



Mardi 27 Juin 2017

10h00

UT3 Paul Sabatier, IRIT, Salle des Thèses

Uras TOS

Equipe PYRAMIDE - IRIT

Réplication de données dans les systèmes de gestion de données à grande échelle

Jury :

Mme. Esther PACITTI – *Professeuse à l'Université Montpellier II, France (Rapporteur)*

M. Ali YAZICI – *Professeur à l'Université Atilim, Turquie (Rapporteur)*

M. Patrick VALDURIEZ – *Directeur de Recherche à l'INRIA, France (Examineur)*

M. Abdelkader HAMEURLAIN – *Professeur à l'Université Paul Sabatier- Toulouse III, France (Directeur de Thèse)*

M. Tolga AYAV – *Professeur Assistant à l'Institut des Technologies d'Izmir, Turquie (co-Directeur de Thèse)*

M. Riad MOKADEM – *Maître de Conférences à l'Université Paul Sabatier- Toulouse III, France (Encadrant)*

Mots-clés : Systèmes cloud, requêtes de base de données, réplication de données, évaluation de performances, profit économique

Résumé : Ces dernières années, la popularité croissante des applications, e.g. les expériences scientifiques, Internet des objets et les réseaux sociaux, a conduit à la génération de gros volumes de données. La gestion de telles données qui de plus, sont hétérogènes et distribuées à grande échelle, constitue un défi important.

Dans les systèmes traditionnels tels que les systèmes distribués et parallèles, les systèmes pair-à-pair et les systèmes de grille, répondre à des objectifs tels que l'obtention de performances acceptables tout en garantissant une bonne disponibilité de données constituent des objectifs majeurs pour l'utilisateur, en particulier lorsque ces données sont réparties à travers le monde. Dans ce contexte, la réplication de données, une technique très connue, permet notamment: (i) d'augmenter la disponibilité de données, (ii) de réduire les coûts d'accès aux données et (iii) d'assurer une meilleure tolérance aux pannes. Néanmoins, répliquer les données sur tous les noeuds est une solution non réaliste vu qu'elle génère une consommation importante de la bande passante en plus de l'espace limité de stockage. Définir des stratégies de réplication constitue la solution à apporter à ces problématiques.

Les stratégies de réplication de données qui ont été proposées pour les systèmes traditionnels cités précédemment ont pour objectif l'amélioration des performances pour l'utilisateur. Elles sont difficiles à adapter dans les systèmes de cloud. En effet, le fournisseur de cloud a pour but de générer un profit en plus de répondre aux exigences des locataires. Satisfaire les attentes de ces locataire en matière de performance sans sacrifier le profit du fournisseur d'un coté et la gestion élastiques des ressources avec une tarification suivant le modèle 'pay-as-you-go' d'un autre coté, constituent des principes fondamentaux dans les systèmes cloud.

Dans cette thèse, nous proposons une stratégie de réplication de données pour satisfaire les exigences du locataire, e.g. les performances, tout en garantissant le profit économique

05 61 55 65 10
info@irit.fr

www.irit.fr



du fournisseur. En se basant sur un modèle de coût, nous estimons le temps de réponse nécessaire pour l'exécution d'une requête distribuée. La réplique de données n'est envisagée que si le temps de réponse estimé dépasse un seuil fixé auparavant dans le contrat établi entre le fournisseur et le client. Ensuite, cette réplique doit être profitable du point de vue économique pour le fournisseur. Dans ce contexte, nous proposons un modèle économique prenant en compte aussi bien les dépenses et les revenus du fournisseur lors de l'exécution de cette requête. Nous proposons une heuristique pour le placement des répliques afin de réduire les temps d'accès à ces nouvelles répliques. De plus, un ajustement du nombre de répliques est adopté afin de permettre une gestion élastique des ressources.

Nous validons la stratégie proposée par une évaluation basée sur une simulation. Nous comparons les performances de notre stratégie à celles d'une autre stratégie de réplique proposée dans les clouds. L'analyse des résultats obtenus a montré que les deux stratégies comparées répondent à l'objectif de performances pour le locataire. Néanmoins, une réplique de données n'est créée, avec notre stratégie, que si cette réplique est profitable pour le fournisseur.

