

1 Cours

Question 1.1. Qu'est ce qui rapproche le *grid computing* du *cloud computing*? Qu'est-ce qui les différencie?

Question 1.2. Qu'est-ce qu'une organisation virtuelle (VO)? Expliquez pourquoi c'est un élément essentiel pour le passage à l'échelle (*scalability*) de la technologie de grille.

Question 1.3. Qu'est-ce qui authentifie un utilisateur sur la grille? Comment sont gérées les autorisations?

Question 1.4. Qu'est ce qu'un *job manager*? Pourquoi est-il couramment utilisé pour le calcul distribué?

2 Problème

Un consortium de grands musées nationaux répartis sur l'ensemble de la planète souhaite mettre en place une plateforme collaborative pour la numérisation, le stockage et le partage à grande échelle de leurs collections. Chaque musée dispose du matériel de numérisation et veut mettre à disposition des membres du consortium les oeuvres numérisées de sa collection.

Gridtek, une jeune entreprise innovante, propose de mettre en place une infrastructure de grille s'appuyant sur le middleware gLite pour répondre au besoin du consortium.

Question 2.1. Quels sont les arguments que Gridtek peut utiliser auprès du consortium pour mettre en avant la technologie de grille?

Question 2.2 (Infrastructure).

- Quels sont éléments de grille que l'entreprise doit déployer sur chaque site?
- Quels sont les éléments de la grille qui peuvent être mutualisés entre tous les sites?
- Quel élément de la grille permet d'établir un catalogue de toutes les oeuvres numérisées?

Question 2.3 (Catalogue). Pour une oeuvre donnée d'une collection, on veut pouvoir conserver plusieurs numérisations. On organise donc le catalogue des numérisations de la façon suivante :

```
/grid/museum/<musée>/<artiste>/<oeuvre>/<fichier_numérisé>
```

- Ecrivez le script (à exécuter sur l'UI) qui permet de lister les oeuvres d'un musée.
- Ecrivez le script qui permet de savoir où est stockée physiquement la numérisation d'une oeuvre.

Question 2.4 (Catégories). Les conservateurs veulent avoir la possibilité d'organiser leurs oeuvres par catégories. Gridtek propose pour cela d'utiliser des alias dans le catalogue.

- a. Proposez un schéma de nommage des catégories dans le catalogue.
- b. Ecrivez le script qui permet au conservateur d'ajouter et de supprimer une oeuvre dans une catégorie.
- c. Ecrivez le script qui permet au conservateur de lister toutes les catégories associées à une oeuvre.

Question 2.5 (Transferts).

- a. Ecrivez le script qui permet à un musée de transférer la numérisation d'une oeuvre sur son propre SE et qui l'enregistre au catalogue.
- b. Ecrivez le script qui permet à un utilisateur de récupérer la numérisation d'une oeuvre.

Question 2.6 (Réplication).

- a. Comment peut-on lister tous les éléments de stockage (SE) disponibles ?
- b. Ecrivez la commande pour répliquer un fichier sur un second SE.
- c. Gridtek propose de mettre un système de cache en utilisant les mécanismes de réplication. Lorsqu'un utilisateur du musée de Paris, par exemple, veut récupérer une numérisation du musée de New York, celle-ci sera préalablement répliquée, si besoin, sur le SE parisien. Modifiez en conséquence le script de la question 2.5 b.

Question 2.7 (Analyses d'images). Certains sites proposent des ressources de calcul pour faire de l'analyse d'image avec le logiciel `img_analyze` :

```
img_analyze -f <image> -p <sensibility>
<image>      : nom du fichier à analyser
<sensibility> : sensibilité de l'analyse (valeur de 0 à 100)
```

Les résultats sont écrits sur la sortie standard.

On veut pouvoir lancer des jobs sur la grille qui exécutent une analyse d'image pour un fichier donné par son LFN et pour une sensibilité donnée. On suppose que l'un des sites fournit un WMS.

- a. Proposez le diagramme de séquence du job.
- b. Ecrivez le script du job qui sera exécuté sur le WN.
- c. Ecrivez le JDL associé.
- d. Ecrivez les commandes qui permettent de lancer le job et de récupérer les résultats.
- e. Modifiez le script du job pour que les résultats soient transférés sur un SE.