

# LCG-France Tier-1 & Analysis Facility

## Relevé des Conclusions

7 Juillet 2009 – 10h30-12h30

---

**Présents** : Claude Charlot, Tibor Kurca, Farida Fassi, Suzanne Poulat, Yvan Calas, Jonathan Schaeffer, Pierre Girard, Fabio Hernandez

**Président et Secrétaire**: Fabio Hernandez

**Agenda** : <http://indico.in2p3.fr/conferenceDisplay.py?confId=2069>

---

### 1. Introduction

Ce document est le résumé des discussions et des conclusions lors de la réunion de suivi des activités CMS au tier-1 & AF LCG-France au CC-IN2P3. L'objectif de la réunion était d'identifier les mécanismes pour séparer complètement les espaces de stockage du tier-1 de ceux du tier-2 CMS.

### 2. Rappel du contexte

Actuellement, 2 sites WLCG sont opérés par le CC-IN2P3 pour CMS. Un tier-1 et un tier-2. La configuration des CEs est différente pour ces 2 sites, tel que montré dans le tableau ci-dessous :

Tier	Computing Elements
Tier-1	cclcgcelio3.in2p3.fr cclcgcelio4.in2p3.fr
Tier-2	cclcgcelio5.in2p3.fr cclcgcelio6.in2p3.fr

Concernant les SEs, un seul end-point SRM est publié pour le stockage des données CMS pour le tier-1 et le tier-2 : `ccrsm.in2p3.fr`. L'ensemble de données CMS est géré par dCache.

Cette situation est particulière et il s'agit de différencier de façon plus stricte les espaces de stockage de données de ces 2 sites, tout en faisant en sorte qu'une seule instance de dCache gère les 2 espaces.

### 3. Séparation tier-1 et tier-2

Afin de séparer les espaces disque du tier-1 et du tier-2, les étapes suivantes ont déjà été accomplies :

- **Storage Element** : le end-point SRM `ccsrmt2.in2p3.fr` a été créé, actuellement sous la forme d'un alias vers le SRM du tier-1 (`ccrsm.in2p3.fr`). Ce SE est publié dans

le système d'information du site. Le CloseSE des computing elements du tier-1 est `ccsrn.in2p3.fr`. Les computing elements du tier-2 voient 2 CloseSEs (`ccsrn.in2p3.fr` et `ccsrmt2.in2p3.fr`) et celui publié comme le défaut est `ccsrmt2.in2p3.fr`.

- **PhEDEx** : Une instance spécifique pour le tier-2 a été déployée sur la VO box `cclcgcms06.in2p3.fr`. Elle est séparée de celle du tier-1 et ne contient qu'un espace disque (pas de MSS attaché). Tibor en est le responsable.
- **Software Area** : les jobs soumis via les CEs du tier2 vont utiliser une software area différente de celle des jobs tier-1 et des fichiers de configuration spécifiques à CMS (par exemple, le mapping entre LFN et PFN) différents entre le tier-1 et le tier-2. Néanmoins, le logiciel CMS n'est installé qu'une seule fois et est utilisable aussi bien par les jobs tier-1 que tier-2. La publication des tags spécifiques à CMS est faite centralement. Tibor est le responsable de la software area CMS pour le tier-1 et le tier-2.

Fabio propose de simplifier l'installation actuelle qui est faite à base de liens symboliques et source potentielle d'erreurs. L'objectif est d'avoir une zone où le logiciel proprement dit est installée et 2 zones séparées (une pour le tier-1 et une autre pour le tier-2) avec les fichiers de configuration et de setup spécifiques pour ces sites. **ACTION** : Tibor explorera les moyens de simplifier cette installation.

CMS a actuellement 2 software areas (gérées par AFS, non répliquées):

- `/afs/in2p3.fr/grid/toolkit/cms`
- `/afs/in2p3.fr/grid/toolkit/cms2`

**TIBOR** : Est-ce que ces 2 zones sont toujours nécessaires ? Actuellement uniquement « cms2 » est utilisée aussi bien pour le tier-1 que pour le tier-2.

Les valeurs de la variable `VO_CMS_SW_DIR` pour les 2 sites sont présentées dans le tableau ci-dessous :

Tier	Valeur de la variable <code>VO_CMS_SW_DIR</code>
Tier-1	<code>/afs/in2p3.fr/grid/toolkit/cms2</code>
Tier-2	<code>/afs/in2p3.fr/grid/toolkit/cms2/T2_VO_CMS_SW_DIR</code>

- **Disk Pools dCache** : les pools dCache du tier-1 et du tier-2 ne sont pas différenciés. C'est la prochaine étape de l'implémentation de la séparation.

#### 4. Séparation des pools dCache tier-1 et tier-2

Après avoir exploré plusieurs pistes, nous sommes arrivés à la conclusion qu'il est nécessaire de :

- Créer un pool group dCache pour les données tier-2 (analyse, monte-carlo, user) du type `ToD1` (disk-only)
- Créer un espace de nommage pour ces données dans la hiérarchie CMS gérée par dCache, dont les fichiers seront physiquement dans les pools du tier-2. Le nommage serait du type :
  - `/pnfs/in2p3.fr/data/cms/data/store/analysis`

- /pnfs/in2p3.fr/data/cms/data/store/**MC**
- /pnfs/in2p3.fr/data/cms/data/store/**user**
- Une fois cette configuration en place, les outils spécifiques CMS (l'agent PhEDEx du tier-2, DLS, etc.) seront configurés pour tenir compte de ce nouveau contexte. Les transferts des fichiers vers et depuis la zone tier-2 seront déclenchés par PhEDEx comme s'il s'agissait d'un site tier-2 distant.
- **ACTION** : En tant que responsable du tier-2, Tibor se mettra en contact avec les experts dCache pour mettre en place cette nouvelle configuration, tout en veillant à ce que les modifications soient cohérentes avec le fonctionnement du tier-1, dont Farida a la responsabilité.

## 5. Accès aux données par les jobs locaux non-grille

L'accès aux données par les jobs locaux non-grille des membres de CMS-France a été discuté. Il est envisagé d'utiliser le service xrootd pour l'accès performant aux données, en profitant des capacités de xrootd de stager à la demande et automatiquement les données depuis les serveurs de disques de dCache.

Cette configuration sera explorée ultérieurement.

## 6. Divers

- Suzanne propose d'augmenter le seuil de la ressource `u_cms_dcache`. Claude propose que les shares BQS entre tier-1 et tier-2 soient programmés selon les engagements du CC-IN2P3.
- Suite à la réunion, Claude a transmis ses souhaits de configuration des shares à l'intérieur de CMS. Le mail de Claude daté du 15/07/2009 est reproduit ci-après :

*Tout peut être calculé depuis nos pledges et le rapport  $T3/(T2+T3)$  ou  $T3$  signifie actuellement accès local via BQS (1). Nous (CMS-France) souhaitons actuellement  $T3/(T2+T3)=40\%$ . Avec les pledges  $T1$  et  $t2$  pour 2009 ( $T1=1825$  kSI2K,  $T2=845$  kSI2K) ça donne:*

- $T1=57\%$
- $T2=26\%$
- $T3=17\%$

*Pour la partie  $T2$ , la répartition entre prod et analyse devrait être 50/50 plutôt que 75/25 comme actuellement.*

*(1)  $T3$  signifie en fait une partie accès local ( $T3$  local) + une partie accès grille privilégié sur le  $T2$  ( $T3$  grille). Nous aurons donc rapidement besoin de mettre en place un système permettant de donner une priorité aux utilisateurs CMS-France sur le  $T2$ . Un groupe VOMS vient d'être créé pour cela. Quand il sera actif (à la rentrée), la fraction 40% devra représenter le total  $T3$  grille (ie acces  $T2$  utilisateurs CMS-France) +  $T3$  local (ie accès BQS).*