

# Résumé CHEP 2010

## Thèmes :

Distributed processing and analysis  
Grid and cloud middleware



# Distributed processing and analysis

Le computing LHC a bien démarré : computing models adaptés

- traitement des données T0 - CAF
- Monte Carlo (simulation)
- formatage des données (skimming, thinning, ...)
- transfert/distribution des données sur la grille
- re-processing / calibration

ATLAS - LHCb  
ALICE - CMS

⇒ les modèles fonctionnent pour les tâches officielles et centralisées  
( tests prévisionnels, outils de monitoring, ... )

⇒ comportement de l'utilisateur  $\lambda$  difficilement modélisable ☺

- évaluer l'impact des utilisateurs sur le modèle
- fournir des outils simples et adaptés
- former les utilisateurs

} Analyse

➡ Il faut être réactif à l'évolution du comportement du modèle (monitoring ↗)  
adaptatation du placement des fichiers, effacement des  
données, protocoles d'accès aux données, ...

↗ luminosité du LHC : comment les modèles vont-ils se comporter ?

# Distributed processing and analysis

Les thèmes abordés (1/2):

WMS / job pilotes / CREAM CE  
WMagent(CMS), Dirac, ALICE, ...

Ganga vs Dirac vs Panda (LHC + Belle + SuperB)  
Phenix : data oriented job submission tool

simulation &  
analyse

Accès aux données :

- vérifier l'accessibilité des données avant de soumettre un job
- outils "data oriented"
- fermes PROOF
- protocoles : xrootd / WebDAV (dCache)
- SQUID / utilisation de caches distribués

Outils de test / de simulation de charge et d'utilisation

Monitorer / dashboards/... (mais un graphe ne fait pas tout...)



Optimisation de l'utilisation des ressources

# Distributed processing and analysis

Les thèmes abordés (2/2):

Virtualisation :

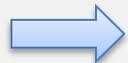
- tests de différents VM / différents OS : efficacités CPU, mémoire, débit réseau ⇒ les résultats sont très dépendants de VM/OS
- tests 2x4 cores machine native / VM : nombre de VM ? pinning des processus ?

Computing models de Belle, SuperB , Fermi, Gamma-Ray Space Telescope, ...

- s'appuient sur l'expérience du LHC
- utilisent les outils de grille



"simplification"



Investir dans les réseaux

//

investir dans le stockage

et "promener" les  
données / job

"multiplier" le stockage et soumettre le  
job là où sont les données

# Grid and cloud middleware

- EGEE / EGI a 10 ans
  - le middleware fonctionne plutôt bien
  - encore et toujours incomplet : processus de développement non prédictible, manque de stratégie
  - cycles de développements, tests, documentation
  - fonctionnalités manquantes : chaque VO se débrouille

**?** : OS unique, machines multi-coeurs, ...

- nouveau middleware EMI (European Middleware Initiative)
  - réunir les différents projets européens : gLite, ARC et Unicore
  - ⇒ présentation du plan de travail du projet

## ➤ Data Management

- tests & intégration : NFS4, WebDAV, ...
- ↗ protocoles existants (DPM,SRM/FTS) et outils de synchronisation (LFC)
- SRM dCache avec Terracotta // ARC DM // Chelonia // StorNet

# Grid and cloud middleware

- nouveau middleware EMI (European Middleware Initiative)
  - WMS vs Condor-G vs ARC CE / CREAM CE vs ...
  - Sécurité
    - ↗ sécurité, procédures en place
    - inter-opérabilité entre grilles
    - jobs pilotes, jobs pilotes multi-users (glexec, Argus, ...)
  - monitoring, accounting
  - meilleur gestion des packages du middleware (disponibilité, déploiement,...)
- visibilité, métrique, chiffres ⇨ availabilities & reliabilities
- outils de gestion de l'information ( # information ↗ )

# Grid and cloud middleware

## **Grid + Virtualisation = Cloud computing**

- Virtualisation (flexibilité des ressources (OS), adaptabilité VO, ...)
  - problématique : création/effacement des images (temps), images partagées, droits sudo, sécurité
  - différentes solutions/projets à l'échelle des labos/clusters
  - évaluation de la perte efficacité
- Cloud computing
  - accès sites clouds privés (sécurité/autorisation ?, coût ?)
  - un cloud public peut-il s'adapter à un usage privé et vice-versa ?



outils +/- matures

( déployés localement, pb de passage à l'échelle)

problème de la dualité cloud privé/public

- Exemples :
- CMS / CRAB : calculs en période normale : grille / pic d'activité : cloud
  - Condor et Dirac proposent des jobs pilots en mode cloud
  - Belle : 1/3 de la production Monte Carlo a été effectuée sur Amazon
  - présentation StratusLab, solutions locales (posters)