

Chères et chers collègues,

Bonjour,

J'ai le plaisir de vous inviter à la soutenance de mon Habilitation à Diriger des Recherches (HDR) intitulée : "**Contribution à la réplication de données dans les systèmes de gestion de données à grande échelle**".

La soutenance aura lieu le Mardi 17 Novembre 2020 à 10h à l'Université Paul Sabatier Toulouse 3, à l'Institut de Recherche en Informatique de Toulouse IRIT (salle des thèses).

La composition du jury de mon HDR est la suivante :

Esther PACITTI, Professeure, Université de Montpellier, Rapportrice

Mohand-Said HACID, Professeur, Université Claude Bernard, Lyon 1, Rapporteur

Farouk TOUMANI, Professeur, Université Blaise Pascal - Clermont-Ferrand 2, Rapporteur

Claudia RONCANCIO, Professeure, Institut Polytechnique de Grenoble - ENSIMAG,

Examinatrice

Franck MORVAN, Professeur, Université Paul Sabatier, Toulouse 3, Examineur

Jean-Marc PIERSON, Professeur, Université Paul Sabatier, Toulouse 3, Examineur

Patrick VALDURIEZ, Professeur, Directeur de recherche, Inria, LIRMM Montpellier,

Examineur

Abdelkader HAMEURLAIN, Professeur, Université Paul Sabatier, Toulouse 3, Directeur de recherche et Parrain de l'HDR

Résumé :

Dans nos travaux de recherche, nous nous intéressons à la conception de stratégies dynamiques de réplication de données, visant l'amélioration des performances dans les systèmes de gestion de données à grande échelle pour des applications décisionnelles.

L'étude de l'état de l'art nous a permis de proposer une classification des stratégies de réplication de données dans les systèmes de grille de données, par exemple en fonction du type d'architecture de la grille, et une autre classification dans les systèmes Cloud, par exemple en fonction de la prise en compte de la rentabilité du fournisseur. Nous avons constaté que seulement quelques stratégies de réplication dans les systèmes Cloud s'intéressent à la satisfaction de l'objectif des performances pour les locataires tout en garantissant un bénéfice économique pour le fournisseur. De plus, la plupart de ces stratégies visent seulement à réduire la consommation des ressources générée par la réplication sans se focaliser sur le coût économique de cette réplication.

Dans cette perspective, nos travaux de recherche visent la conception de stratégies dynamiques de réplication de données prenant en compte le compromis entre la garantie de performances pour les locataires et la rentabilité économique pour le fournisseur. D'abord, la création d'une réplique est envisagée uniquement si l'objectif de temps de réponse n'est pas atteint pour le locataire. Ensuite, une heuristique est proposée afin de trouver un placement acceptable pour cette réplique, qui satisfait cet objectif tout en étant profitable pour le fournisseur. Pour cela, nous avons étendu les modèles de coûts classiques en intégrant les coûts économiques, notamment le coût engendré par la réplication. Enfin, un algorithme d'ajustement dynamique du nombre de répliques permet au fournisseur de s'adapter à la demande des locataires.

Les performances de nos propositions ont été comparées à celles d'autres stratégies de réplication de données proposées dans les systèmes Cloud. L'analyse des résultats a montré que nos propositions ne se contentent pas seulement de satisfaire les objectifs des locataires. Elles prennent aussi en compte le profit du fournisseur, notamment dans des conditions

défavorables, par exemple des taux importants d'arrivée de requêtes et des seuils de temps de réponse stricts.

Mots clés. Gestion de données, bases de données, réplication de données, grille de données, Cloud Computing, modèles de coûts, performances, élasticité, profit économique.

Abstract

We present a summary of our research activities, focusing on the design of dynamic data replication strategies that aim to improve performance in large-scale data management systems for OLAP applications.

The study of the state of the art allowed us to propose a classification of data replication strategies in data grid systems, e.g., according to the type of grid architecture for which a strategy is designed, and another classification in cloud systems, e.g., based on the consideration of the cloud provider's economic profit. Only some replication strategies in cloud systems focus on meeting performance objectives for tenants while ensuring an economic benefit for the provider. In addition, most of these strategies aim only to reduce the consumption of resources generated by replication without focusing on the economic cost of this replication.

In this perspective, we aim to design data replication strategies that take into account the compromise between performance satisfaction for tenants and economic profitability for the cloud provider. First, a new replica creation is considered only if the tenant's response time (RT) objective is not reached. Then, a heuristic is proposed in order to find an acceptable placement for this replica, which satisfies this objective while being profitable for the provider. For this aim, we have extended the classic cost models by integrating economic costs, in particular the costs generated by replication. Finally, a dynamic replica factor adjustment algorithm allows the provider to adapt to tenant demand.

We compared the performance of our proposals alongside other data replication strategies proposed in cloud systems. Analysis of the results shows that our proposals do not only satisfy the tenants' objectives. They also ensure the provider's profit, especially under unfavourable conditions such as high query arrival rates and strict RT thresholds.

Keywords. Data management, databases, data replication, data grids, cloud computing, cost models, performance, elasticity, economic profit.

Bien à vous,

Riad Mokadem, PhD, Associate Professor
PYRAMID team - IRIT lab.
Paul Sabatier University,
118 route de Narbonne, 31062 Toulouse, France
Phone: +33 (0)561 557 442
[E-mail:mokadem@irit.fr](mailto:mokadem@irit.fr)