

Analyse et ré-exécution de jobs sur un cluster

Fabien Hermenier `fabien.hermenier@unice.fr`

Fabrice Huet `fabrice.huet@unice.fr`

11 février 2014

Nombre d'étudiants souhaités : 3-4

Description du sujet

Les plateformes de calculs distribués sont souvent divisées en deux catégories. La première, dite de production permet aux utilisateurs de lancer des calculs dont le but est la production de résultats, comme par exemple prévoir la météo, optimiser la forme d'une coque de bateau... La seconde, dite d'expérimentation sert à développer et tester de nouveaux algorithmes, protocoles, logiciels...

Cette différence a un impact sur la façon dont les machines sont utilisées. Dans le cas des expérimentations les utilisateurs utilisent beaucoup de noeuds pendant une durée relativement courte (1-2h) alors que pour la production, les durées de calcul sont conséquentes (12-24h ou plus).

Une question récurrente est de savoir si une unique plateforme de type Cloud permettrait de satisfaire les différents types d'utilisateurs. En pratique, un utilisateur réserverait des noeuds qui seraient automatiquement installés dans la configuration souhaitée (production ou expérimentale) avant qu'il puisse les utiliser. Une façon de mesurer l'impact sur les utilisateurs est de rejouer leurs réservations sur un Cloud virtuel simulant l'unique plateforme et vérifier les conséquences (manque de ressources, trop d'attente...). Pour avoir des résultats crédibles, nous utiliserons les logs de toutes les réservations effectuées sur deux clusters d'INRIA Sophia Antipolis.

Travail demandé

1. À partir des logs des deux plateformes qui sont dans des formats différents, proposer une représentation commune permettant l'analyse
2. Analyser les logs pour fournir des informations quantitatives (nombre de jobs/an, évolution des demandes en fonction de la période...)
3. Implémenter un Cloud virtuel permettant de ré-exécuter les calculs des utilisateurs

Lieu

INRIA Sophia Antipolis

Prérequis

Informations complémentaires