

La fragmentation dans les entrepôts de données : une approche basée sur les algorithmes génétiques

Ladjel Bellatreche*, Boukhalfa Kamel **

* LISI/ENSMA
Futuroscope - France
bellatre@ensma.fr

** Université de Laghouat Route de Ghardaia, BP 37G
Laghouat Algérie
k.boukhalfa@mail.lagh-univ.dz

Résumé. La fragmentation horizontale est une technique d'optimisation non redondante de requêtes décisionnelles de type ROLAP. L'utilisation de cette technique dans les entrepôts de données représente un enjeu plus important que dans un contexte de bases de données traditionnelles. Cette importance est due au différents choix des tables (de dimensions ou des faits) à fragmenter. Dans le contexte des entrepôts, la fragmentation n'a un sens que si la table des faits est partitionnée en fonction des schémas de fragmentation des tables de dimensions. Mais ce type de fragmentation de la table des faits pourrait engendrer un nombre important de fragments qui rendrait le processus de maintenance très coûteux. Afin de réduire ce nombre ou le rendre contrôlable par l'administrateur de l'entrepôt, nous proposons l'utilisation d'un algorithme génétique. Ce dernier a pour but de sélectionner les tables de dimension à fragmenter pour (1) éviter l'explosion du nombre de fragments de la table des faits et (2) garantir une meilleure performance d'exécution des requêtes. Notre algorithme génétique est développé sous visual C et validé par une étude expérimentale en utilisant le banc d'essai APB-1 release II.

1 Introduction

Les principales caractéristiques des entrepôts de données sont leur grande taille et la complexité des requêtes décisionnelles dues aux opérations de jointure et d'agrégation. Plusieurs techniques d'optimisation ont été proposées pour réduire le coût d'exécution de requêtes. Ces techniques peuvent être divisées en deux catégories : *structures redondantes* et *structures non redondantes*. Les structures redondantes nécessitent un espace de stockage et un coût de mises à jour. Parmi ces dernières, on peut citer les vues matérialisées et les index [Kotidis et Roussopoulos, 1999, Gupta, 1999]. La deuxième catégorie ne nécessite pas de coût de stockage. C'est le cas des techniques de fragmentation (horizontale et verticale) [Sanjay *et al.*, 2004]. La plupart des systèmes commerciaux comme ORACLE, DB2 proposent les deux structures [Zilio *et al.*, 2004, Sanjay *et al.*, 2004]. Dans ce papier nous nous intéressons à une structure non redondante, à savoir la fragmentation horizontale. Plusieurs travaux de recherche et des produits commerciaux ont montré son utilité dans le processus d'optimisation des requêtes