

WACA: Politique de répartition de charge des services web dans une architecture de type Cloud

Sylvain Lefebvre*, Sathya Prabhu Kumar *, Raja Chiky *

*LISITE - ISEP 28, rue notre des Champs, 75006, PARIS
prénom.nom@isep.fr,
<http://www.isep.fr>

Résumé. De plus en plus d'applications font appel à des services web déployés sur des architectures de type cloud. Ces services peuvent traiter des données massives et des fichiers en grandes quantités pour répondre à un nombre croissant de demandes (requêtes) des utilisateurs. Ces données peuvent être de différents types allant des données multimédia aux données numériques provenant de capteurs. En plus de la volumétrie des données, les applications web doivent également gérer la satisfaction des utilisateurs ce qui impose une garantie de la qualité du service fourni.

Pour ces raisons, la répartition des traitements sur une grappe de machines devient nécessaire pour un grand nombre d'applications. L'efficacité de ces applications dépendra particulièrement de la politique de répartition de charge utilisée.

Cet article présente une nouvelle politique de répartition de charge appelée WACA (Workload And Cache Aware load balancing) qui prend en compte la charge des machines et le contenu de leur cache. Cette politique permet de réduire le temps d'exécution des requêtes des utilisateur en se basant sur des résumés de cache construits sous forme de filtre de Bloom. WACA est un algorithme de répartition de charge centralisé qui affecte les requêtes vers la machine susceptible de contenir les données visées dans son cache en évitant de la surcharger. La politique proposée a été développée et évaluée en utilisant un framework de répartition des applications web appelé CLOUDIZER et la stratégie a été comparée à des algorithmes de répartition de charge traditionnels.