Stockage distribué @ LCPQ

Capitoul – 23 juin 2016

David Sanchez
<david.sanchez@irsamc.ups-tlse.fr>

Laboratoire de Chimie et Physique Quantiques

L'infra en 2014

- **4 clusters sous Rocks Clusters, en v5 et v6.**
- **Dont un cluster qui dispose d'un stockage lustre sur InfiniBand.**
- **Les autres n'ont que du NFS entre le maître et les nœuds de calcul.**

Sur le cluster Lustre

- **Compliqué à mettre en place à l'époque de l'installation du cluster l'utilisant.**
- **Upgrade très compliqué (kernel lustre nécessaire sur tous les nœuds clients).**
- **Lustre version 2009 : instable.**
- **Module noyau, kernel panic, reboot oss.**

- **Et si on regardait ailleurs ?**

Critères de sélection

- **Sous GNU/Linux.**
- **Non lié au kernel.**
- **Utiliser divers protocoles réseaux : ethernet, InfiniBand.**
- **Performances.**
- **Stabilité.**
- **Facilité d'administration.**

Le choix : BeeGFS

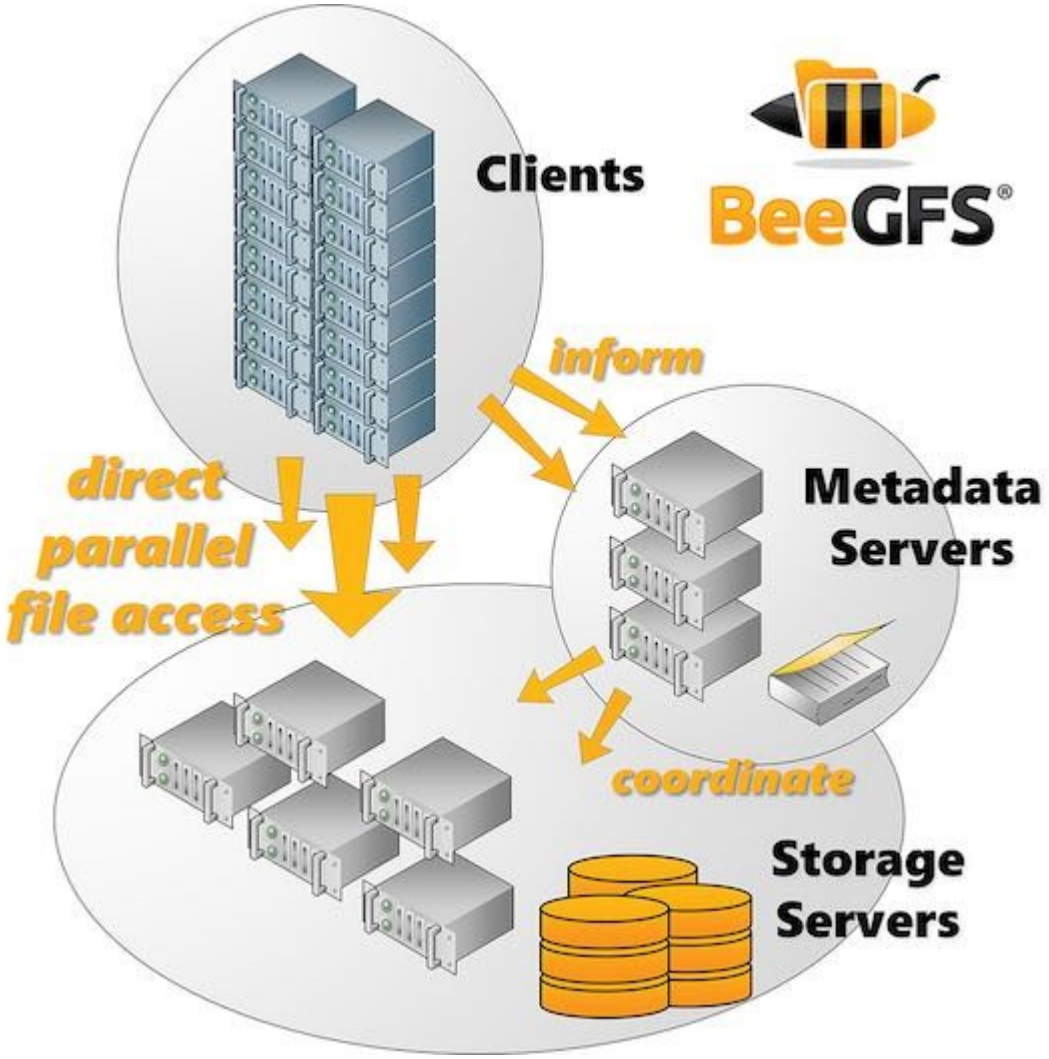
- **Ex-FhGFS.**
- **Système de fichiers conçu pour le calcul distribué.**
- **Découpage des fichiers pour stockage sur multiples serveurs de stockage.**
- **Utilisation des disques locaux des nœuds de calcul (ex : Strasbourg).**
- **Certifié pour Intel Omni-Path.**

Architecture d'une solution BeeGFS

- **beegfs-mgmtd : démon de gestion du cluster de stockage (sur le serveur maître).**
- **beegfs-storage : démon pour le stockage des données (sur les oss).**
- **beegfs-meta : démon pour la gestion des métadonnées (sur le mds).**
- **beegfs-client : démon pour monter le système de fichiers sur les nœuds de calcul.**
- **beegfs-helperd : démon pour la partie espace utilisateur du client (sur nœuds de calcul).**

Schéma architecture

- **http://insidehpc.com/2016/06/beegfs-omni-path-certification-12gbs-per-server/**



Performances – parallèle

- **Sur 3 nœuds en ethernet, 3 jobs par nœud :**
  - 259 Mo/s de moyenne par job.
- **Débit cumulé : 2,2Go/s, saturation des interfaces 10GbE des OSS.**

Fin

Merci !