

D-WorM : Middleware pour le traitement des requêtes massives dans les entrepôts de données massives

Rado Ratsimbazafy*, Omar Boussaid*, Fadila Bentayeb*

*Laboratoire ERIC Université Lumière Lyon 2,
5 Avenue Pierre Mendès France, 69676 Bron Cedex
prenom.nom@univ-lyon2.fr, <http://www.eric.univ-lyon2.fr>

Résumé. Avec l'accroissement massif des données, plusieurs problèmes connus des entrepôts de données sont à repenser à l'échelle du *Big Data* et des environnements distribués. Dans cet article nous proposons D-WorM (Data warehouse Workload Manager), une solution pour résoudre la problématique posée par le traitement des requêtes massives. Notre Middleware a pour objectif d'améliorer la gestion des montées en charge.

1 Introduction

Les entrepôts de données sont au cœur des systèmes d'informations décisionnels. Ils sont constamment interrogés par plusieurs utilisateurs avec des requêtes décisionnelles à des fins d'analyse pour la prise de décision (Inmon, 1992). Il n'est pas rare qu'ils subissent des montées en charge (requêtes massives ou *workloads* en anglais). Avec l'avènement du *Big Data* la taille des entrepôts de données devient de plus en plus grande. Le développement de nouvelles technologies type *cloud computing* nous amène à repenser les mécanismes de stockage et de traitement des très grands entrepôts de données.

Le traitement des requêtes massives n'est pas un problème récent. En effet, de nombreux chercheurs s'y sont déjà intéressés, que ce soit pour les bases de données (Sellis, 1988) ou pour les entrepôts de données (Gacem et Boukhalfa, 2013).

Des travaux de recherche récents ont été menés pour résoudre le problème, à l'échelle du *Big Data*. Des solutions ont alors été proposées sous forme de boîtes noires qui regroupent tout un écosystème qui ne laisse pas beaucoup de liberté à l'utilisateur pour personnaliser l'infrastructure selon ses besoins (Bajda-Pawlikowski et al., 2011; Giannikis et al., 2012).

Le prototype D-WorM que nous proposons dans cet article, a pour objectif de fournir une couche applicative capable de traiter les montées en charge pour réduire le temps de réponse moyen en mutualisant certaines tâches, en partageant certaines données et en utilisant des vues. L'objet de ce papier est de présenter le fonctionnement de notre solution à partir de plusieurs charges de requêtes de tailles différentes.

Cet article est organisé comme suit. Le principe de fonctionnement de D-WorM est expliqué dans la section 2. Les choix technologiques pour l'implémentation et le scénario de démonstration sont détaillés dans la section 3. Dans la section 4 nous concluons et discutons des perspectives de recherche à venir.