Entrevue Oracle – Centre de Calcul de l'IN2P3

Le batch au CC-IN2P3

Le batch au CC-IN2P3

- Historique de BQS
- Chiffres et mots clés de BQS
- Migration BQS → GE

Historique (1/2)

- BQS = Batch Queuing System

  Développement interne, initié en 1991, lors de la migration VM -> Unix.
  Base fonctionnelle de BMON (VM)
  Développement = 1 FTE

- 2001 : Introduction d'une base de données relationnelles (MySql/InnoDB),

  Montée en puissance du calcul au CC et donc de l'utilisation du batch
  Développement = 1.5 FTE
  En 2002 : 300 machines (=600 cores), 600 jobs simultanés, 3000 jobs /jour

- 2004 : Amplification des développements

  Pour répondre aux challenges de scalabilité et robustesse, d'ouverture à la grille, d'évolution fonctionnelle (ex les jobs parallèles), d'évolution logicielle de BQS
  Développement = 2.2 FTE

Historique (2/2)

- 2008 : Pause dans les développements

  Stabilité fonctionnelle

  Stablité dans les évolutions logicielles (ex : L'intégralité de la configuration ainsi que les jobs sont stockés en base de données)

  Mise en oeuvre d'une **revue approfondie** de notre système de batch (2008~2009)

- 2009 : Décision de migrer de BQS vers une solution déjà utilisée dans la communauté HEP

  Lancement d'une étude des divers sytèmes de batch

Chiffres et mots clés (1/2)

- BQS est actuellement déployé sur 2 fermes de calcul :
  Une ferme de jobs séquentiels : 1300 machines (10,000 cores), 10,000 jobs simultanés, +100,000 jobs /j
  Une ferme de jobs parallèles : ~64 machines (512 cores), avec réseau 10 Gige

- Chiffre d'évolution de 2004 à 2010 en nombres de machines, cores et volume de jobs et pour la ferme anastasie (jobs séquentiels)

	march 04	march 05	march 06	oct. 06	dec. 07	oct. 08	may 09	july 10
Numbers of machines	300	650	720	850	1000	1100	1300	**1300**
Numbers of cores	600	1300	1500	2250	4500	8000	9500	**10000**
Simultaneous jobs	1000	2000	2700	3800	5000	8500	9000	**10000**
Flow of jobs, per day	5000	10000	15000	23000	35000	50000	70000	**110000**

Chiffres et **mots clés** (2/2)

- Scalabilité, robustesse et efficacité du repository d'information (indispensable au fonctionnement des CEs)

- Outils pertinents pour répondre à l'exploitation quotidienne

  Régulation des accès aux services de stockage. Outils de détection et de prise d'action pour les jobs et machines dit 'pathologiques'. Nombreux paramètres de régulation de la charge du système

- Mutualisation

  Un même ferme pour servir ~60 groupes que ce soit des utilisateurs locaux ou via la grille de calcul

- Fonctionnalités

  Fairshare sur groupes et rôles intra-groupe

  Gestion des tokens AFS

  Gestion des jobs paralleles

  Delégation de gestion de sous –ensembles (share intra-groupe, ressources logiques) au responsible de groupe

  Gestion des limites en CPU, mémoire et espace disque

  Expression unifiée du CPU quelque soit la puissance de la machine

  …

Migration BQS → GE (1/2)

- Les étapes

  Décision de migration en juin 2009

  Investigation des produits utilisés dans la communauté HEP et choix de la solution GE en février 2010

  Investigation technique de GE au cours du 2eme trimestre 2010

- La méthode (phase 1)

  Recensement des caractéristiques d'un système de batch et avec pour trame une expérience de 20 ans en exploitation dubatch

  Constitution d'une grille de 60 critères regroupés sous 15 thémes majeurs

  Enquète (pageWEB) sur les retours d'expérience

  L'étude s'est rapidement concentrés sur 4 produits : LSF, SGE, PBS-Pro, TORQUE/Maui

- La méthode (phase 2)

  Prise en main des 2 "meilleures" solutions (1 FTE pendant 10 jours sur chaque système)

  Adoption de GE au regard de critères techniques mais aussi avec l'idée d'être 'moteur' dans la communauté HEP

Migration BQS → GE (2/2)

- La situation actuelle
  Exploration fonctionnelle de GE (version Open Source, 6.2u5)
  Mise en oeuvre d'une ferme de test
  Mise en oeuvre de sessions de formation interne

- La suite :
  Une ferme de pré-prod (limitée à quelques utilisateurs) pour le 4eme trimestre 2010
  Mise en production prévue au 1er trimestre 2011
  Arrêt de BQS au plus tard fin 2011

Conclusion

- Domaines possibles de collaboration:
  - Capacité de développement dans nos domaines d'expertise (Interfaçages File-systems (AFS), Grilles, ..)
  - Développement d'une communauté d'utilisateurs dans le domaine HEP