

Bilan du workshop

« analyse des données au CC-IN2P3 »

Catherine Biscarat (CC-IN2P3)
pour Ghita Rahal (CC-IN2P3)

Comité de direction LCG-France, mai 2008

PRESENTATION DU WORKSHOP

But : optimiser l'utilisation (et le développement) des infrastructures du CC

Moyen : échange direct entre le CC et les analyseurs de données

Date et place : 17 avril au CC [1]

Participation : 27 inscrits (dont 1/3 du CC); en pratique **> 35 personnes** ; de nombreux **analyseurs**; large fraction des **expériences « qui tournent »**

Chairs : Cristinel Diaconu (chargé de mission pour l'informatique à l'IN2P3) et Ghita Rahal (responsable du groupe de support aux utilisateurs au CC)

Format :

- présentation des ressources du CC
- utilisation du CC par les expériences pour l'analyse
- discussions et bonne interaction entre les utilisateurs et les experts

[1] <http://indico.in2p3.fr/conferenceDisplay.py?confId=722>

AGENDA : PREMIERE PARTIE

- Présentation des ressources offertes par le CC (utilisation optimale)

09:45	Café-croissants	
10:15	Bienvenue (05') Ambition du workshop	Cristinel Diaconu (CPPM)
10:20	Vue d'ensemble des services de calcul et de stockage au Centre de Calcul (35') ( Slides ) Facilities; storages, access to the data, CPU evolution	Fabio Hernandez (CC-IN2P3)
10:55	Outils d'information et de suivi pour l'utilisateur (15') ( Slides ) Comment utiliser le centre de calcul, Recherche de documentation, ressources, utilisation des ressources.	Ghita Rahal (CC-IN2P3)
11:10	Utilisation de la ferme de calcul (20') ( Slides ) Quel est la manière optimale d'utiliser les ressources CPU: possibilités du service batch pour les jobs d'analyse (profil des machines, queues, jobs, prioritization,...)	Suzanne Poulat (CC In2p3)
11:30	Tour de table - discussion (20')	

AGENDA : SECONDE PARTIE

- Parole aux expériences (en cours de prises de données et futures); points forts et points à améliorer dans le cadre de l'analyse des données

11:50	D0 (20') (Slides)	Jan Stark (LPSC Grenoble, CNRS/IN2P3)
12:10	BaBar (20') (Slides)	Georges Vasseur
12:30	Déjeuner	
14:00	Visite de la salle du calcul (30')	Xavier Canehan (CC-IN2P3)
14:30	H1 (20') (Slides)	Emmanuel Sauvan (CPPM)
14:50	ATLAS (20') (Slides)	Jean-Baptiste de Vivie (LAL Orsay)
15:10	CMS (20') (Slides)	Nicola de Filippis (LLR)
15:30	LHCB (20') (Slides)	Marie-Helene Schune (LAL)
15:50	Café	
16:00	ALICE (20') (Slides)	Renaud Vernet
16:20	Tour de table - discussion (1h00')	
17:20	Conclusions (10')	Cristinel Diaconu (CPPM)

REMARQUES INTERMEDIAIRES

- Voix d'analyseurs de caractères différents, présentations **inhomogènes** mais de **grandes lignes** se sont distinguées.
- En général, CC **tres bien perçu** dans son rôle de production programmée à grande échelle (reprocessing de données, production Monte Carlo)
- **Retour mitigé** en ce qui concerne l'analyse (expériences en cours et à venir)
- Spécificités de l'analyse:
 - peu planifiables
 - demande une bonne réactivité
 - beaucoup de petits jobs à la fois
- Expériences LHC :
 - jusqu'à présent, CC considéré dans son rôle de T1 (production)
 - à présent on développe les infrastructures pour analyseurs finaux
 - Toutefois, mêmes observations que celles des expériences en cours
- Bilan tiré de ce workshop:
 - **Limitations actuellement rencontrées**
 - **Besoin de nouveaux services**

i. problèmes de fiabilité et de stabilité : lorsqu'un service s'arrête, l'ensemble de la chaîne de travail est impacté. Certains centres ont des arrêts mensuels programmés et peu d'interruption de service en dehors. Une amélioration de la fiabilité a été observée depuis quelques mois.

- HPSS, dCaChe, AFS, sps
- Travail rendu difficile
- Perception générale du **CC négative**
- **Fuite** vers d'autres centres (CERN, DESY, FNAL)
- « Une *downtime* pendant le week-end est toujours une *downtime* »

ii. lenteurs AFS/dCache : ont poussé certaines expériences à faire leur analyse dans d'autres sites

- AFS : améliorations récentes (ATLAS : duplication de volumes)
- dCache : fichiers introuvables, jobs restent « pendus »

iii. Durée de compilation au CC : la compilation prend plus de temps au CC que dans d'autres centres (CERN, DESY, FNAL).

- Exemple de DØ pour la compilation de packages standards d'analyse
 - Clued0 **11 min** (cluster interactif FNAL)
 - Ccali **27 min** amélioré à **15 min** avec le doublement du cache AFS dans les jours suivants le workshop
- AFS contribue à cette lenteur de compilation mais performances de compilation **non encore optimales** (136%).

iv. site commun T1, T2, T3 : c'est un avantage pour ce qui est de la disponibilité des données mais cela pourrait être un désavantage en ce qui concerne la concurrence production-analyse.

- ATLAS : en 2009/2010 utilisation intensive des ESD (et RAW) pour compréhension du détecteur; en mode stable (> 2010) utilisation de formats plus digestes (distribués sur les T2)

v. monitoring des jobs de batch : manque de visibilité sur ce que font les jobs sur les workers et pas d'accès aux worker nodes par l'utilisateur pour inspection des logs.

- pour examen de jobs malades (hors personnel du CC et production)
- Et aussi pour développement de script batch
- Phénomènes de «lancement en aveugle » et gachis de temps/CPU
- Mise en place d'une interface d'activité pour les utilisateurs ? ssh ?

vi. accès simultané aux données : beaucoup de jobs d'analyse n'ont besoin d'un accès aux données qu'en **début** et en **fin** de job. Il n'en demeure pas moins que la ressource BQS utilisée pour la **régulation** de la mise en exécution des jobs est bloquée pendant **toute la durée du job, empêchant l'entrée** en machine des jobs suivants.

- Attente trop longue en queue
- Avec une sous-utilisation des ressources du CC

Cette partie est un inventaire des besoins qui, à l'heure actuelle, ne sont pas remplis par les infrastructures en place au CC.

i. plateforme de travail interactif : ce besoin se manifeste dans le cadre de l'analyse finale de ntuples (itérations fréquentes sur les fichiers) et aussi en phase de debugging des analyses; par exemple sessions ROOT avec CPU intensif.

- Le physicien a recours au batch pour les actions fastidieuses à répéter, pas pour tout le reste (génération ponctuelle, analyse de 1000 événements, gmake + 20 minutes CPU + ROOT).
- Besoin de **grand volumes de disk** et d'une **grande vitesse d'accès**
- Ni sps tel qu'il existe, ni les ccali ne permettent cet usage
- « Pourquoi lancer des jobs sur la ferme quand je ne peux pas regarder le résultat ? »

ii. ferme d'analyse : il est important de disposer de suffisamment de ressources pour offrir une bonne **réactivité** et un **turnover** rapide des analyses (plusieurs fois par jour).

- **Création de ntuples** (quelques heures, avec dernier filtrage avant analyse interactive)
- Forme un ensemble avec une plateforme interactive
- Balance entre réactivité et taux de remplissage

iii. Evaluation d'une solution basée sur PROOF : structure intégrée à BQS ou bien structure propre ?

- ATLAS se tourne vers cette solution (CMS intéressé)

iv. priorités des jobs d'analyse : les jobs d'analyse ne sont pas soumis en flot constant mais par **paquets** et ont besoin de rentrer **tous rapidement** en machine; les priorités sont de nature différentes de celles de la production.

- Capacité de BQS à absorber rapidement les pics

v. forum d'échange des informations : besoin d'échange d'information à l'intérieur d'un groupe / expérience et entre expériences. **Hypernews** a été cité comme un outil connu de la communauté HEP permettant ces échanges.

- Meilleure information lors des arrêts (programmés ou pas)
- Aide à l'utilisateur par les utilisateurs (amélioration de la réactivité en terme de support)
- Hypernews : archivage, authentification, interface web moderne

REMARQUES FINALES

- La liste des points mentionnés ici représente les points forts mise en évidence parmi des **avis divers**.
- Cette liste, telle qu'elle existe, va être examinée par le CC; le CC prévoit la mise en place de « **groupe de travail** » par point; leur charge sera de **converger** vers une solution convenable aux expériences (**consultation préalable**).
- Le CC a suggéré aux expériences de s'accorder si possible sur la mise en œuvre d'une infrastructure commune pour les analyses (PROOF a été cité).

BACK-UP SLIDES