

La Fragmentation Horizontale Revisitée: Prise en Compte de l'Interaction de Requêtes

Amira Kerkad*, Ladjel Bellatreche*, Dominique Geniet*

*LIAS/ISAE-ENSMA, Futuroscope, Poitiers, France
(amira.kerkad, bellatreche, dominique.geniet)@ensma.fr

Résumé. La principale caractéristique des requêtes définies sur un entrepôt de données relationnel (\mathcal{EDR}) est le fait que leurs jointures passent systématiquement par la table des faits. Cette situation favorise l'interaction entre les requêtes. Cette interaction a été largement exploitée par les algorithmes d'optimisation de requêtes dans les bases de données traditionnelles, connue sous le nom d'optimisation multi-requêtes. Dans les \mathcal{EDR} , cette interaction a été utilisée pour définir des méthodes de sélection des vues matérialisées. Dans cet article, nous revisitons le problème de sélection de schéma de fragmentation horizontale. Après un état de l'art, où nous soulignons le fait que les algorithmes existants ignorent l'interaction entre les requêtes, nous proposons un nouvel algorithme basé sur cette interaction. Sa principale caractéristique est l'utilisation d'une structure de données incrémentale considérant l'interaction. Il utilise le principe d'élection des requêtes pour aiguiller le processus de fragmentation et propager le gain au mieux sur l'ensemble des requêtes de la charge. Finalement, une étude expérimentale est conduite pour prouver l'efficacité de notre approche.

1 Introduction

Comme pour tout développement de projet informatique, la conception des applications consommatrices des données (bases de données, entreposage de données, etc.) passe par un cycle de vie. Dans la première génération de bases de données, ce cycle contenait quatre étapes principales : (1) la collecte des besoins, (2) la modélisation conceptuelle, (3) la modélisation logique et (4) la modélisation physique. La dernière phase a connu une importance particulière avec l'apparition des applications décisionnelles. Ces dernières manipulent une masse importante de données, accédée par des requêtes complexes exigeant un temps réponse rapide pour satisfaire les décideurs. Plus la compétitivité entre entreprises augmente plus cette exigence est primordiale. Avec l'évolution des réseaux et des systèmes de stockage et de traitement, le cycle de vie a été étendu pour prendre en considération la phase de déploiement. Cette dernière inclut les architectures matérielles sur lesquelles les données sont déployées (systèmes distribués, systèmes parallèles, clusters, grille de calcul, informatique dans les nuages, etc.).

La fragmentation horizontale (\mathcal{FH}) a la particularité d'être utilisée au niveau logique, physique et déploiement. Au niveau logique, elle a été largement considérée comme une technique de conception de bases de données centralisées et réparties Karlapalem (1996). Dans les entrepôts de données, elle est positionnée comme une technique d'optimisation. Contrairement aux