

Soumission de jobs de calcul

David Bouvet

IN2P3-CC

Observatoire de Paris, Meudon - 10/12/2008

- **Rappel nœuds de la grille**
- **Soumission de job : *proxy* et scénario**
- **JDL**
- **Commandes de soumission**

- **UI (*User Interface*)** : point d'accès à la grille WLCG/EGEE
 - 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*)** : nœuds 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

- 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 est composé d'un *RB (Resource Broker)* : le courtier de ressources de la grille, responsable de trouver les « meilleures » ressources pour soumettre les jobs.
- Le WMS interagit avec les noeuds suivants :
 - *UI (User Interface)* : point d'accès pour les 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
 - *LB (Logging and Bookkeeping)* : stocke les infos concernant le job pour des requêtes utilisateurs.

- **Transition entre 2 générations de RB/WMS**
 - LCG RB: le plus déployé actuellement, peu de fonctionnalités avancées, performances limitées
 - Commandes : edg-job-xxx
 - gLite WMS : déploiement en cours
 - Commandes : glite-wms-xxx
 - Proxy delegation (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

- Rappel nœuds de grille
- **Soumission de job : *proxy* et scénario**
- **JDL**
- **Commandes de soumission**

- **voms-proxy-init -voms vo.rocfr.in2p3.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.rocfr.in2p3.fr" Done
Creating proxy ..... Done
Your proxy is valid until Sat Nov 4 02:56:14 2006
```

- **voms-proxy-info**

```
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     : 11:58:53
```

- **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 egeode extension information ===

```
V0           : vo.rocfr.in2p3.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    : /egeode/Role=NULL/Capability=NULL
timeleft     : 11:58:25
```

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

Certificate:

Data:

Version: 3 (0x2)

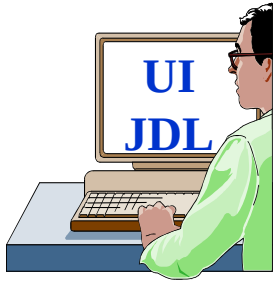
Serial Number: 2239 (0x8bf)

Signature Algorithm: md5WithRSAEncryption

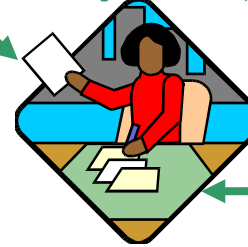
Issuer: C=IT, O=GILDA, OU=Personal Certificate,
L=CLERMONT-FERRAND, CN=CLERMONT-FERRAND01/Email=emmanuel
.medernach@clermont.in2p3.fr

Validity...

Soumission de jobs : scénario



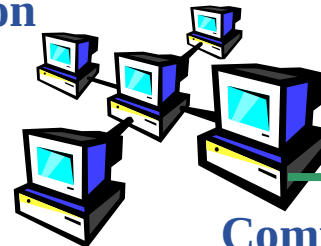
Information System (IS)



Resource Broker (RB)

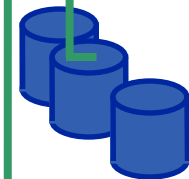


Job Submission Service (JSS)



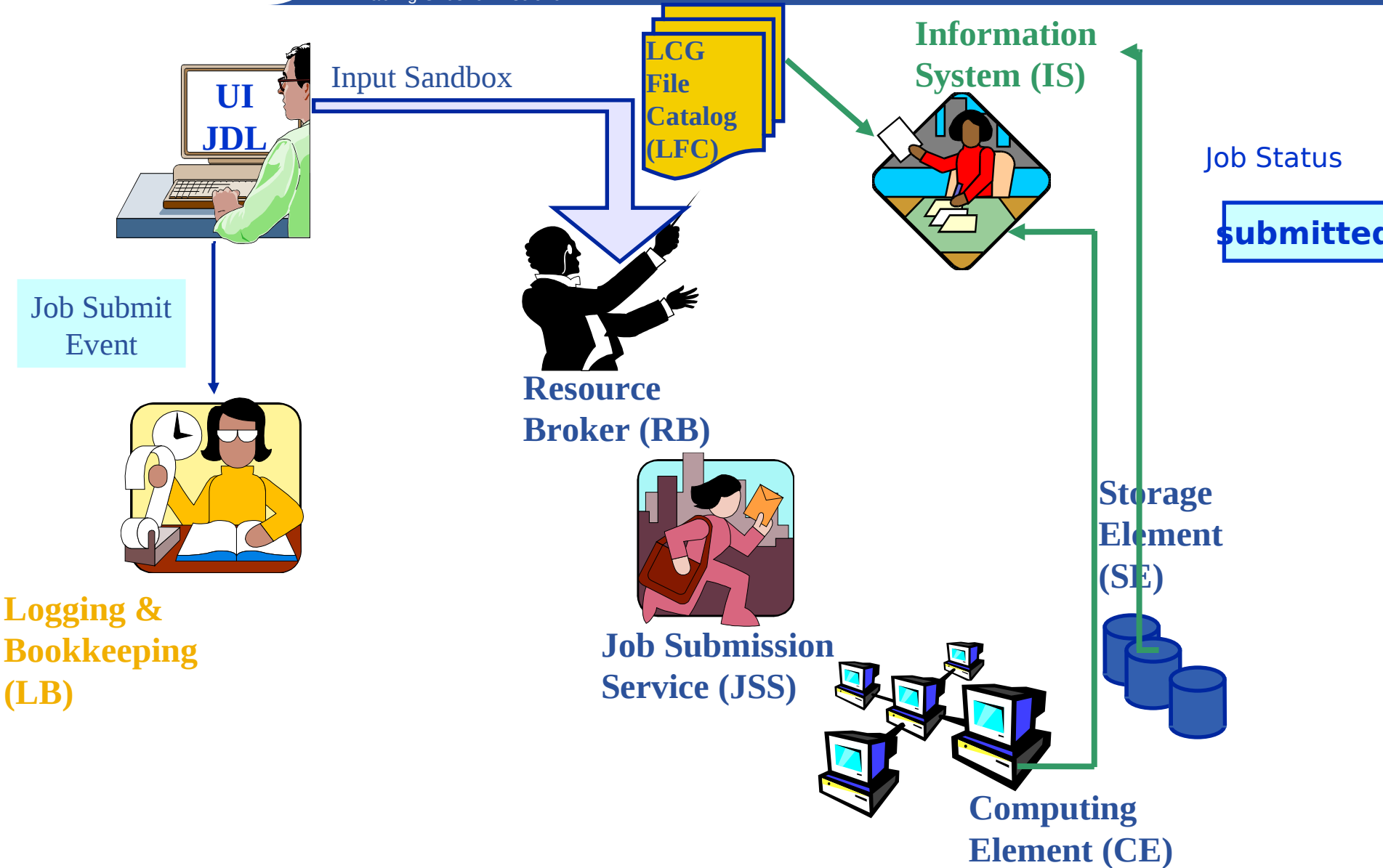
Computing Element (CE)

Storage Element (SE)

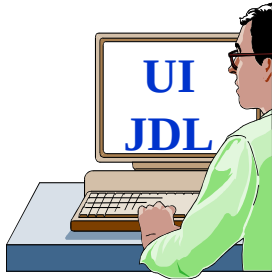


Logging & Bookkeeping (LB)

Soumission de jobs : scénario



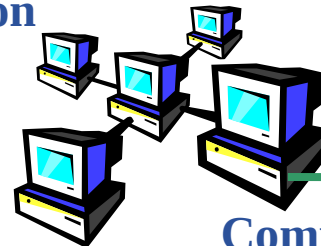
Soumission de jobs : scénario



Resource Broker (RB)

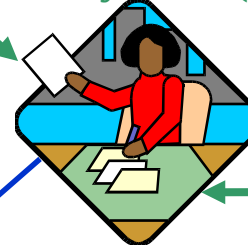


Job Submission Service (JSS)

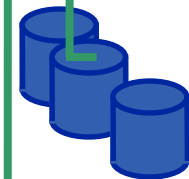


Computing Element (CE)

Information System (IS)



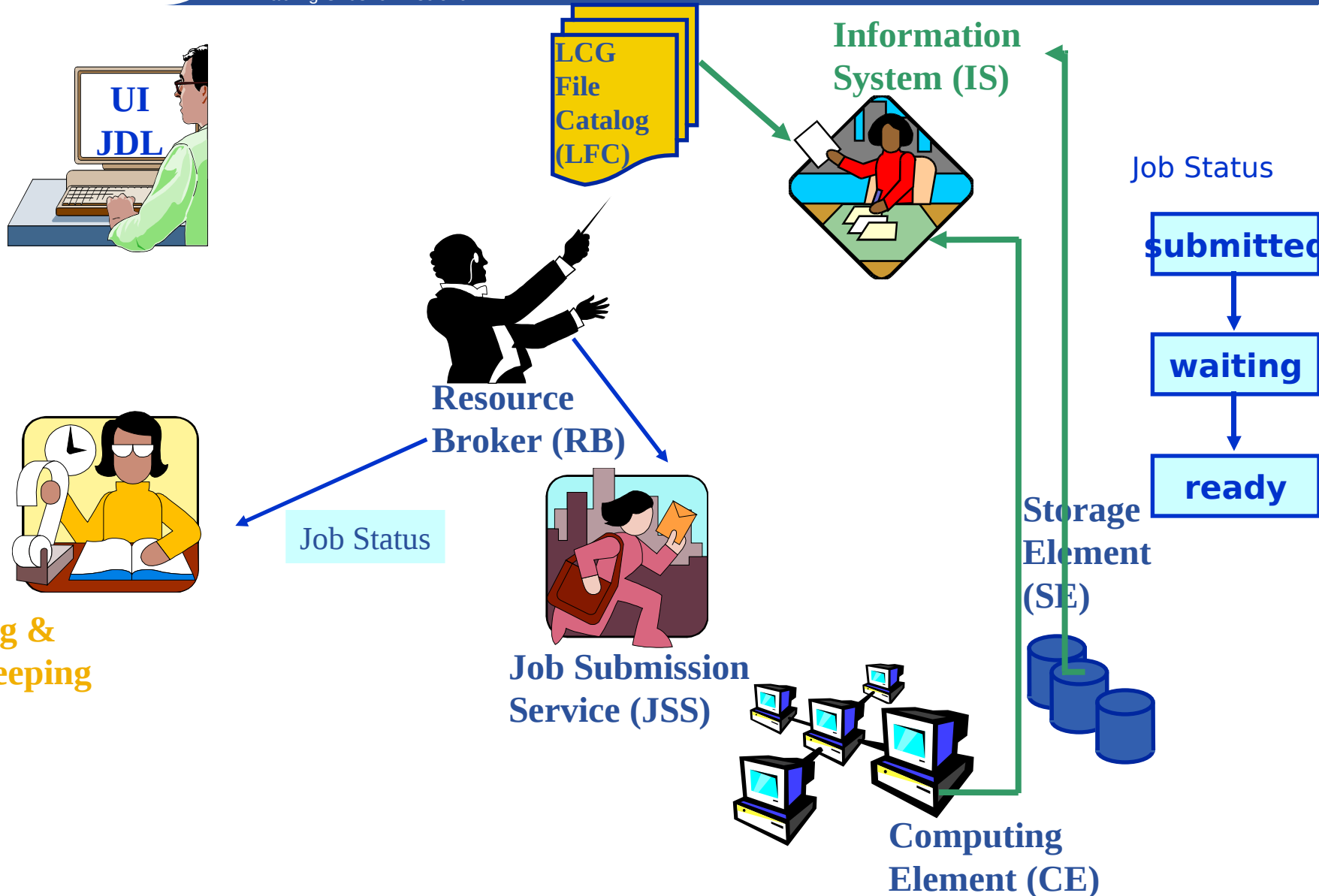
Storage Element (SE)



Job Status

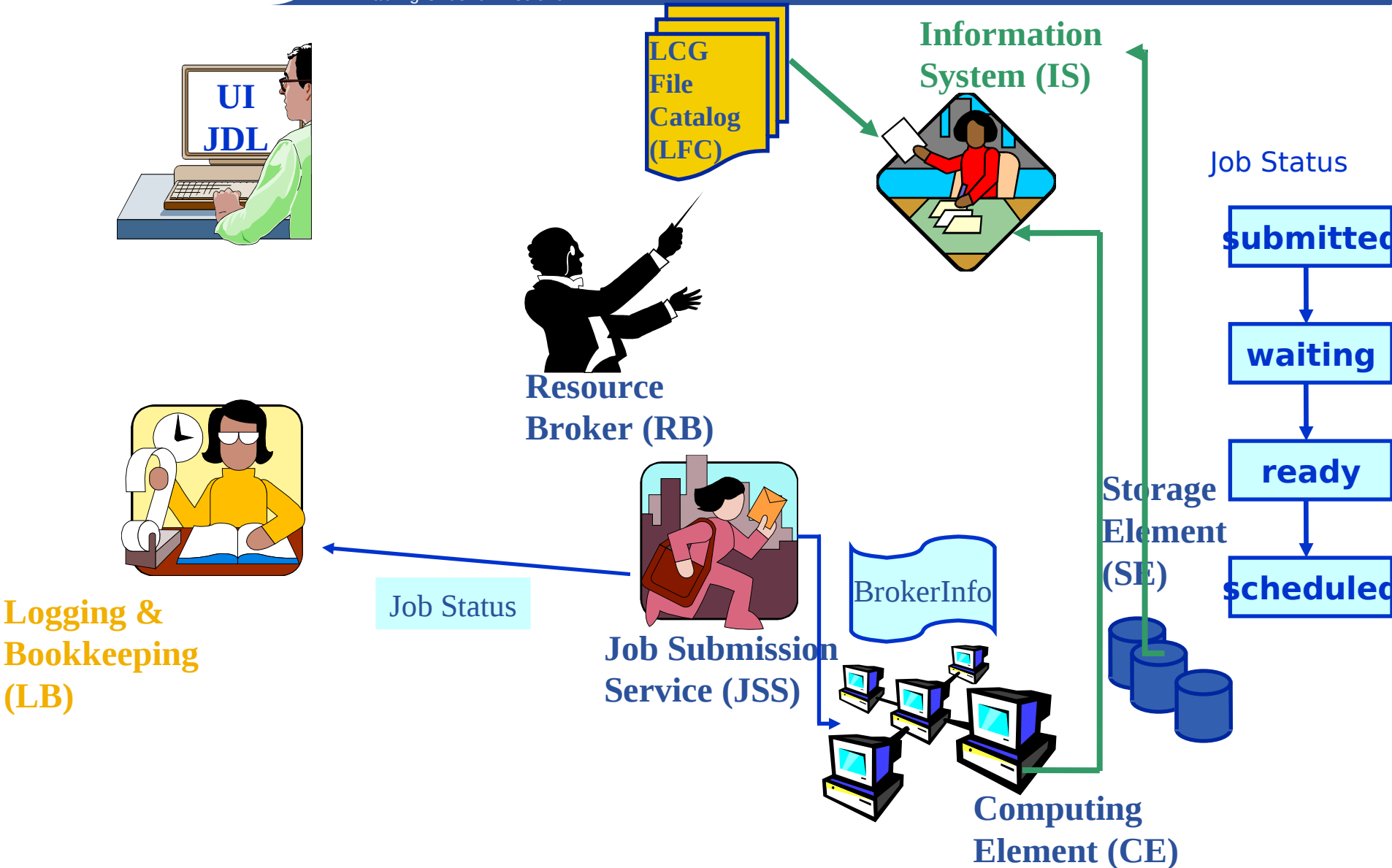


Logging & Bookkeeping (LB)

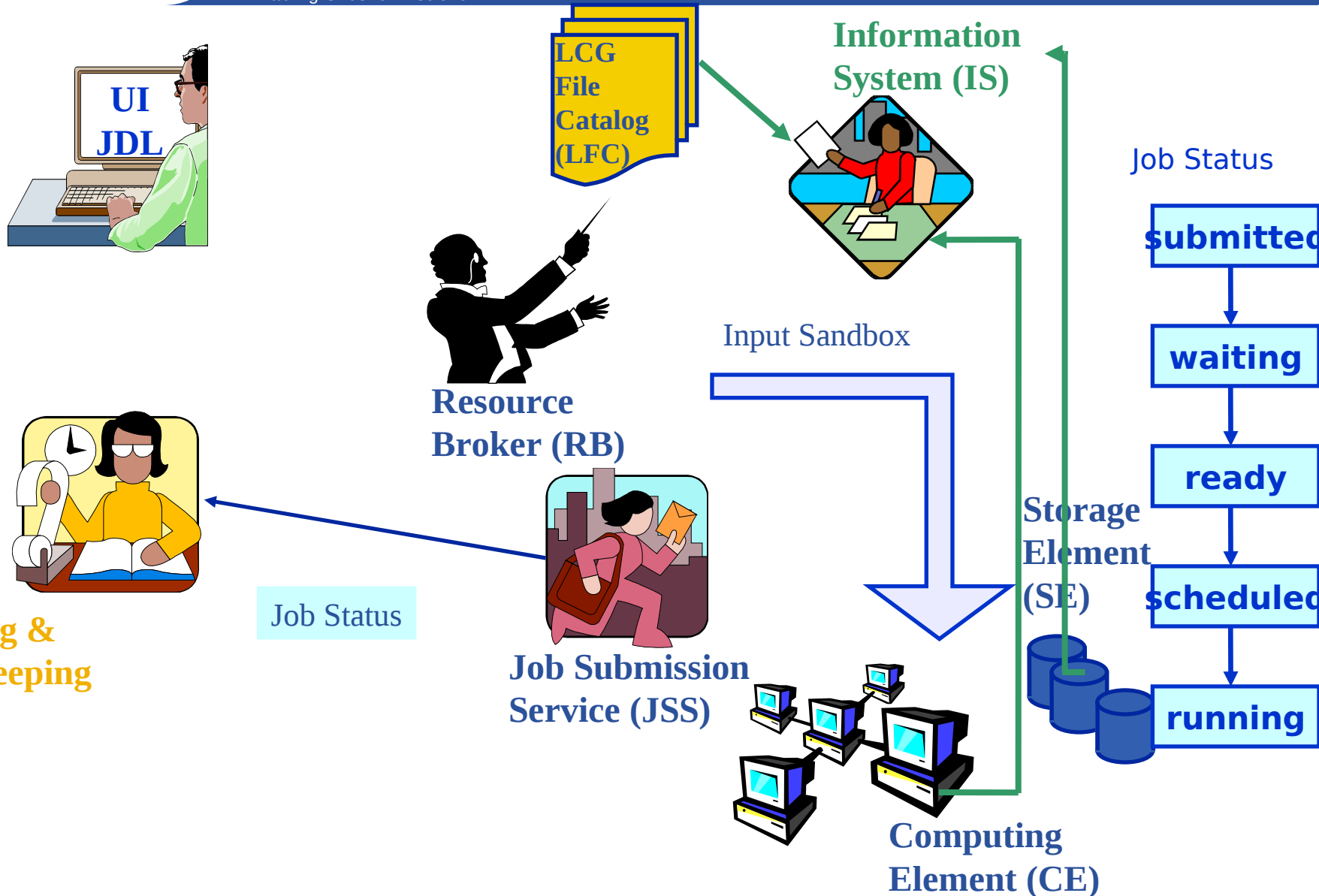


Logging & Bookkeeping (LB)

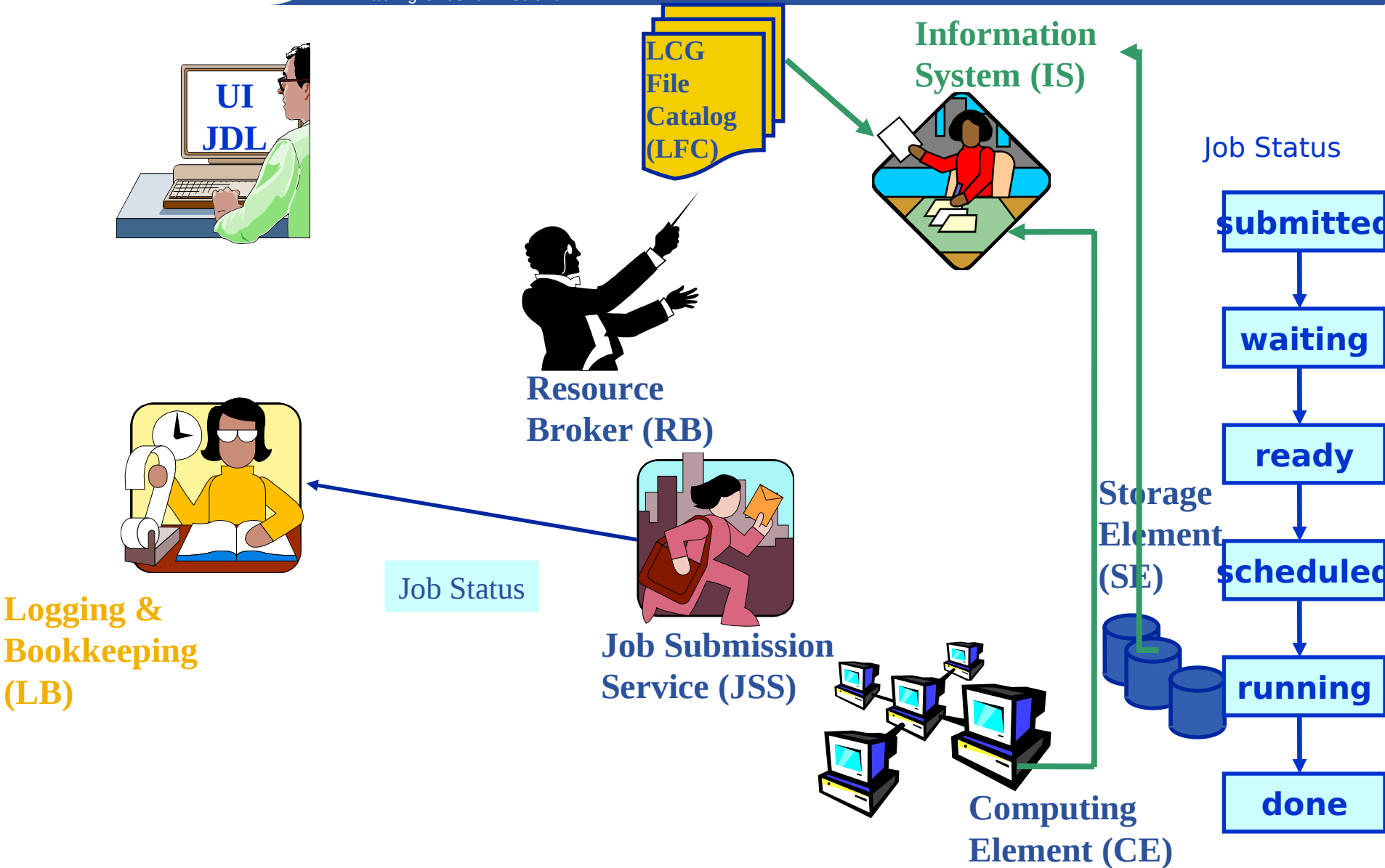
Soumission de jobs : scénario



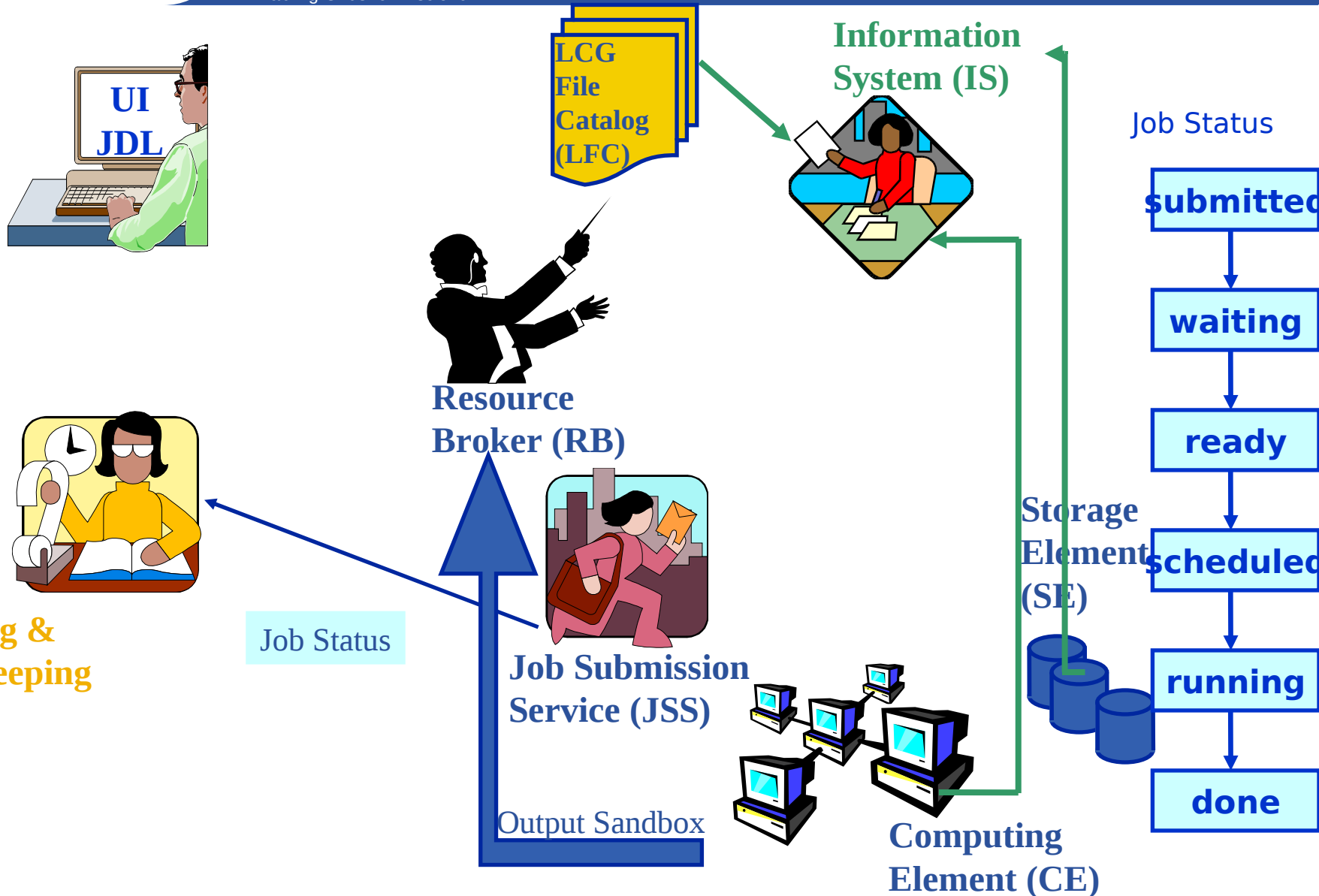
Soumission de jobs : scénario



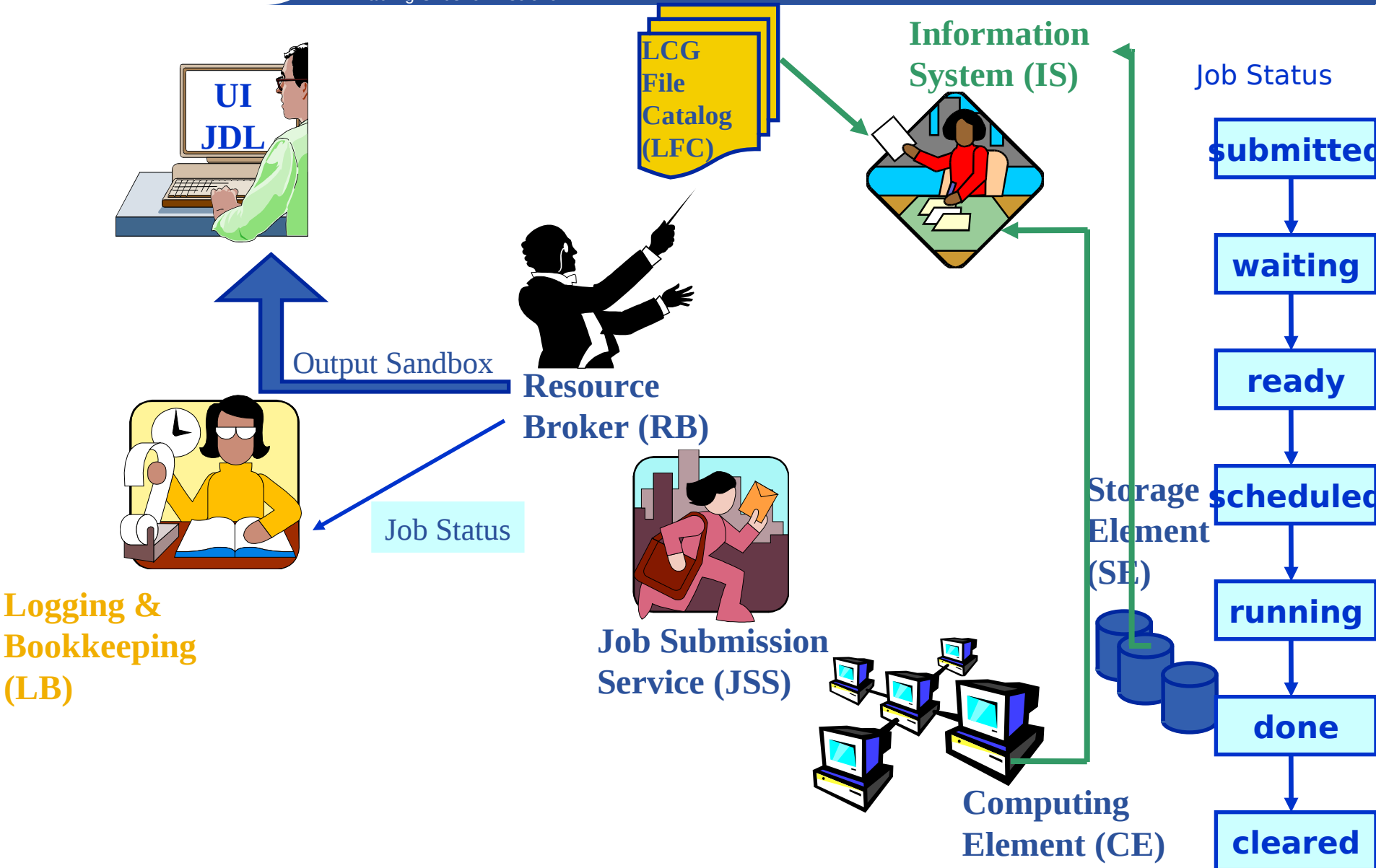
Soumission de jobs : scénario



Logging & Bookkeeping (LB)



Soumission de jobs : scénario



- Rappel nœuds de grille
- Soumission de job : *proxy* et scénario
- **JDL**
- **Commandes de soumission**

- **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 l'output (OutputSandbox)

- **cat HelloWorld.jdl**

```
Executable = "/bin/echo ";  
Arguments = "Hello World ";  
StdOutput = "message.txt ";  
StdError = "stderr ";  
OutputSandbox = {"message.txt", "stderr"};
```

- **Les attributs supportés sont groupés en 2 catégories :**
 - *Job*
 - défini le job lui-même
 - *Ressources*
 - pris en compte par le RB et utilisé par l'algorithme de *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)
 - Données en entrées utilisées, SE où stocker les données en sortie, protocoles utilisés par les applications pour accéder aux SE*

- Executable (***obligatoire***)
 - le nom de la commande
- Arguments (***optionnel***)
 - arguments de la ligne de commande du job
- StdInput, StdOutput, StdErr (***optionnel***)
 - 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 gridFTP nécessaires lors de l'exécution du job
 - les fichiers listés sont envoyés depuis l'UI sur le CE
- 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 ressource de calcul
 - 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.GlueCEStateStatus == "Production"` (la ressource doit être dans la grille de production)
- Rank
 - exprime la préférence (comment ranger les ressources qui ont déjà rempli les conditions de l'attributs 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)

- **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*)**
 - le *hostname* du catalogue LFC a utilisé

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";
```


- Rappel nœuds de grille
- Soumission de job : *proxy* et scénario
- JDL
- **Commandes de soumission**

- `edg-job-submit glite-wms-job-submit -a`
Soumets un job
Retourne le jobID
- `edg-job-list-match glite-wms-job-list-match -a`
Liste les ressources compatibles avec la description du job
Effectue le *matchmaking* sans soumettre le job
- `edg-job-cancel glite-wms-job-cancel`
Annule un job
- `edg-job-status glite-wms-job-status`
Donne le statut du job
- `edg-job-get-output glite-wms-job-output`
Récupère les fichiers spécifiés dans l'attribut *OutputSandbox* en local sur l'UI
- `edg-job-get-logging-info glite-wms-job-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

```
$ edg-job-submit --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
```

```
*****
```

```
$ edg-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
```

```
*****
```

```
$ edg-job-get-output --dir resultats https
://lxn1177.cern.ch:9000/j7BaJWDA11AYYGYvbRR1Uw
```

```
Retrieving files from host: lxn1177.cern.ch ( for
https://lxn1177.cern.ch:9000/j7BaJWDA11AYYGYvbRR1Uw )
```

```
*****
```

```
JOB GET OUTPUT OUTCOME
```

```
Output sandbox files for the job:
```

```
- https://lxn1177.cern.ch:9000/j7BaJWDA11AYYGYvbRR1Uw
have been successfully retrieved and stored in the directory:
/home/manu/resultats/manu_j7BaJWDA11AYYGYvbRR1Uw
```

```
*****
```

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

```
$ cat ~/resultats/manu_j7BaJWDA11AYYGYvbRR1Uw/std.*
```

```
...
```

```
$ edg-job-submit -o jobsid --vo gilda helloworld.jdl
$ edg-job-submit -o jobsid --vo gilda helloworld.jdl
$ edg-job-submit -o jobsid --vo gilda helloworld.jdl
$ edg-job-submit -o jobsid --vo gilda helloworld.jdl
$ edg-job-submit -o jobsid --vo gilda helloworld.jdl

$ edg-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) :**

```
$ edg-job-submit --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 Icg-RB ou le glite-WMS 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 RB/WMS toutes les informations requises pour établir la correspondance
- **Le CE choisi par le RB/WMS doit remplir les conditions du job**
- **Si 2 CE ou plus satisfont toutes ces requêtes, celui qui a le meilleur rang est choisi**

```
$ edg-job-list-match --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
```

```
...
```

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

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

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

- **But : permettre d'examiner les fichiers produits par un job pendant son exécution**
 - 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 performance du WMS**

- **Chaque VO dispose d'un espace spécifique pour installer ses applications sur un CE**
 - espace partagé par le CE et ses WNs
 - référencé à travers une variable d'environnement :

`VO_VONAME_SW_DIR`
 - VONAME est le nom de la VO avec les '.' remplacés par des '_'
- **Droit d'écriture restreint au seul VO Software Manager**
 - accessible en lecture à tout le monde (toutes les VOs)
 - Software Manager définit 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 ...
```

