

Fragmentation Primaire et Dérivée: Étude de Complexité, Algorithmes de Sélection et Validation sous ORACLE10g

Kamel Boukhalfa *, Ladjel Bellatreche* Pascal Richard *

* LISI/ENSMA - Université de Poitiers
Futuroscope 86960 France
(boukhalk, bellatreche, richardp)@ensma.fr

Résumé. La fragmentation horizontale a été largement adoptée par la communauté des bases de données. Elle a une place à part entière dans la conception physique. Plusieurs systèmes de gestion de bases de données (SGBD) commerciaux ont proposé un langage de définition de données pour partitionner des tables relationnelles en utilisant différents modes. Dans ce papier, nous présentons d'abord l'évolution de la fragmentation ces dernières années au sein des SGBDs. Deuxièmement, nous étudions le problème de sélection de schéma de fragmentation d'un entrepôt de données relationnel, et nous montrons qu'il est NP-complet. Vu sa complexité, nous développons un algorithme de hill climbing (méthode de voisinage) pour sélectionner un schéma de fragmentation quasi optimal. Nous effectuons des expérimentations afin de comparer cet algorithme avec deux autres algorithmes: un génétique et un recuit simulé en utilisant un modèle de coût mathématique. Finalement, nous effectuons une validation réelle de nos algorithmes sous ORACLE10g en utilisant les données issues du banc d'essai APB1.

1 Introduction

Les entrepôts de données sont connus par leur volumétrie et requêtes complexes caractérisées par des jointures, sélections et agrégations. Pour optimiser ces opérations et faciliter la gestion de ces données, la fragmentation horizontale est devenue un candidat sérieux (Papadomanolakis et Ailamaki, 2004; Sanjay et al., 2004). Elle permet de partitionner les tables, les vues matérialisées ou les index en plusieurs ensembles disjoints de tuples stockés physiquement et généralement accédés séparément. Une caractéristique intéressante liée à la fragmentation horizontale est le fait qu'elle ne duplique pas les données, et par conséquent, elle réduit les coûts de stockage et de maintenance.

Deux versions de la fragmentation horizontale ont été définies (Özsu et Valduriez, 1999) : primaire et dérivée. La fragmentation horizontale primaire d'une relation est effectuée en utilisant les attributs définis sur cette table. La fragmentation horizontale dérivée consiste à partitionner une table en utilisant les attributs définis sur une ou plusieurs autres(s) table(s). En d'autres termes, la fragmentation horizontale dérivée d'une table est basée sur le schéma de