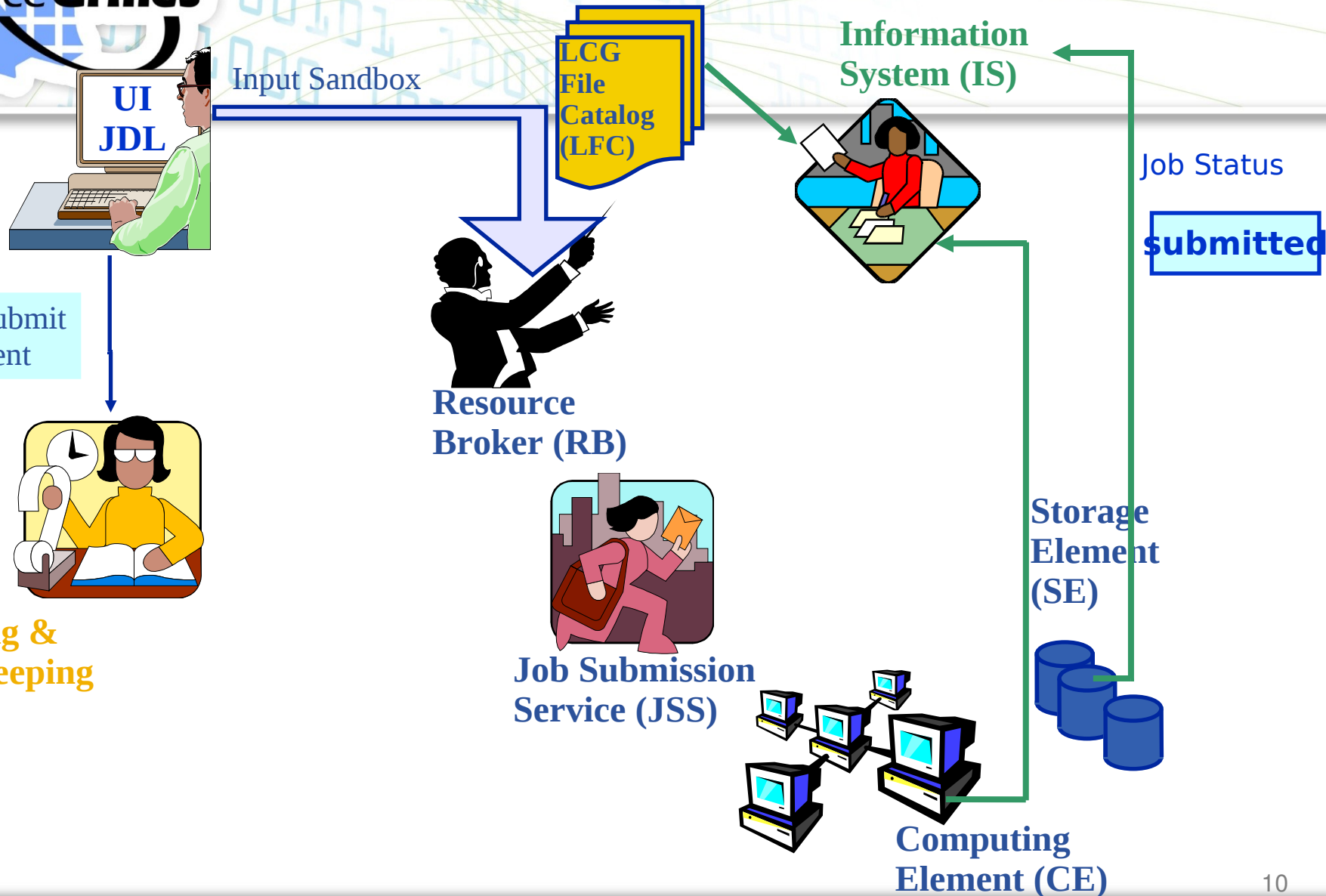


Computing Element (CE)

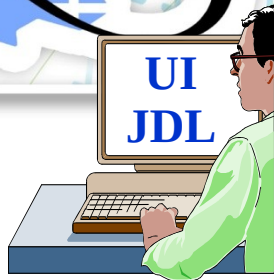
Storage Element (SE)



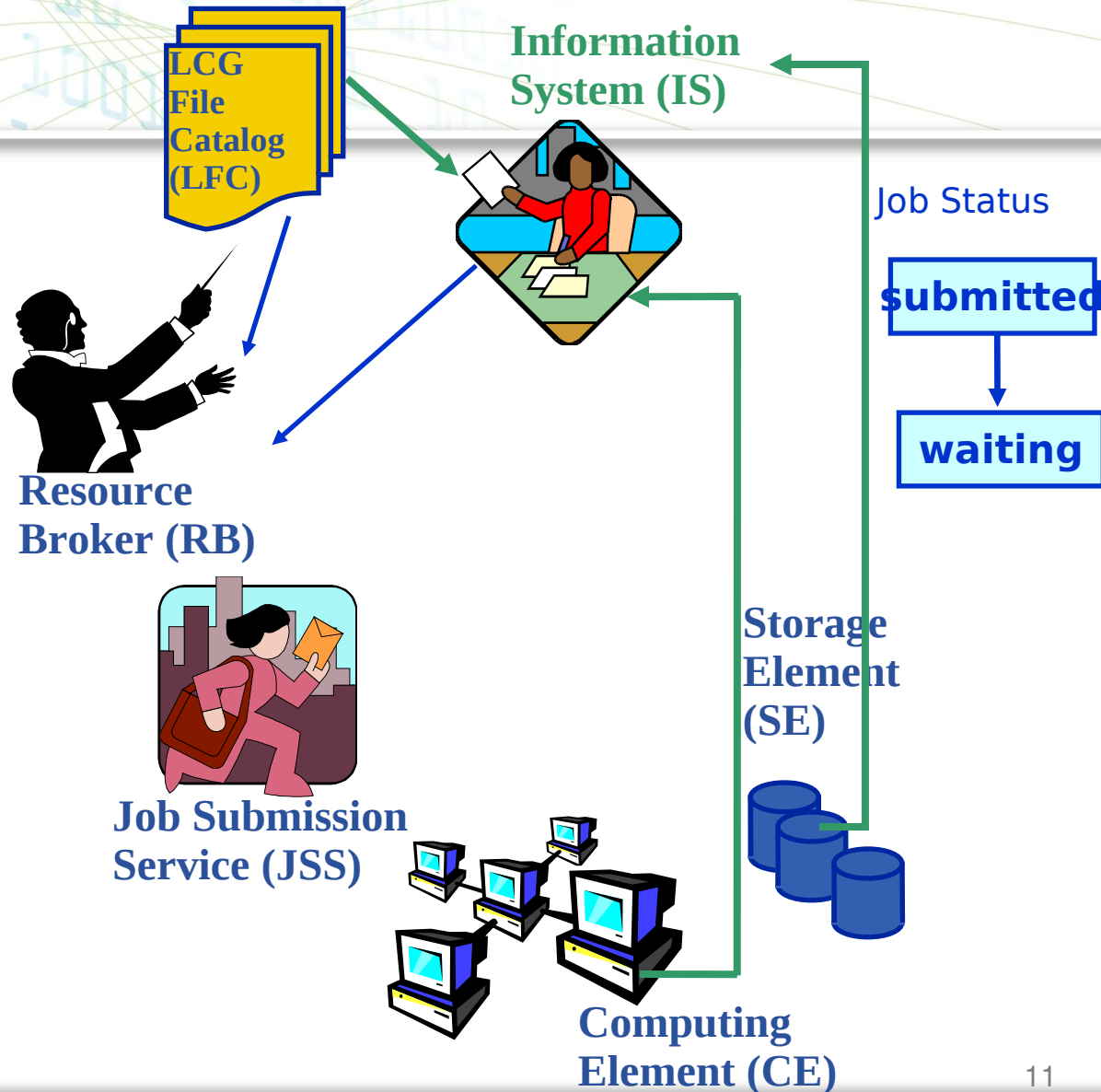
Soumission de jobs : scénario



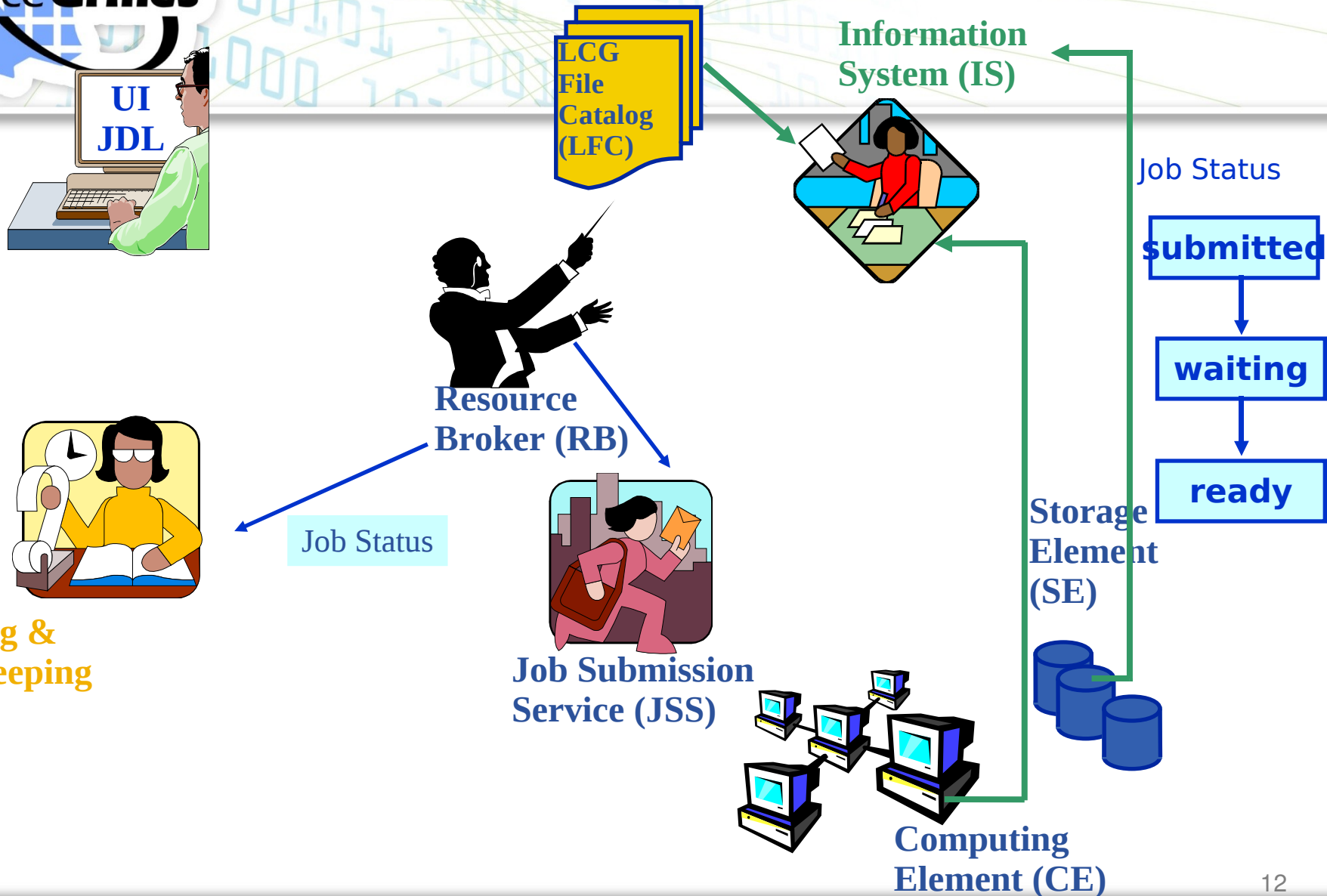
Soumission de jobs : scénario



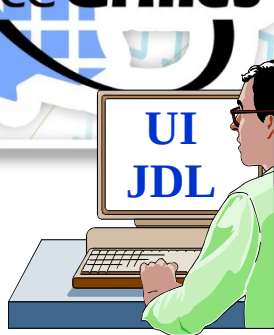
Logging &
Bookkeeping
(LB)



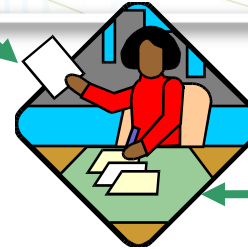
Soumission de jobs : scénario



Soumission de jobs : scénario



Information System (IS)



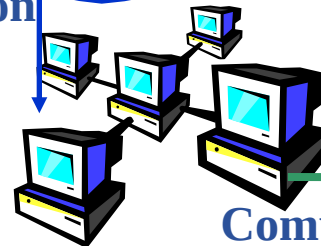
Resource Broker (RB)



Logging & Bookkeeping (LB)

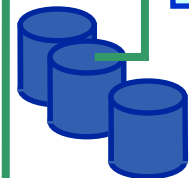


Job Submission Service (JSS)



Computing Element (CE)

Storage Element (SE)



Job Status

submitted

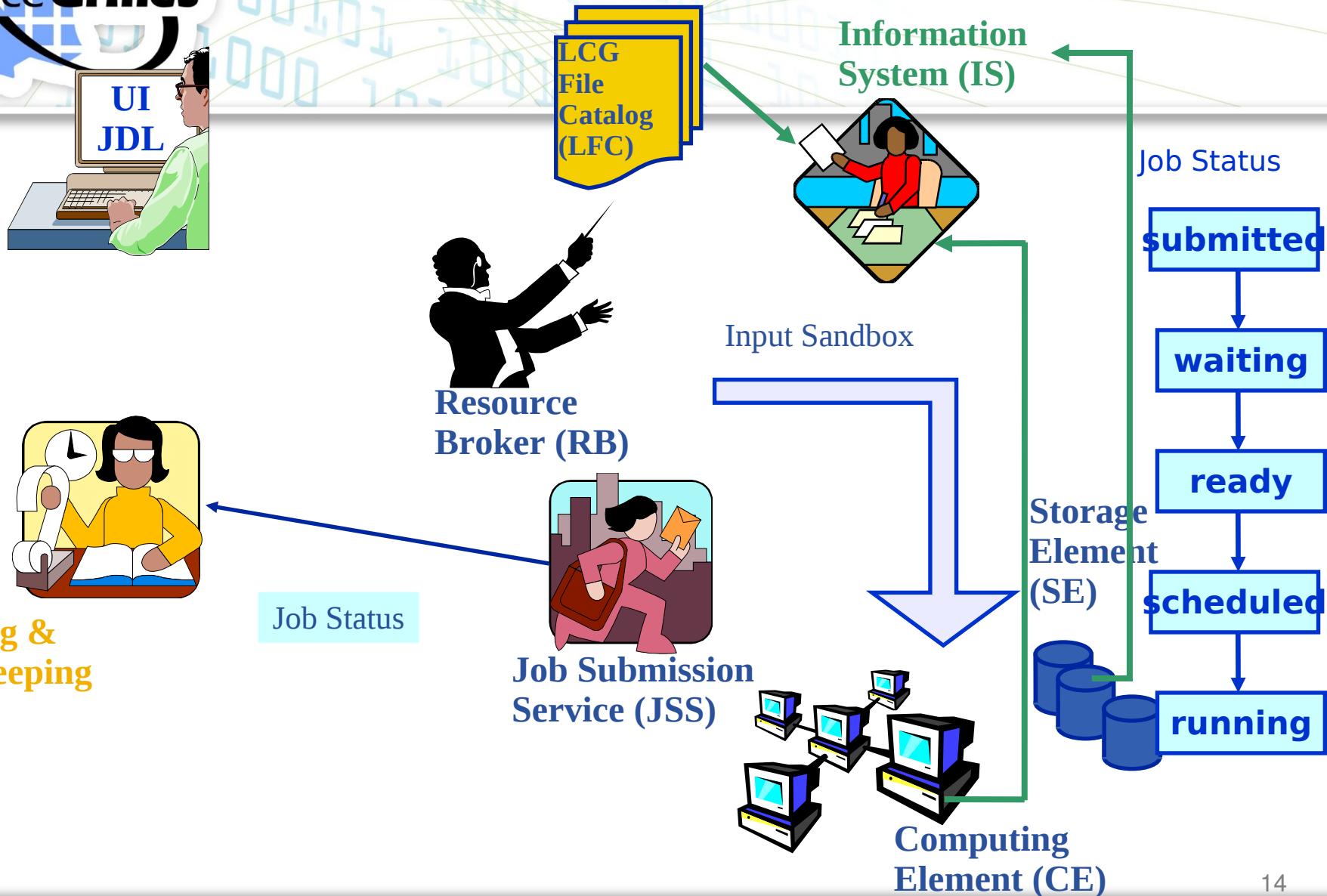
waiting

ready

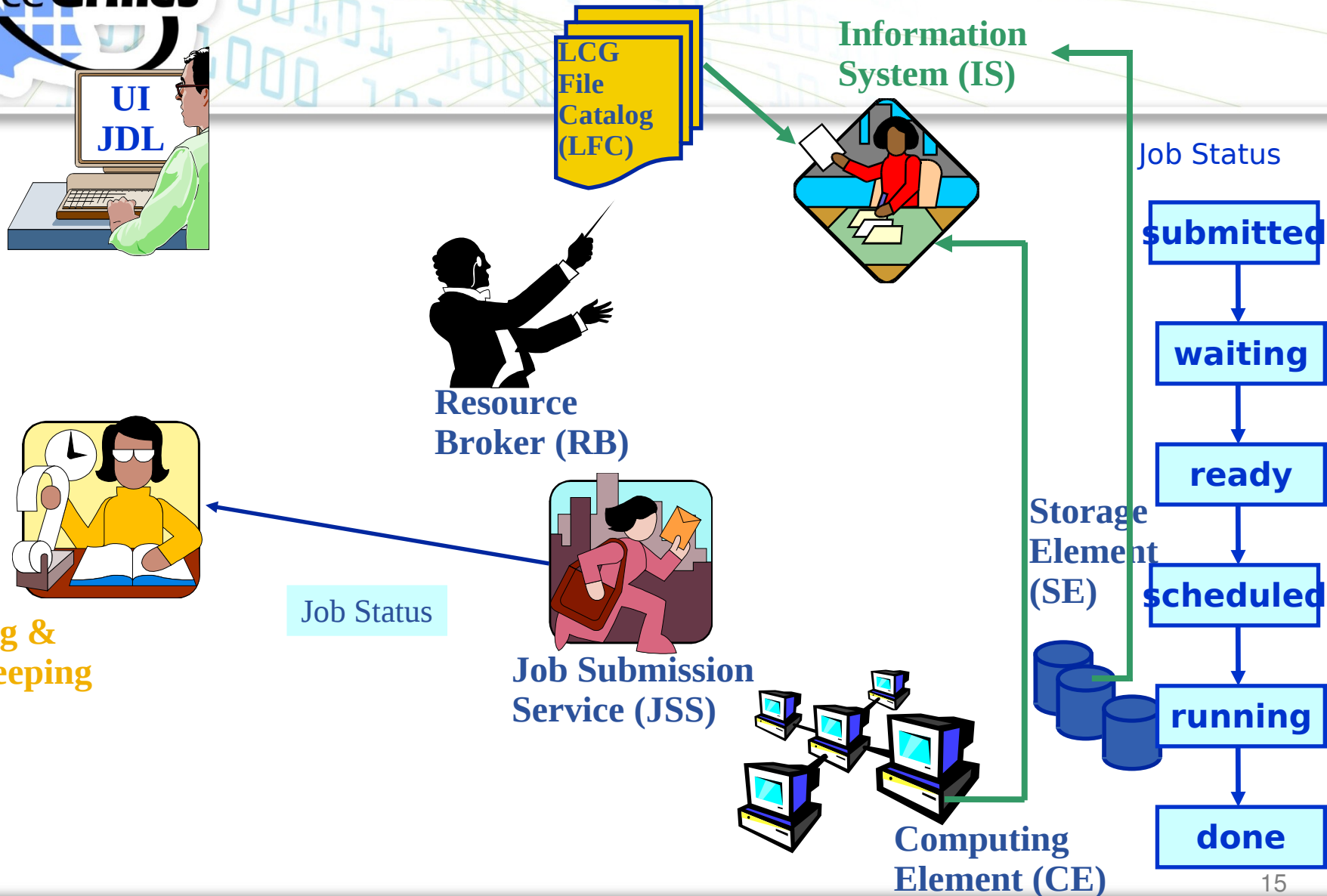
scheduled

Job Status

Soumission de jobs : scénario

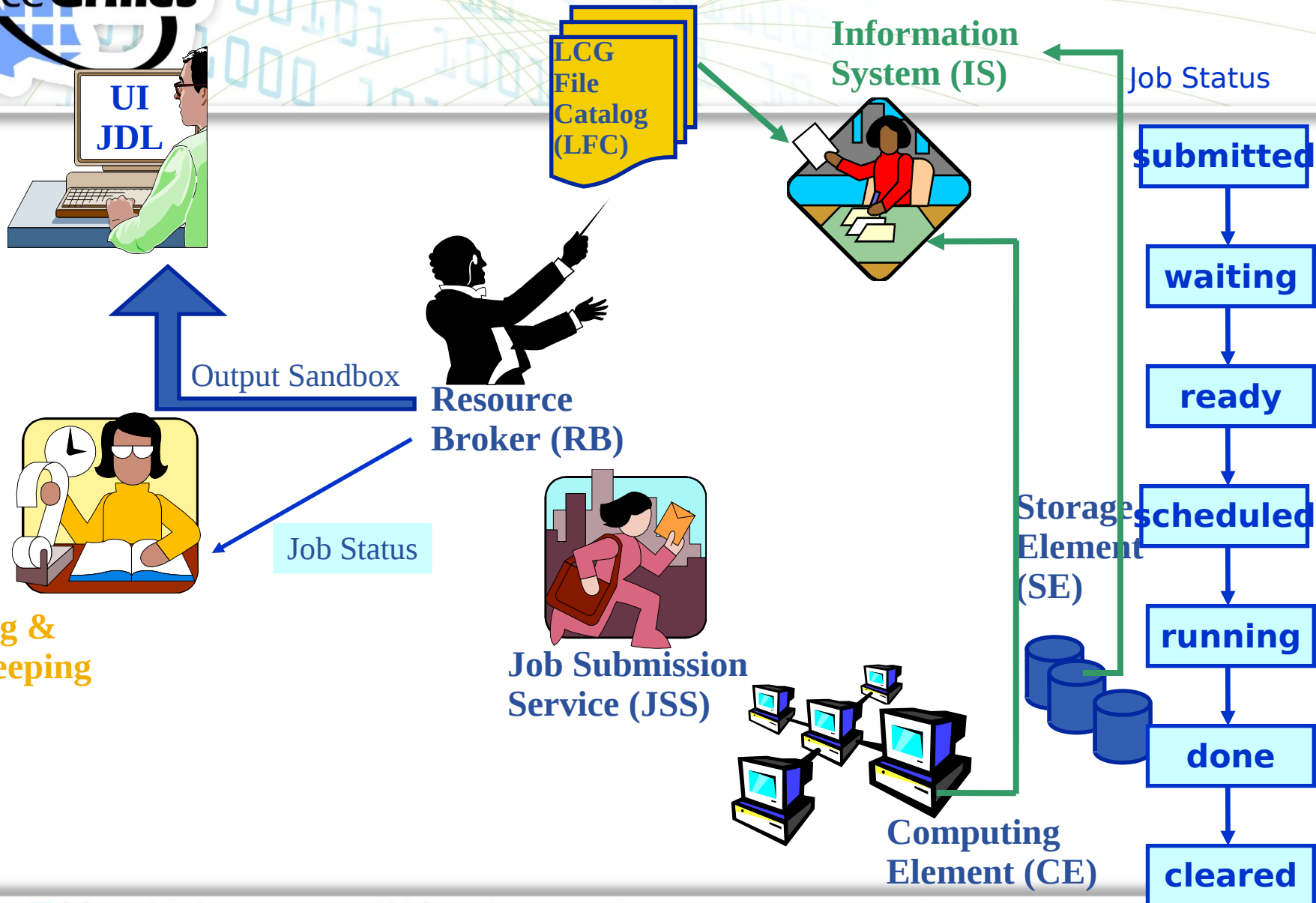


Soumission de jobs : scénario





Soumission de jobs : scénario





Plan : vous êtes ici

- Rappel nœuds de grille
- Soumission de job : *proxy* et scénario
- **JDL**
- **Commandes de soumission**
- *Job perusal*
- *VO software area*

➤ JDL : Job Description Language

- on spécifie (**minimum**) :
 - le programme et ses arguments
 - **redirection des outputs et erreurs dans des fichiers**
 - ce qu'on fait de la sortie (OutputSandbox)

➤ cat HelloWorld.jdl

Executable = **"/bin/echo "**;

Arguments = "Hello World ";

StdOutput = **"message.txt "**;

StdError = **"stderr "**;

OutputSandbox = {"message.txt", "stderr "};

JDL : types d'attribut

- **Les attributs supportés sont groupés en 2 catégories :**
 - Job
 - définit le job lui-même
 - Ressources
 - proviennent du système d'information, pris en compte par le WMS et utilisé par l'algorithme de correspondance (*matchmaking*)
 - ressources de calcul (Attributs)
 - Utilisé pour exprimer les attributs *Requirements* et/ou *Rank* par l'utilisateur
 - Doivent être préfixés par "other."
 - ressources de données et de stockage (Attributs), nécessitent l'interrogation des catalogues de fichiers.
 - Données en entrées utilisées, protocoles utilisés par les applications pour accéder aux SE

➤ **Arguments (*optionnel*)**

- arguments de la ligne de commande du job

➤ **StdInput (*optionnel*), StdOutput et StdError (*obligatoires*)**

- standard input/output/error du job

➤ **Environment (*optionnel*)**

- liste de variables d'environnement

➤ **InputSandbox (*optionnel*)**

- liste de fichiers sur le disque local de l'UI ou sur un serveur ([grid]FTP, http, ...) nécessaires lors de l'exécution du job
- les fichiers listés sont envoyés depuis l'UI sur le WN

➤ **OutputSandbox (*optionnel*)**

- liste des fichiers, générés par le job, qui seront récupérés
- ces fichiers sont envoyés depuis le RB sur l'UI

➤ **Requirements: besoin du job en ressources**

- spécifié en utilisant les attributs des ressources publiées dans le système d'information
- la valeur par défaut définie dans le fichier de configuration de l'UI est ajoutée (ET logique) :
 - par défaut : `other.GlueCEStateStatus == "Production"` (la ressource doit être dans la grille de production)

➤ **Rank: exprime la préférence (ordonner les ressources qui ont déjà rempli les conditions de l'attribut Requirements)**

- spécifié en utilisant les attributs des ressources publiées dans le système d'information
- si non spécifié, la valeur par défaut définie dans le fichier de configuration de l'UI est considérée :
 - par défaut : `-other.GlueCEStateFreeCPUs` (le plus grand nombre de CPU libres)



JDL : attributs pour les données

➤ **InputData** (*optionnel*)

- fait référence aux données utilisées en entrée d'un job : ces données sont publiées dans le catalogue LFC et stockées sur un SE
- PFN et/ou LFN

➤ **DataAccessProtocol** (*obligatoire si InputData spécifié*)

- le protocole ou la liste des protocoles avec lesquels l'application est susceptible d'accéder aux *InputData* sur un SE donné

➤ **OutputSE** (*optionnel – uniquement avec LCG-RB*)

- le hostname du SE sur lequel sera copié les *OutputData*
- le RB utilise cet attribut pour choisir un CE qui est compatible avec le job et proche du SE (notion de *closeSE*)

➤ **OutputData** (*optionnel – uniquement avec LCG-RB*)

- données en sortie qui seront enregistrées sur un SE à la fin du job

➤ **DataCatalog** (*optionnel, **recommandé***)

- le point d'accès du service DLI (Data Location Interface) du catalogue LFC a utilisé : <http://<hostname>:8085>

attribut job

```
Executable = "gridTest";  
StdError = "stderr.log";  
StdOutput = "stdout.log";  
InputSandbox = {"/home/joda/test/gridTest"};  
OutputSandbox = {"stderr.log", "stdout.log"};
```

attribut
données

```
InputData = "lfn:testbed0-00019";  
DataAccessProtocol = "gridftp";
```

attributs
ressources

```
Requirements = other.Architecture=="INTEL" && \  
                other.OpSys=="LINUX" && other.FreeCpus\  
                >=4;  
Rank = "other.GlueHostBenchmarkSF00";
```



Plan : vous êtes ici

- Rappel nœuds de grille
- Soumission de job : proxy et scénario
- JDL
- **Commandes de soumission**
- *Job perusal*
- *VO software area*



Commandes sur une UI

(glite-WMS / **lcg-RB**)

- **glite-wms-job-submit -a / **edg-job-submit****
 - Soumets un job
 - Retourne le jobID
- **glite-wms-job-list-match -a / **edg-job-list-match****
 - Liste les ressources compatibles avec la description du job
 - Effectue la correspondance (*matchmaking*) sans soumettre le job
- **glite-wms-job-cancel / **edg-job-cancel****
 - Annule un job
- **glite-wms-job-status / **edg-job-status****
 - Donne le statut du job
- **glite-wms-job-output / **edg-job-get-output****
 - Récupère les fichiers spécifiés dans l'attribut OutputSandbox en local sur l'UI
- **glite-wms-job-logging-info / **edg-job-get-logging-info****
 - Donne des informations de *logging* sur les jobs soumis (tous les événements répertoriés par les divers composants du WMS) - Très utile pour déboguer

```
$ glite-wms-job-submit -a --vo gilda helloworld.jdl
```

```
Selected Virtual Organisation name (from --vo option): gilda
Connecting to host grid004.ct.infn.it, port 7772
Logging to host grid004.ct.infn.it, port 9002
```

```
*****
```

JOB SUBMIT OUTCOME

```
The job has been successfully submitted to the Network Server.
Use edg-job-status command to check job current status. Your job
identifier (edg_jobId) is:
```

```
- https://grid004.ct.infn.it:9000/PKw6dRR-0ziUf8r217TZoA
```

```
*****
```



Soumission de jobs (ex.) : statut

```
$ glite-wms-job-status
```

```
https://grid004.ct.infn.it:9000/PKw6dRR-0ziUf8r217TZoA
```

```
*****
```

```
BOOKKEEPING INFORMATION:
```

```
Status info for the Job :
```

```
https://grid004.ct.infn.it:9000/PKw6dRR-0ziUf8r217TZoA
```

```
Current Status:      Scheduled
```

```
Status Reason:       Job successfully submitted to Globus
```

```
Destination:         grid006.cecalc.ula.ve:2119/jobmanager-lcgpbs-  
long
```

```
reached on:          Fri Sep  2 08:21:16 2005
```

```
*****
```




Soumission de jobs (ex.) : output

```
$ glite-wms-job-output --dir resultats  
https://lxn1177.cern.ch:9000/j7BaJWDA11AYYGYvbRRlUw
```

```
Retrieving files from host: lxn1177.cern.ch ( for  
https://lxn1177.cern.ch:9000/j7BaJWDA11AYYGYvbRRlUw )
```

```
*****
```

```
JOB GET OUTPUT OUTCOME
```

```
Output sandbox files for the job:  
- https://lxn1177.cern.ch:9000/j7BaJWDA11AYYGYvbRRlUw  
have been successfully retrieved and stored in the directory:  
/home/manu/resultats/manu_j7BaJWDA11AYYGYvbRRlUw
```

```
*****
```

- **L'option --dir est optionnelle : l'UI est configurée pour rediriger les fichiers d'output vers un répertoire par défaut.**

```
$ cat ~/resultats/manu_j7BaJWDA11AYYGYvbRRlUw/std.*
```

```
...
```



Soumission de jobs (ex.) : stockage des JobID

```
$ glite-wms-job-submit -a -o jobsid --vo gilda helloworld.jdl
$ glite-wms-job-submit -a -o jobsid --vo gilda helloworld.jdl
$ glite-wms-job-submit -a -o jobsid --vo gilda helloworld.jdl
$ glite-wms-job-submit -a -o jobsid --vo gilda helloworld.jdl
$ glite-wms-job-submit -a -o jobsid --vo gilda helloworld.jdl
```

```
$ glite-wms-job-status -i jobsid
```

```
-----
1 : https://grid004.ct.infn.it:9000/UcDXhD6z3yRGzBQt1k_Z6Q
2 : https://grid004.ct.infn.it:9000/-mfCNPcCcpCf5u0e3D6JkQ
3 : https://grid004.ct.infn.it:9000/D24Fo3VbfHzpHFXau2WZeg
4 : https://grid004.ct.infn.it:9000/2SPkbdH0D8j2faVBXzU3qQ
5 : https://grid004.ct.infn.it:9000/WwPvzNZAyDd1HhnJkvBGgQ
a : all
q : quit
-----
```

➤ Soumission directe à un CE (option -r) :

```
$ glite-wms-job-submit -a --vo gilda -r gilda-ce- \
01.pd.infn.it:2119/jobmanager-lcgpbs-infinite \
helloworld.jdl
```

➤ \$ cat hostnamerank.jdl

```
Type = "Job";
JobType = "Normal";
Executable = "/bin/hostname";
Arguments = "-f";
StdOutput = "hostname.out";
StdError = "hostname.err";
OutputSandbox = {"hostname.err", "hostname.out"};
RetryCount = 7;
Rank=(other.GlueCEStateFreeCPUs == 0 ? - \
    other.GlueCEStateWaitingJobs : other.GlueCEStateFreeCPUs);
Requirements = (other.GlueCEPolicyMaxCPUTime<=3600) && (RegExp \
    ("infn", other.GlueCEUniqueId));
```

- 1 CPU libre et job de plus de 2 heures :

```
Requirements = other.GlueCEInfoTotalCPUs > 1 \ && other.GLUECEPolicyMaxCPUTime > 120;
```

- On peut spécifier un CE particulier avec le JDL :

```
Requirements = other.GlueCEUniqueID == \ "lxshare0286.cern.ch:2119/jobmanager-pbs- \ short";
```

- **Le glite-WMS (ou le lcg-RB) est le composant principal du WMS.**
 - Son rôle est de trouver la meilleure ressource (CE) possible où le job pourra être exécuté
 - Il interagit avec le service de gestion des données et le système d'information
 - ils fournissent au WMS toutes les informations requises pour établir la correspondance
- **Le CE choisi par le WMS doit remplir les conditions du job**
 - Si 2 CE ou plus satisfont toutes ces requêtes, celui qui a le meilleur rang est choisi



Soumission de jobs (ex.) : ressources disponibles

```
$ glite-wms-job-list-match -a --vo gilda helloworld.jdl
```

```
Selected Virtual Organisation name (from --vo option): gilda  
Connecting to host grid004.ct.infn.it, port 7772
```

```
*****
```

COMPUTING ELEMENT IDs LIST

The following CE(s) matching your job requirements have been found:

CEId

```
ced-ce0.datagrid.cnr.it:2119/jobmanager-lcgpbs-infinite  
ced-ce0.datagrid.cnr.it:2119/jobmanager-lcgpbs-long  
ced-ce0.datagrid.cnr.it:2119/jobmanager-lcgpbs-short  
gilda-ce-01.pd.infn.it:2119/jobmanager-lcgpbs-short  
grid-ce.bio.dist.unige.it:2119/jobmanager-lcgpbs-infinite  
grid-ce.bio.dist.unige.it:2119/jobmanager-lcgpbs-long  
grid-ce.bio.dist.unige.it:2119/jobmanager-lcgpbs-short
```

```
...
```



Soumission de jobs (ex.) : info sur les ressources

```
$ lcg-infosites --vo gilda ce
```

```
*****
These are the related data for gilda: (in terms of queues and
CPUs)
*****
```

#CPU	Free	Total	Jobs	Running	Waiting	Computing	Element
36	36	0		0	0		
grid010.ct.infn.it:2119/jobmanager-lcgpbs-long							
14	14	0		0	0		
grid011f.cnafr.infn.it:2119/jobmanager-lcgpbs-long							
6	6	0		0	0		
ced-ce0.datagrid.cnr.it:2119/jobmanager-lcgpbs-long							
...							

- **But : examiner les fichiers produits pendant un job**
 - peut s'appliquer à tout fichier
 - requiert 2 lignes supplémentaires dans le JDL :
 - `PerusalFileEnable = true;`
 - `PerusalTimeInterval = 120; # In seconds, not too low`
- **Définition et récupération des fichiers à examiner : `glite-wms-job-perusal [--set|--get|--unset] -f file jobid`**
 - `--set` définit les fichiers à examiner
 - `--get` récupère la différence avec la version précédente
 - `--all` force la récupération de tous les fichiers
 - `--nodisplay` stocke le fichier plutôt que de l'afficher
 - `--unset` : annule l'examen (la récupération périodique) du fichier
- **A utiliser avec modération : peut avoir un impact important sur les performances du WMS**

- **Chaque VO dispose d'un espace spécifique pour installer ses applications sur un CE**
 - espace partagé par les WNs
 - référencé par variable d'environnement: `VO_VONAME_SW_DIR`
 - VONAME est le nom de la VO avec les '.' et '-' remplacés par des '_'
- **Droit d'écriture restreint au seul VO Software Manager**
 - accessible en lecture à tout le monde [(toutes les VOs)]
 - Software Manager défini avec un rôle VOMS (au choix de la VO)
- **Mise à jour de la SW area effectuée en soumettant des jobs avec le rôle Software Manager**
- **Contenu de la SW area peut être publié en définissant des tags depuis 1 UI ou 1 WN (job)**
 - `lcg-ManageVOTag -host CE -vo voname ...`



Merci

➤ Liens utiles:

- gLite User Guide : <https://edms.cern.ch/document/722398/>
- WMPProxy (version actuelle du WMS) user guide
<https://edms.cern.ch/document/674643>
- JDL attributes <https://edms.cern.ch/document/590869>