

# Test de DIRAC pour le DC2013

Johann Cohen-Tanugi, LUPM

réunion LSST-France, 12-19 déc. 2013

## Rappels “historiques”

- C. Arnault mentionne DIRAC lors de sa présentation au meeting LSST-France au LPC (11-12 fév. 2013)
- J'étais dans la salle, et j'ai pris la balle au bond...
- 1ère téléconf. le 18 mars
- Mise en place d'un environnement de test, et à participation à l'effort DC2013 de cet été.
- (J'espère faire rentrer le LUPM officiellement dans LSST d'ici fin 2014)



# DIRAC ? 2 slides d'Andreï (CPPM)

- ▶ General purpose middleware based on the large user communities experience
- ▶ Complete set of services:
  - ▶ Computing **resources management** system
  - ▶ Computing infrastructure neutral **workload management** system
  - ▶ Complete **data management** including data placement, bookkeeping, classification ( metadata )
  - ▶ Production management for massive workload and data manipulation
  - ▶ Various friendly interfaces including web portal
- ▶ Open software framework easy to expand for the needs of new communities.
- ▶ Offered as a general purpose service by France-Grilles



# Resource Management

- ▶ Conventional Grid computing resources
  - ▶ gLite middleware based grids: EGI, GISELA, OSG, ...
  - ▶ The work is in progress for ARC middleware based grids
  - ▶ NorduGrid
  - ▶ Other types of grids can be supported
  - ▶ As long we have customers needing that
- ▶ Clouds
  - ▶ Providing a cloud broker to integrate various cloud managers
  - ▶ OpenStack, OpenNebular, Amazon, ...
  - ▶ Work in progress as the cloud environment is not yet stable
- ▶ **Standalone CC/clusters**
  - ▶ **Using SSH access, no grid middleware required**
- ▶ Desktop grids



# DIRAC au LUPM ?

- Compétence grille on-site (noeud de grille EGI MSFG-Open)... mais surtout
- Management de la production Monte-Carlo de CTA sur grille
  - Effort continu depuis ~2012 :
    - 2.2 M jobs
    - 600 TB écrits
    - 3.7 M d'enregistrements dans le DFC avec MD
  - Expertise locale pour déployer les analyses finales sur grille également : plusieurs chaînes d'analyse tournent également sur la grille via DIRAC.



# DIRAC pour le DC2013

[http://lsst.in2p3.fr/wiki/index.php/Tests\\_de\\_DIRAC\\_pour\\_LSST](http://lsst.in2p3.fr/wiki/index.php/Tests_de_DIRAC_pour_LSST)

- Objectif volontairement limité :
  - Utilisation de la VO France-Grille plutôt que LSST (qui n'est pas encore dans EGI)
  - Le stack difficile à compiler et nécessite SL6 : restriction du test d'abord à la machine cclsst, puis au sous-cluster sl6 au printemps → solution “DIRAC sur cluster”
  - Mise en place +/- en même temps que la préparation “officielle” DC2013 (see Philippe et Dominique) : restriction du déploiement aux scripts de création des calexp dans le filtre z.



# Retour d'expérience I

- DIRAC déployé et utilisable en “configuration France-Grille”
  - Les scripts bash sont extrêmement simples à déployer sous DIRAC, qui fournit en sus l'interface web de monitoring et d'audit (CPU time, efficacité, etc...), et permet de lancer des prods sous login personnel
  - Par contre demande évidemment l'“effort” initial d'obtenir un certificat
- Très simple d'enregistrer un espace sps comme “site storage”, mais jugé pour l'instant inutile.
- En parallèle, un effort au stade beta d'implémentation d'un service iRods.



# Retour d'expérience II

- Clairement un premier test limité :
  - Pas le temps pour Philippe de lancer une partie de sa prod sous DIRAC, pour avoir un retour d'un utilisateur lambda
  - Le format cluster limite forcément l'impact d'un framework orienté grille comme DIRAC
  - Pas de gestion de storage ou de catalogue de fichiers.
- Mais un test positif!
  - Le système est déployé, et peut être configuré plus finement (expertise d'Andreï et de Luisa, plus discussions permanentes avec Ricardo Graciani au sein de CTA)
  - La mise en route d'une séquence de production (calexp filtre z) a réellement été très simple
  - L'ergonomie du portail DIRAC n'a plus qu'à être mise à profit



# La suite ?

- Optique “French Data Challenges” :
  - Poursuivre DIRAC pour interfacer le CC en mode cluster et bénéficier à peu de frais de son ergonomie
    - Synergie avec CTA dans le développement d'un système de production DIRAC pour les besoins de ces 2 expériences.
    - DIRAC=python, comme le STACK : poursuivre une meilleure intégration des codes.
  - Mais surtout passer en mode grille
    - VO LSST ou pas ? (pas d'obligation stricte à ce stade)
    - GRIF, AuverGrid, MSFG-Open en plus du CC (en mode grille)
  - Entamer la réflexion sur l'accès disque/le catalogue/les codes d'analyse. Et quid des Data DB ?

