

Projet de stage

Sujet : Architecture d'accès à des ressources de Calcul et de Données (mots clés : Cluster, Grille, Calcul et données répartis, Workflow) : étude et mise en œuvre.

Durée : 6 mois à partir de début mars 2006
Laboratoire : CDS, Observatoire astronomique de Strasbourg
Tuteur : André Schaaff (schaaff@astro.u-strasbg.fr)

Contexte du stage :

Le *Centre de Données astronomiques de Strasbourg* (CDS) développe des services de référence à forte valeur ajoutée, très utilisés par les astronomes du monde entier. Placé sous la responsabilité de l'Institut National des Sciences de l'Univers et de l'Université Louis Pasteur, il fait partie de l'Observatoire astronomique (UMR 7550). Il joue un rôle moteur dans la mise en place de liens entre les services d'information en astronomie et dans le développement de standards, en partenariat avec les autres producteurs de services, en particulier avec les grands observatoires et les journaux. Il participe très activement aux projets d' "Observatoire Astronomique Virtuel" actuellement proposés en Europe, aux Etats-Unis, en Grande Bretagne, ...

Description du sujet :

Parmi ses différentes ressources matérielles le CDS dispose d'un cluster (une dizaine de machines) mixte de Calcul (essentiellement utilisé par le service Aladin) et d'accès à des catalogues (service Vizier) de données astronomiques. Ce cluster est actuellement accédé via un outil propriétaire développé en interne. Cet outil permet de gérer le « loadbalancing », il est assez sommaire mais efficace en terme de temps de réponse.

A l'occasion de l'arrivée d'un nouveau cluster de 8 machines biprocesseurs 64bits, nous souhaitons revoir globalement l'architecture et les outils associés (éclatement en deux clusters « Calcul » et « Données », ...).

Nous collaborons avec des collègues travaillant dans le cadre des Grilles et nous souhaitons intégrer cela dans la nouvelle architecture.

Le stage débutera par un état de l'art portant sur les outils de gestion de clusters et par un apprentissage des concepts liés aux grilles.

Le stagiaire proposera une architecture (outils de gestion du (des) cluster(s), intergiciel de soumission de tâches, etc.) permettant de gérer le cluster de Calcul et d'intégrer l'accès à des grilles (Grid5000, EGEE). Des tâches de calcul pourront par exemple être traitées localement sur le cluster de Calcul ou sous-traitées à une grille.

Le stagiaire étudiera également l'intégration du cluster de données dans cette architecture. Plus généralement, la soumission de chaînes de traitements (Workflow) sera également abordée. Les services du CDS fonctionnent « en production » et les aspects disponibilité, fiabilité, facilité de maintenance, extensibilité,..., sont primordiaux.

Cheminement souhaité : état de l'art, implémentation d'un prototype (avec quelques machines), évaluation, développements, fiabilisation d'une architecture opérationnelle.

Lien : <http://cdsweb.u-strasbg.fr>