



Conclusion et exemple

Exercices pour le Module 12
INF8480 Systèmes répartis et infonuagique

Michel Dagenais

École Polytechnique de Montréal
Département de génie informatique et génie logiciel

Google Build System

Un usager peut tester ses modifications récentes à un gros logiciel en téléchargeant l'ensemble du code source, le compilant et l'exécutant sur son poste de travail. Le Google Build System donne l'illusion de ce comportement à l'utilisateur mais en fait la compilation est faite sur les serveurs. Quels sont les gains d'efficacité que permet le Google Build System?



Google Build System

Premièrement, ce système évite de copier les fichiers entre les serveurs et la station de l'utilisateur, tout se fait sur les serveurs. Deuxièmement, les fichiers sont partagés entre tous les développeurs, de sorte qu'un fichier déjà compilé avec les bonnes options peut être fourni directement à un second utilisateur au lieu d'effectuer la même compilation une deuxième fois. Troisièmement, le fichier de sortie est vérifié avec la version antérieure, s'il n'y a pas de changement, il n'y a pas besoin de faire d'opérations supplémentaires; par exemple, si on modifie les commentaires dans un fichier de programme, le code compilé ne sera pas affecté et il n'y a pas besoin de refaire les tests ou l'édition de liens. Finalement, la compilation peut se faire en parallèle sur un grand nombre de serveurs, ce qui réduit les délais d'attente.



Google Wide Profiling

Le système Google Wide Profiling permet avec un surcoût très faible de mesurer et de comparer la performance de différents matériels et de différents modules et versions de logiciels. Comment cela est-il possible?



Google Wide Profiling

Une approche par profilage permet d'ajuster la fréquence d'échantillonnage de manière à avoir un surcoût qui peut être aussi faible que voulu. Toutefois, si le nombre d'échantillons est très faible, la précision diminue. Dans le cas des systèmes Google, le nombre d'ordinateurs et de requêtes est gigantesque et il est donc possible de prendre l'information sur une très petite proportion des requêtes et néanmoins d'avoir beaucoup d'échantillons. Etant donné le très grand nombre de requêtes, on peut comparer la performance obtenue sur différents ordinateurs et sur différentes versions des logiciels. L'hypothèse sous-jacente est que toute différence sur la moyenne d'un grand nombre de requêtes sera causée par la différence de matériel ou de logiciel et non pas par les spécificités de chaque requête.

