Formation Utilisateur grille de calcul : FU

United Nations
Educational, Scientific and
Cultural Organization

FU5 : 20 et 21 Juin 2013
CNRST - Rabat

Plan de Formation

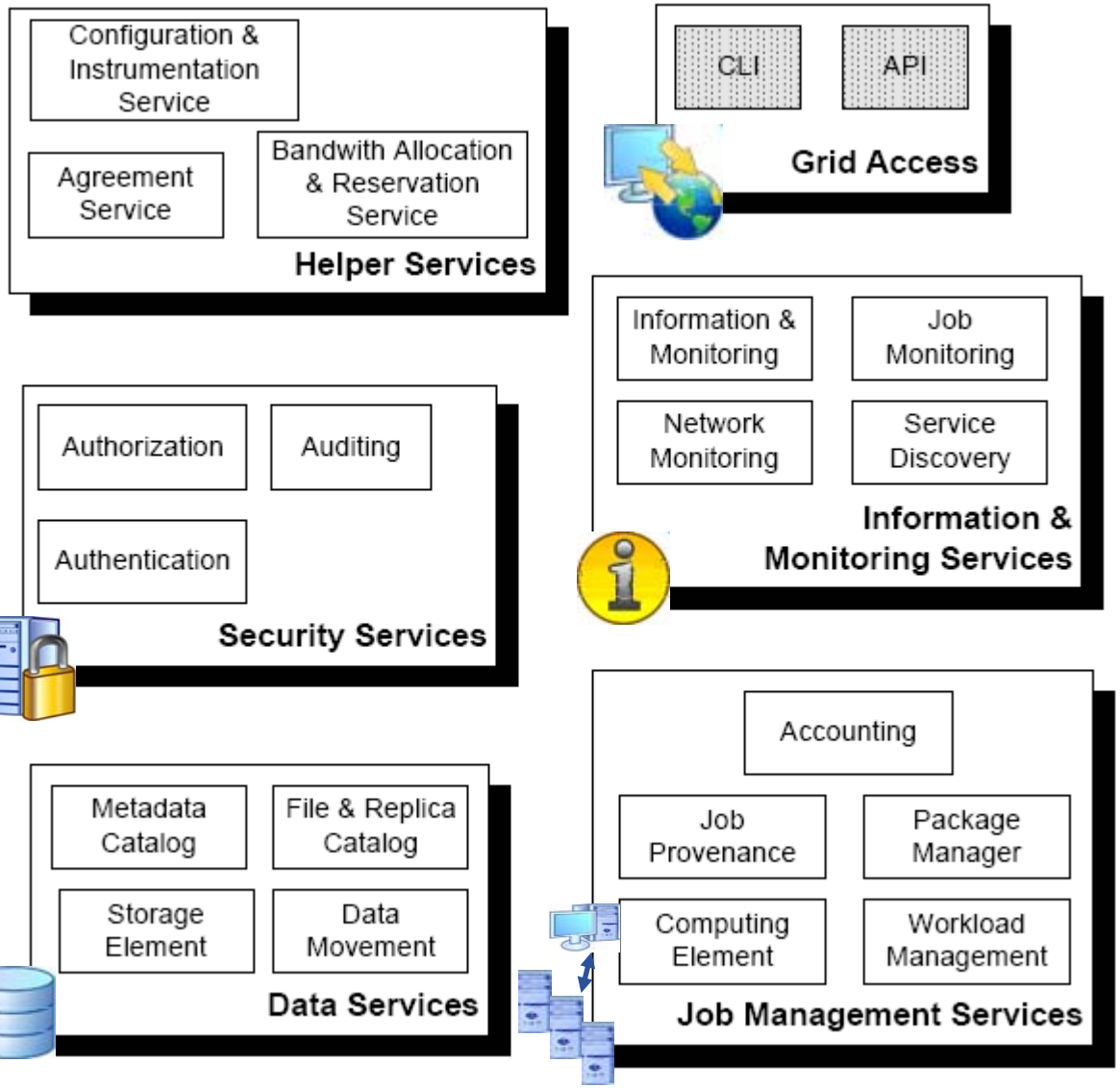
- Introduction aux grilles de calcul
- Projet MaGrid & UNESCO-HP
- Services de base de gLite
- Gestion des jobs sur la grille

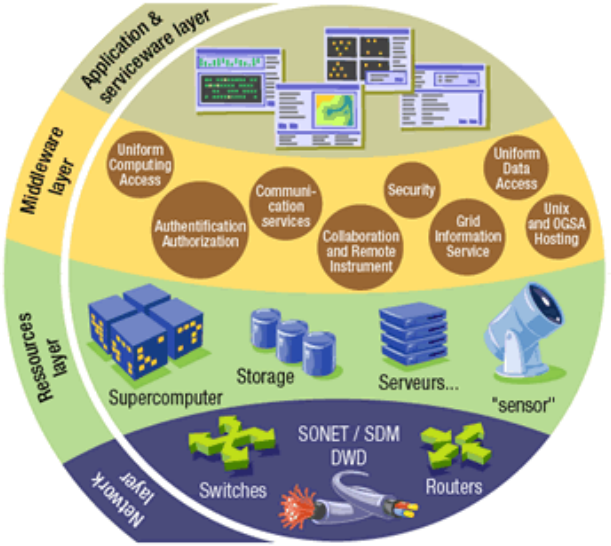
Nabil Talhaoui
grid@magrid.ma
Division TIC – CNRST, Rabat

Services de base de gLite

Services de base de gLite

Vue générale





Services de base de gLite

Accès à la Grille

Que faut-il pour travailler sur une Grille de calcul?

Un **certificat électronique personnel** délivré par une **Autorité de Certification (CA)** [Authentification]

Une adhésion à une **Organisation Virtuelle (VO)** [Autorisation]

Un compte sur une **Interface Utilisateur (UI)** ou sur un **Service Web** (portail) [Accès aux services]

Services de base de gLite

Accès à la Grille

Authentification/Autorisation

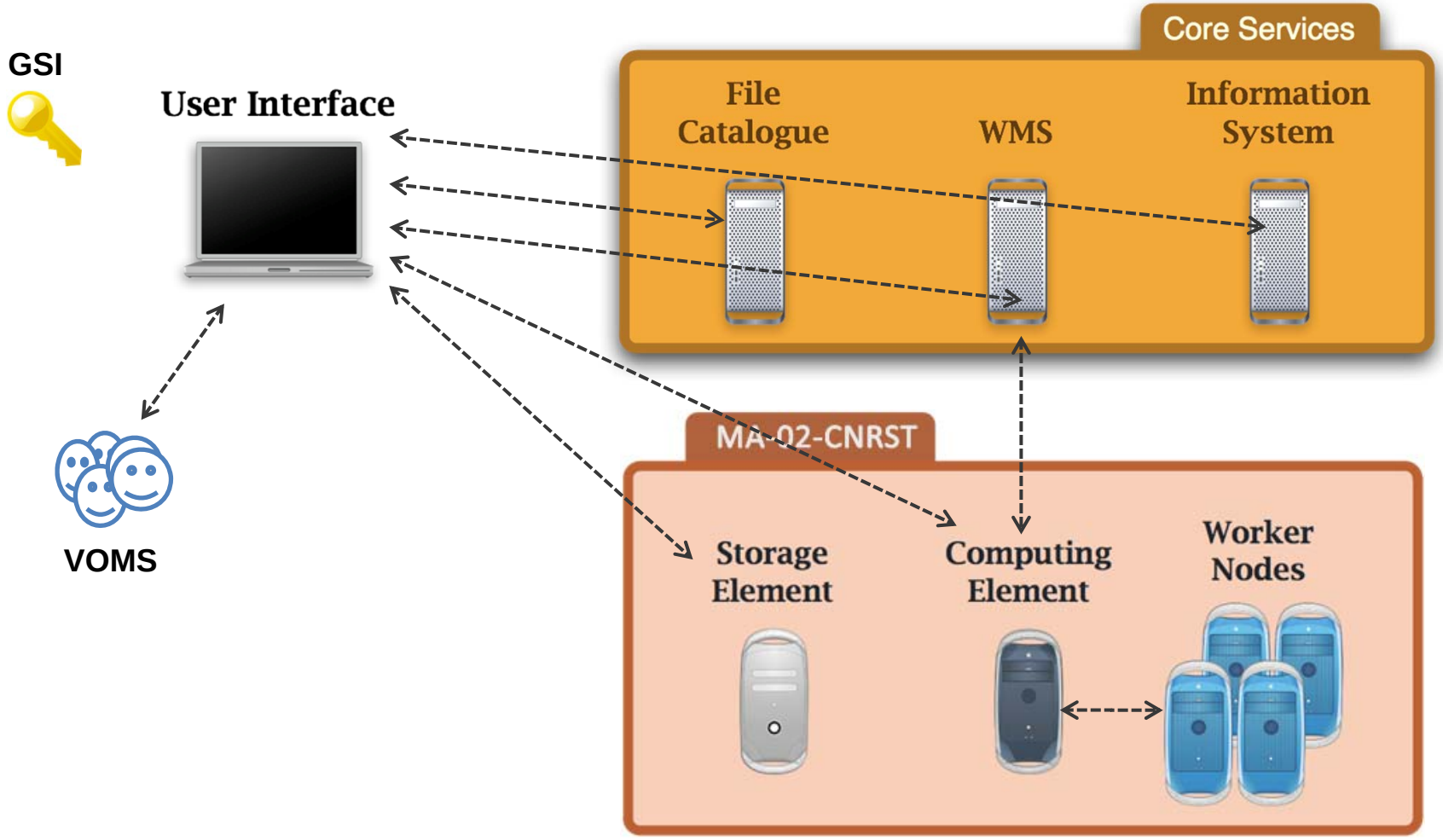
- La sécurité sur la Grille **(GSI)** est basée sur l'utilisation des certificats **X509** et la **PKI**.
- Chaque **utilisateur**, **serveur** ou **service** est identifié au moyen d'un certificat numérique (X509) certifiant son identité (**Authentification**) délivré par une **Autorité de Certification**.
- Les ressources sont en possession des VOs
- Chaque VO associe des droits d'accès aux ressources selon le «group» et le «role» de l'utilisateur **(Autorisation)**

- Le **VO Membership System (VOMS)** est le service qui traque l'adhésion des utilisateurs aux VOs. C'est le service de base de l'autorisation
- L'accès aux ressources se déroule en toute sécurité (intégrité, confidentialité), en utilisant une granularité qui peut aller jusqu'au niveau d'un seul utilisateur.

Services de base de gLite

Accès à la Grille

User Interface (UI)



Services de base de gLite

Accès à la Grille

User Interface

- **UI** est le point d'accès à la Grille. A travers cette interface, l'utilisateur peut interagir avec le WMS/CE pour la soumission des jobs, avec le SE/FC pour le transfert des données volumineuses et avec le IS pour la découverte des ressources .
- Opérations de base depuis le UI:
  - Lister les ressources convenables pour exécuter un job donné;
  - Opérations sur les jobs (soumission, suivi de l'état, annulation) ;
  - Récupérer les résultats d'un calcul ;
  - Copier, répliquer ou supprimer des données sur la Grille.
- L'accès à la Grille peut être via **CLI**, un **Web portal** ou via les **Sciences gateways**

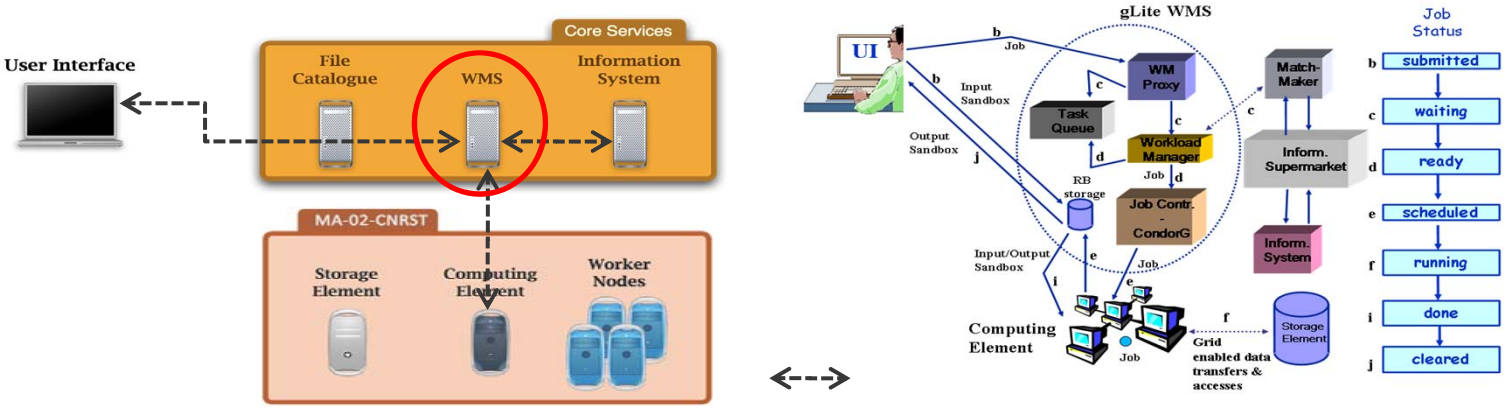
Services de base de gLite

Services de Gestion de Job

Workload Management System (WMS)

WMS est un ensemble de services effectuant toutes les tâches requises pour exécuter les jobs de l'utilisateur, tout en cachant la complexité de la Grille.

- C'est un meta-scheduler faisant la correspondance entre besoins et ressources
- Sélectionne un site (CE) en s'appuyant sur le système d'information qui connait l'état réel des ressources
- Intègre des contraintes sur la localisation des données
- Fonctionnalités avancées: job paramétrique, DAG…



Services de base de gLite

Services de Gestion de Job

Workload Management System (WMS)

Composantes principales du WMS:

Workload Manager (WM) : accepte et satisfait les prérequis du job (Matchmaking)

Logging & Bookeeping (LB) : garde la trace d'exécution du job en fonction de son statut: (Submitted, Running, Done,...)

WMProxy : chargé d'accepter les requêtes entrantes depuis l'interface utilisateur.

Services de base de gLite

Services de Gestion de Job

Computing Element (CE)

- **CE** est le service représentant les ressources de calcul localisé sur un site donnée.
- Il inclut :

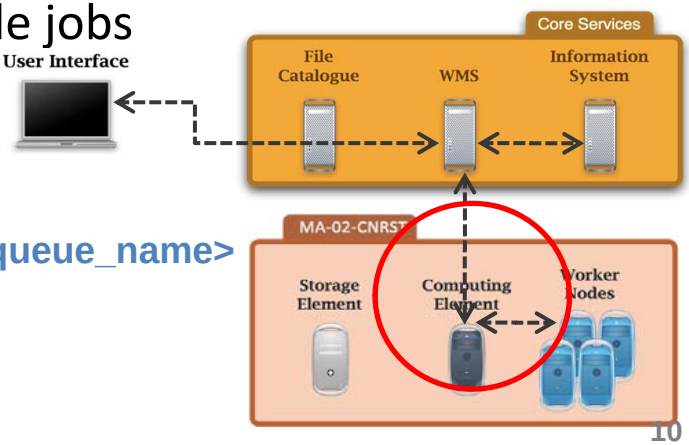
Grid Gate (GG) → interface générique aux ressources de calcul (ex: LCG CE /CREAM CE)

+ **Local Resource Management System (LRMS)** → gestionnaire de queues (ex: OpenPBS, LSF, Maui/Torque..)

+ **Worker Nodes (WN)** → le lieu d'exécution des jobs

- Gère les jobs (soumission et control de jobs)
- Renseigne le WMS des mises à jour des statuts de jobs
- Publie les informations relatives au site

(les queues, statut des CPUs, etc..)



CEId = <gg_hostname>:<port>/<gg_type>-<LRMS_type>-<batch_queue_name>
Ex: ce3.cnrst.magrid.ma:8443/cream-pbs-magridschool

Services de base de gLite

Services de Gestion de Job

WMS/CE

- Le CE peut être utilisé par :
  - **Un client générique** : quand l'utilisateur interagit directement avec le CE (soumission via CE)
  - **WMS** qui soumet le job à un CE choisi suite à un processus de Machmaking (soumission via WMS)
- La soumission d'un job vers le WMS a plusieurs avantages par rapport à la soumission directe vers le CE, en effet, le WMS:
  - Interagit avec plusieurs CEs, il peut ainsi choisir le CE qui répond parfaitement aux besoins du job soumis.
  - En utilisant le service LB, il fournit un suivi global des jobs.
  - Supporte des jobs plus complexes qui ne peuvent être traités directement par les CEs (DAG, collections, parametric)
  - Gère les échecs des jobs (possibilité de re-soumission automatique etc..)

Services de base de gLite

Services de gestion des données

Défis

Hétérogénéité

- Les données sont stockées sur des systèmes de stockage différentes à l'aide de différentes technologies d'accès

Distribution

- Les données sont stockées dans des endroits différents (dans la plupart des cas il n'y a aucun système de fichiers partagé ou un espace de nom commun)
- Les données doivent se déplacer entre différents endroits

Description des données

- Les données sont stockées comme des fichiers (besoin de les décrire et de les localiser selon leur contenu)

Storage Resource Manager — **SRM**

File Catalogue — **FC**

File Transfer Service — **FTS**

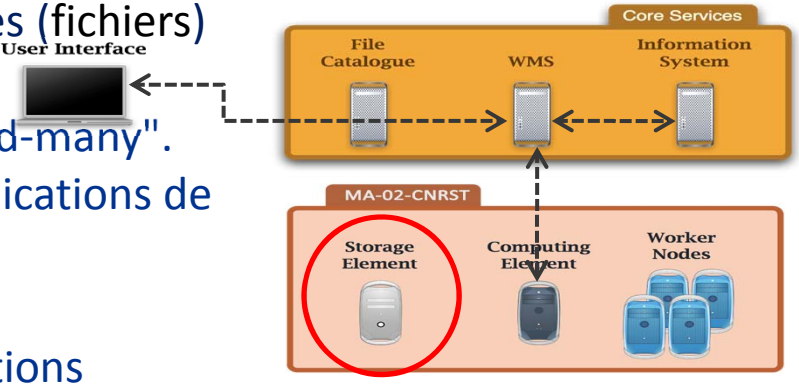
Metadata Service — **MS**

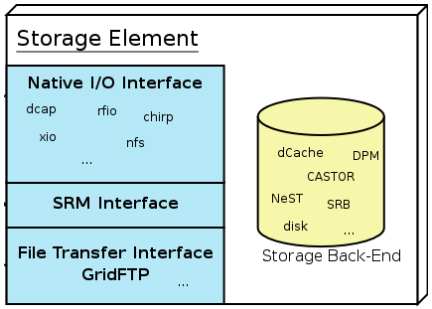
Services de base de gLite

Services de gestion des données

Storage Element (SE)

- **SE** est l'élément d'une Grille représentant la ressource de stockage. C'est le service de gLite qui permet aux utilisateurs et aux applications de stocker/récupérer les données (fichiers)
- Les fichiers localisés sur les SEs…
  - Sont dans la majeur des cas "write-once, read-many".
  - Accessibles par les utilisateurs et par les applications de n'importe où sur la Grille.
  - Peuvent être répliquées sur plusieurs sites.
  - Lecture et suppression sont les seules opérations recommandées
- Les SEs…
  - Fournissent un espace disque dédié aux stockage des fichiers.
  - Fournissent un protocole de transfert (**GSIFTP**)
  - Fournissent une interface pour la gestion des stockages hétérogènes : Storage Resource Manager (**SRM**)
- Types..
  - **dCache**, **Storm**, CERN Advanced STORage manager (**CASTOR**), Disk Pool Manager (**DPM**)



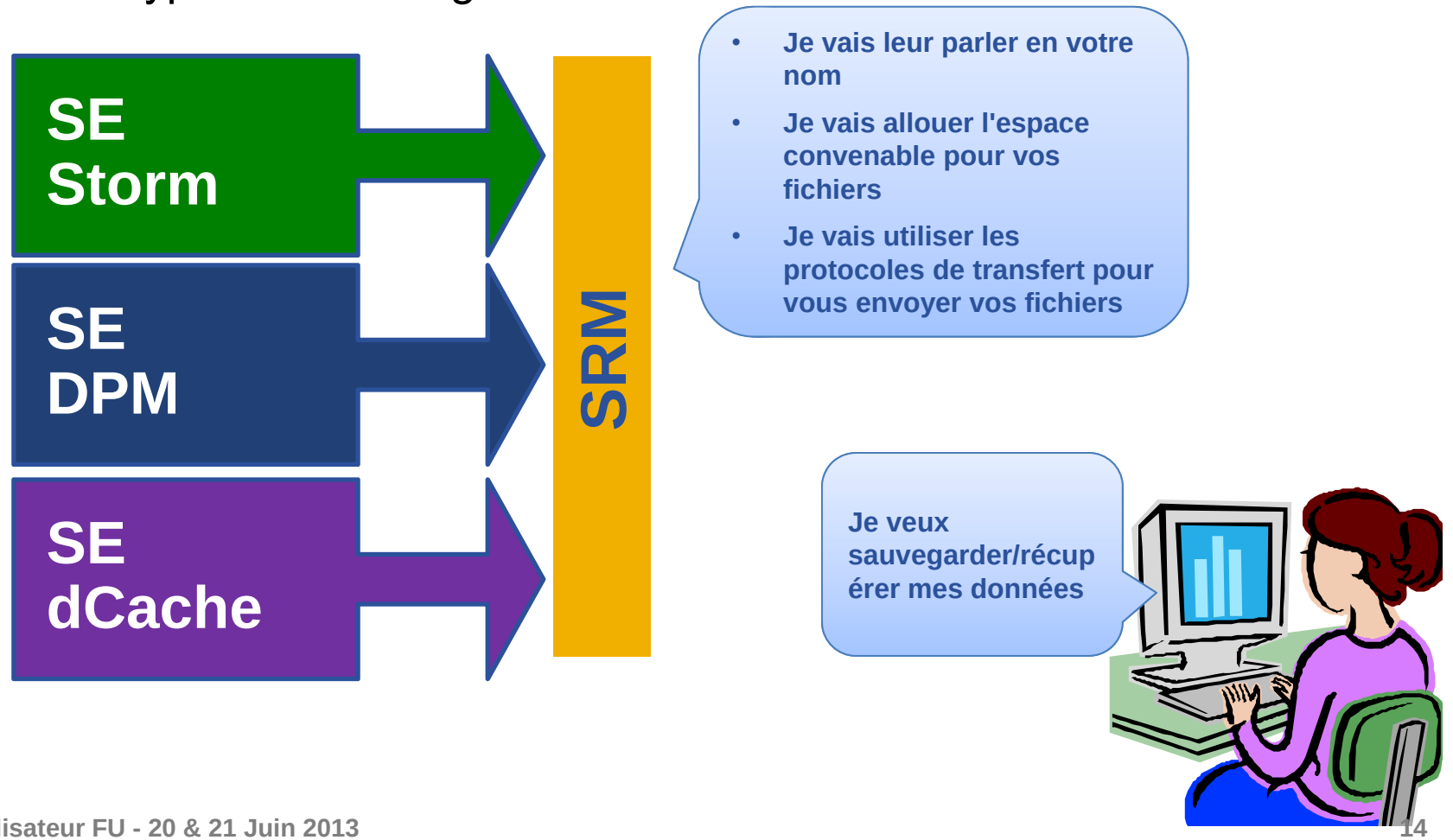


Services de base de gLite

Services de gestion des données

Storage Resource Manager (SRM)

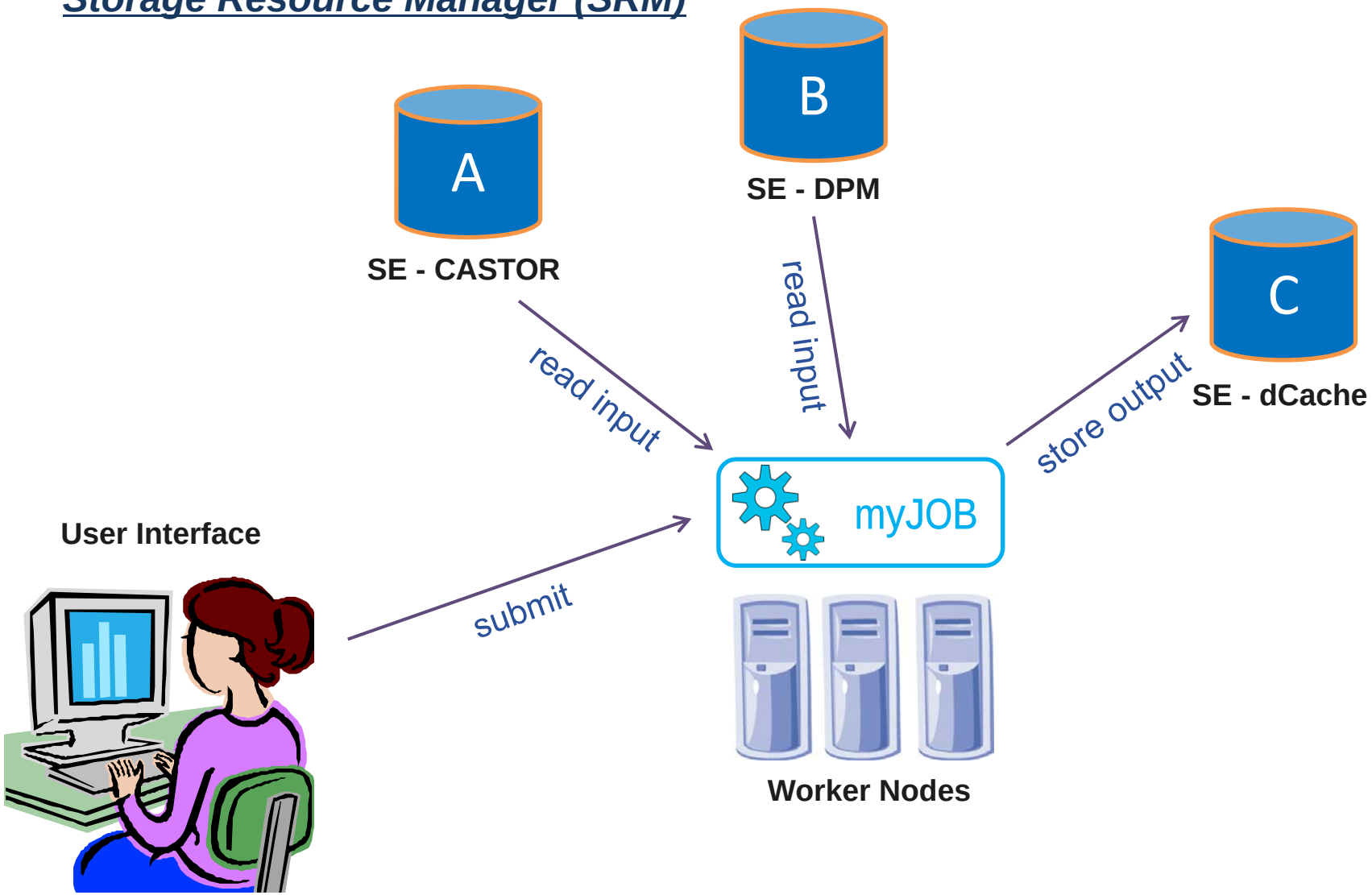
Le SRM est une interface unique qui prenne en charge l'interaction avec les différents types de stockage sur la Grille



Services de base de gLite

Services de gestion des données

Storage Resource Manager (SRM)



Services de base de gLite

Services de gestion des données

Terminologie

- Les fichiers sur la Grille peuvent être référéenciés par plusieurs noms : GUID, LFN, SURL, TURL
- **GUID** : Grid Unique IDentifier
  - Identifiant unique d'un fichier sur la Grille
  - guid:<36_bytes_unique_string>

    Ex: guid:38ed3f60-c402-11d7-a6b0-f53ee5a37e1d
  - Toutes les réplicas d'un fichier se partagent le même GUID
- **LFN** : Logical File Name
  - Un alias plus intuitif qui peut être utilisé à la place du GUID
  - lfn:<any_string>

    Ex: lfn:importantResults/Test1240.dat
  - lfn:/grid/<vo>/<directory>/<file> (cas du LFC)

    Ex : lfn://grid/magrid/elkharrim/myfile.dat

Services de base de gLite

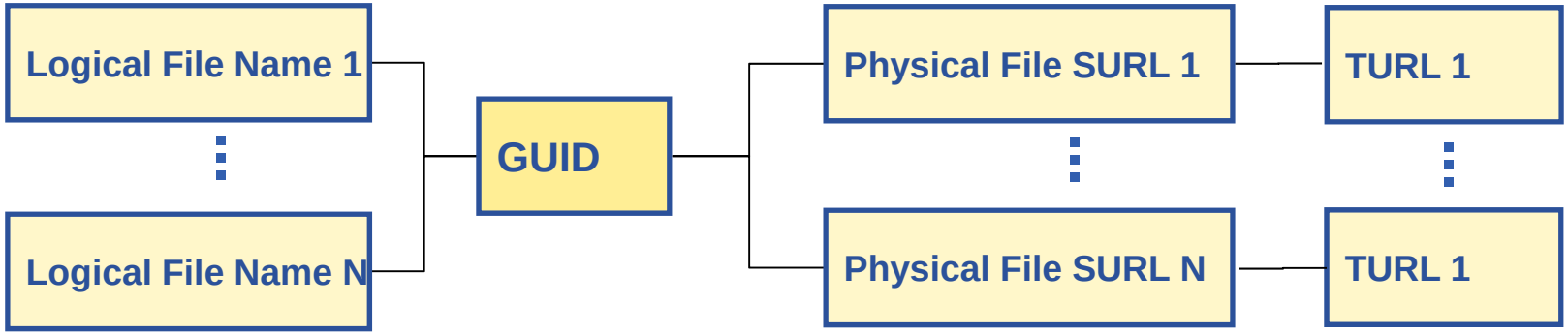
Services de gestion des données

Terminologie

- **SURL/PFN :** Storage URL / Physical File Name
  - La location physique des réplicas d'un fichier
  - srm://<SE_hostname>:<port>/<some_string>

    Ex: srm://se1.cnrst.magrid.ma/dpm/home/magrid/project1/test.dat

  (cas des SEs supportant SRM version 2)
- **TURL** : Transport URL
  - Le lien complet pour accéder un fichier sur un SE (chemin, port et protocole d'accès)
  - Exemple : rfio://lxshare0209.cern.ch//data/alice/ntuples.dat



Services de base de gLite

Services de gestion des données

FC (File Catalog)/LCG File Catalog (LFC)

- **Utilisateurs** et **applications** ont besoin de localiser les réplicas sur la Grille.
- **FC (File Catalog)** est le service qui maintient le mapping entre LFN(s), GUID et SURL(s)
- Il garde l'information sur les réplicas et interagit avec le IS.
- **LFC** (LCG File Catalog, LCG = LHC Computing Grid, LHC = Large Hadron Collider) est le FC adopté par gLite 3.2
- C'est un catalogue intuitif dont l'espace des noms ressemble à la structure d'un dossier dont la / est la racine globale de la Grille (**/grid/<vo_name>/**): Le LFN d'un fichier est de la forme :

lfn:/grid/<vo_name>/<chemin et nom de fichier>

lfn:/grid/magridschool/elkharrim/datasets/ds001.txt

Services de base de gLite

Services de gestion des données

Gérer les fichiers sur la Grille

- Un fichier n'est considéré un «**Grid file**» que s'il est :
  - Physiquement présent sur un SE
  - Enregistré sur un catalogue
- Commandes pour gérer le LFC et les réplicas :

lfc-*, lcg-*

- Les commandes **lcg-*** assurent la consistance entre les fichiers sur les SEs et les entrées sur le FC
- L'utilisation des outils de gestion de données de bas niveau peuvent causer la perte de données sur les SEs et corrompre les données sur le FC → déconseillée sauf si nécessaire.
- Pour interagir avec le LFC depuis le UI, la variable LFC_HOST doit être défini dans l'environnement:

Ex: export LFC_HOST=lfc.magrid.ma

Services de base de gLite

Système d'Information

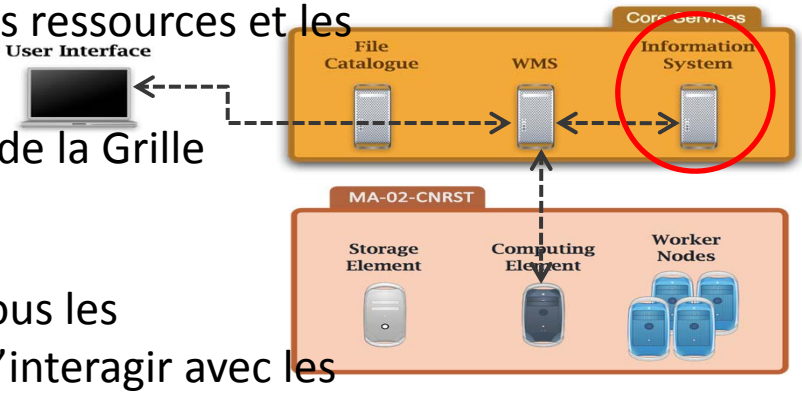
Définition

Qu'est-ce?

- 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 les sites de la Grille
- Vérifier l'état de santé des ressources et services de la Grille



Comment?

- En adoptant un modèle de données commun à 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
- En supervisant localement l'état et la description des ressources/services, et en publiant les données fraichement collectées sur le système d'information

Services de base de gLite

Système d'Information

Utilisateur du SI

- **Utilisateurs finaux de la Grille**
  - Disposer des informations pertinentes sur les ressources disponibles (CPUs libres, Espace de stockage, Applications)
- **Administrateurs du site**
  - Publier des informations sur les ressources et les services qu'ils fournissent
- **Middleware**
  - **WMS**: faire correspondre les exigences du job et l'allocation des ressources
  - **Services de monitoring**: Récupération d'informations sur l'état et la disponibilité des ressources

Services de base de gLite

Système d'Information

Principe

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

- Les "Tags" du software installé

Services de base de gLite

Système d'Information

GLUE Schema

- Pour réponde aux exigences liées à l'hétérogénéité des ressources et leur dispersion géographique, le modèle de données adopté par gLite est le **GLUE Schema (Grid Laboratory Uniform Environment)**
- Le GLUE Schema est une abstraction des ressources de la Grille développé par OGF (Open Grid Forum)
- Implémentations : LDAP, RDBMS,XML,…
- L'implémentation la plus utilisée est le **BDII (Berkeley DB Information Index)**

Services de base de gLite

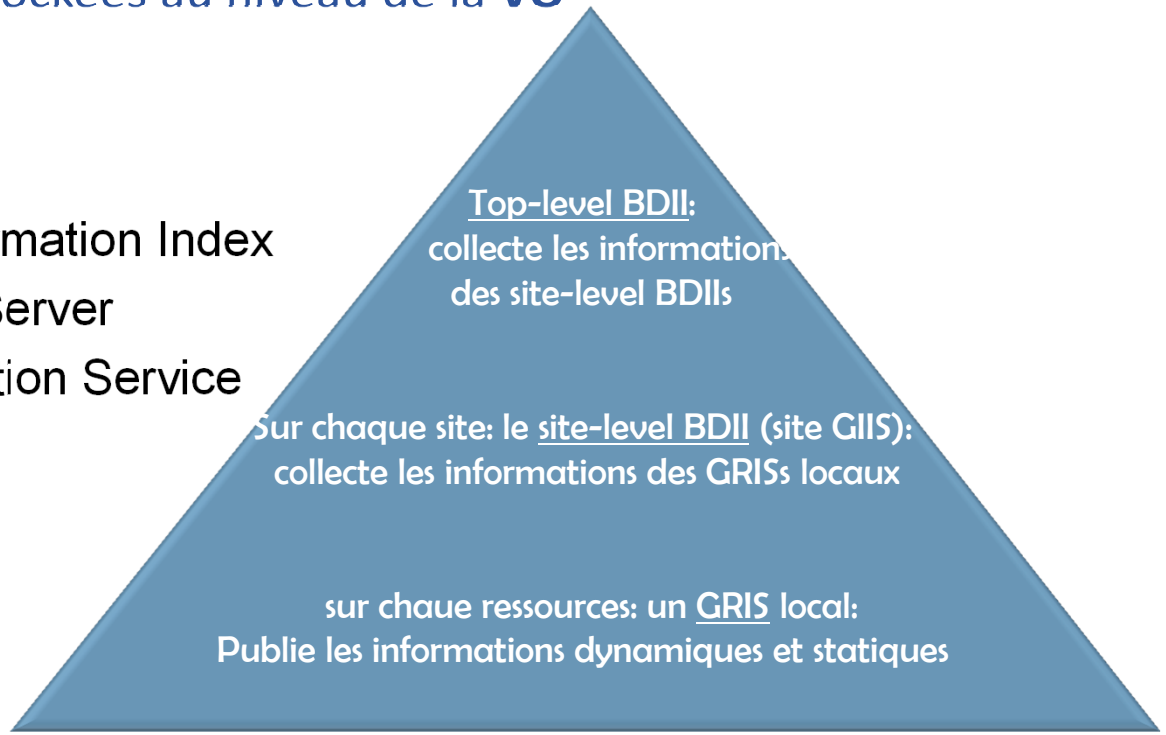
Système d'Information

Architecture

L'information est organisée selon une architecture à **3 niveaux**:

- GRIS : Informations stockées au niveau des **ressources**
- GIIS/Site BDII : Informations stockées au niveau du **site**
- Top BDII : Information stockées au niveau de la **VO**

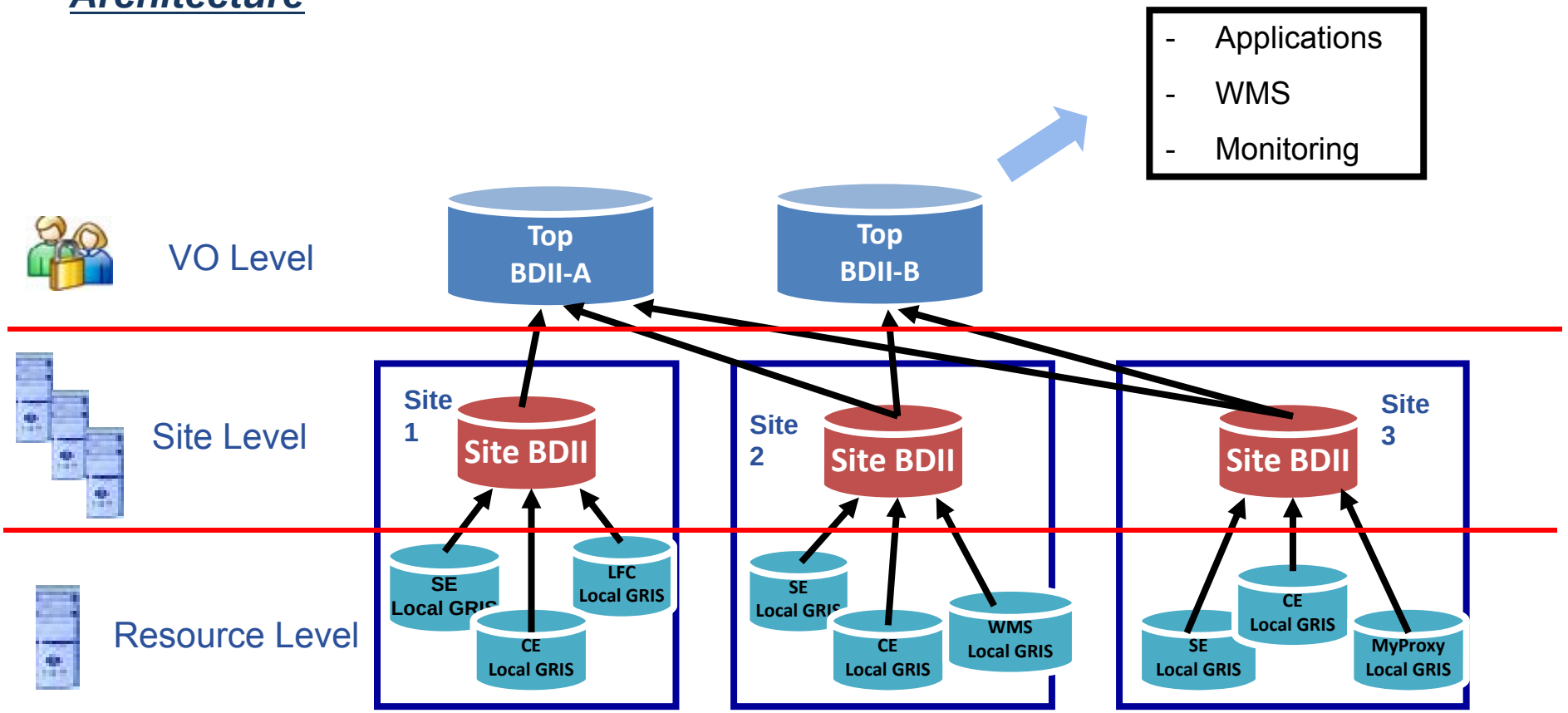
BDII : Berkeley Database Information Index
GIIS : Grid Index Information Server
GRIS : Grid Resource Information Service



Services de base de gLite

Système d'Information

Architecture



Services de base de gLite

Système d'Information

lcg-infosites et lcg-info

- 2 commandes principales qui servent comme outils d'interrogation du SI:
  - **lcg-infosites :** pour récupérer des informations sommaires et formatées sur les ressources de la Grille correspondant aux cas d'utilisation les plus courants.

    (Ex: lcg-infosites --vo atlas ce)

  - **lcg-info** supportant des requettes plus complexe

    (Ex: lcg-info --vo atlas --list-ce --query 'PlatformArch=x86_64')
- L'information est tirée du **BDII** spécifié par le la variable d'environnement **LCG_GFAL_ INFOSYS** ou bien dans la ligne de commande. (Ex: export LCG_GFAL_ INFOSYS =bdii.magrid.ma)

Services de base de gLite

Infrastructure de Sécurité sur la Grille

Grid Security Infrastructure (GSI)

GSI est basé sur les **Certificats X.509** et les **PKI**

- Implémente :
  - **Single Sign-On**: le mot de passe n'est donné qu'une seule fois
  - **Délégation**: un service peut-être utilisé au nom d'une autre personne c-à-d autoriser une autre entité à utiliser son authentification et ses autorisations
  - Processus d'**Authentification mutuelle**
- Introduction des **certificats proxy**
  - Certificat à durée de vie courte, contenant la clé privée, signé avec le certificat de l'utilisateur
  - Le proxy peut se déplacer sur le réseau
  - Le proxy est utilisé pour l'authentification effective aux services de la Grilles

Services de base de gLite

Infrastructure de Sécurité sur la Grille

Certificat X.509

- Repose sur l'utilisation de la Cryptographie asymétrique (**RSA**) et l'accréditation par l'Autorité de Certification (**CA)**

- Un certificat **X.509** peut être issu pour:
  - Une personne physique (certificat personnel)
  - Une machine (certificat hôte)
  - Un programme (certificat de service)

Certificat X.509

- Les certificats sont conservés dans des **FICHIERS**
- Il existe plusieurs formats de représentation des certificats
  - **PKCS12** (format navigateur web)
    - Extensions **.p12** ou **.pfx**
    - La clé privée et la clé publique sont dans **un même fichier**
    - Le fichier est chiffré et protégé par un mot de passe
    - La plupart des CA délivrent les certificats personnels dans ce format
  - **PEM** (format « Grille »)
    - Extensions **.pem ou .crt et .key**
    - La clé privée et la clé publique sont dans **2 fichiers distincts**
    - Le fichier contenant la clé privée est chiffré et protégé par un mot de passe

Certificat X.509

- **La Clé Publique (le Certificat)**
  - Signée par l'Autorité de Certification après vérification de l'identité du destinataire
  - Publiée sur le réseau via le service de publication de la CA
- **La Clé Privée**
  - Conservée par le navigateur de l'utilisateur et dans son dossier home sur l'UI
  - Chiffrée et protégée par un mot de passe

Services de base de gLite

Infrastructure de Sécurité sur la Grille

Certificat X.509

- Un couple de clés indissociables
- Les clés sont générées ensembles
- Impossibilité de retrouver une clé à partir l'autre
- Durée de vie d'un certificat : un an à compter de la date de la demande
- Renouvelable

Certificat X.509

Informations importantes contenues dans un **certificat (clé publique)**:

- Le sujet ou DN du certificat
- Le numéro de série du certificat
- La période de validité du certificat
- L'Autorité de Certification émettrice
- La clé publique
- Des extensions X509v3
  - Les utilisations autorisées du certificat
  - L'email
  - ...
- La signature de la CA émettrice

Services de base de gLite

Infrastructure de Sécurité sur la Grille

openssl

Obtenir la clé privée

openssl pkcs12 -nocerts -in cert.p12 -out userkey.pem

Obtenir la clé publique

openssl pkcs12 -clcerts -nokeys -in cert.p12 -out usercert.pem

Changer le mot passe de la clé privée

openssl rsa –in userkey.pem -des3 -out userkey.pem.new

Visualiser une clé publique

openssl x509 -text -noout -in usercert.pem

openssl pkcs12 -info -in cert.p12

Services de base de gLite

Infrastructure de Sécurité sur la Grille

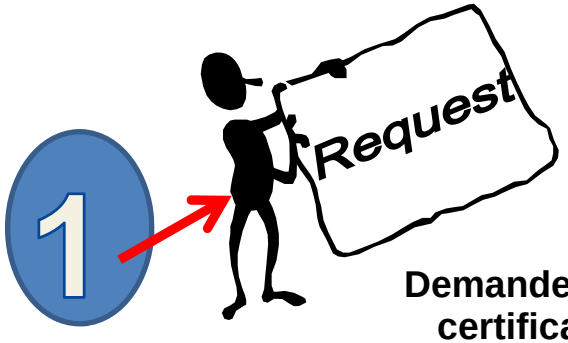
Autorités de Régistration (RA) et de Certification (CA)

- L'**Autorité de Régistration (RA)** valide les demandes de création et de révocation de certificat
- L'**Autorité de Certification (CA)** est une entité digne de confiance qui se charge de délivrer les certificats
- Politique de gestion des autorités : **Grid PMA (Policy Management Authority)**
  - Etablir des obligations minimales pour les CA
  - Accréditer les CA
  - Auditionner les CA

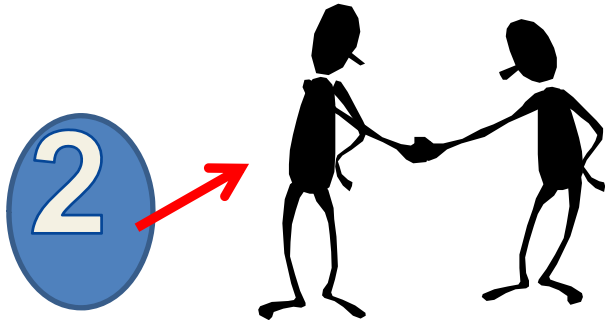
Services de base de gLite

Infrastructure de Sécurité sur la Grille

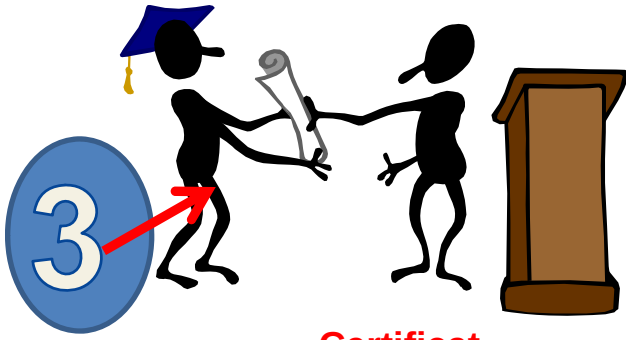
Processus schématisé de demande de Certificat



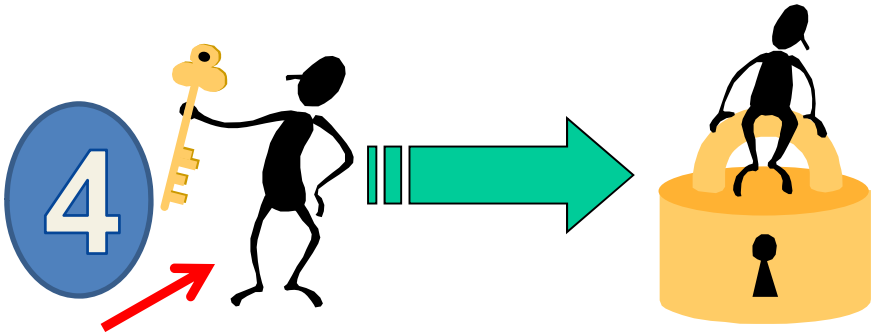
Demande de certificat



Utilisateur identifié et confirmé par la RA



Certificat émise par la CA



Le certificat est utilisé pour accéder à la grille

Proxy

Les jobs:

- Interagissent avec les différents services de la Grille.
- Utilisent les ressources de la Grille.
- Doivent avoir les mêmes privilèges que l'utilisateur.
- S’exécutent ailleurs là où la clé privée de l’utilisateur n'est pas disponible.

L’identité de l’utilisateur doit être déléguée au job

Proxy

- Un **Proxy Certificate (Proxy)** est conçu pour agir au nom de l'utilisateur (délégation).
  - Certificat à durée de vie courte, contenant la clé privée, signé avec le certificat de l'utilisateur
  - Peut se déplacer sur le réseau
  - Doit être crée ou valide avant de soumettre un job sur la Grille.
- Créer un Grid proxy standard:

grid-proxy-init

Services de base de gLite

Infrastructure de Sécurité sur la Grille

Virtual Organisation Management Service (VOMS)

VOMS est le service qui permet à un proxy d'avoir des extentions contenant des informations sur le VO, les groupes dont l'utilisateur fait partie dans un VO et ses rôles.

- Donne l'autorisation aux utilisateurs à accéder aux ressources au niveau de la VO
- La base de données du VOMS contient l'ensemble des membres avec leur niveau d'autorisation
- Un utilisateur peut avoir plusieurs niveaux d'autorisation dans chaque VO et peut faire partie de plusieurs VO

Services de base de gLite

Infrastructure de Sécurité sur la Grille

group :

sous ensemble du VO contenant des membres qui se partagent des privilèges ou des responsabilités dans un projet.

- Les groupes sont hiérarchiques, profondeur non limitée
- Permet de moduler les droits des membres de la VOMS en fonction de leur groupe
- Le groupe par défaut est /<vo-name>

role :

est un attribut typique qui permet à un utilisateur d'avoir certain privilèges pour accomplir certain taches.

- Software manager, VO-Administrator, …
- Les rôles ne sont pas hiérarchiques : il n'existe pas de sous-rôle
- Les rôles doivent être explicitement spécifiés lors de la création du proxy

Créer un VOMS proxy (courte durée)

- Créer un proxy

  voms-proxy-init --voms <vo-name>
- Obtenir des informations sur un proxy

  voms-proxy-info -all
- Créer un proxy en spécifiant un groupe

  voms-proxy-init --voms <vo-name>:/group/subgroup
- Créer un proxy en spécifiant un rôle

  voms-proxy-init

  --voms <vo-name>:[/group]/role=production
- Détruire un proxy

  voms-proxy-destroy

proxy de longue durée

- Pour des raison de sécurité, c'est une mauvaise idée de prolonger la vie d'un VOMS proxy (par défaut **12h**).
- Cependant, un job peut avoir besoin d'autorisations plus longues.
- Solution :**serveur MyProxy**, serveur dédié pour stocker les proxys longs.
- Le WMS utilise les proxys longs pour effectuer le renouvellement automatique des proxys pour les jobs soumis sur la Grille.
- Créer et stocker un proxy de long durée (7j par défaut/ modifiable par l'option –c):

myproxy-init -s <myproxy_server> -d –n