

Une Sélection Multiple des Structures d'Optimisation Dirigée par la Méthode de Classification K-means

Rym Bouchakri *, Ladjel Bellatreche **, Kamel Boukhalfa***

* Ecole nationale Supérieure d'Informatique, Oued-Smar Alger - Algérie
r_bouchakri@esi.dz

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

*** USTHB, Alger, Algérie
boukhalk@ensma.fr

Résumé. Le volume d'information contenu dans un entrepôt de données s'accroît sans cesse, augmentant de ce fait le temps d'exécution des requêtes décisionnelles. Pour y remédier, l'administrateur doit, durant la phase de conception physique de l'entrepôt, effectuer une sélection de structures d'optimisation (index, vues matérialisées ou fragmentation), puis assurer leur gestion et maintenance. Pour optimiser un nombre maximum de requêtes, il est indispensable d'opter pour une sélection multiple de structures ayant une forte similarité. Dans la littérature, deux principales similarités entre les structures d'optimisation ont été identifiées : une entre les vues et les index et l'autre entre la fragmentation horizontale dérivée et les index de jointure binaire. Dans ce travail, nous proposons une approche de sélection multiple des index de jointure binaire et de fragmentation. Vue la complexité de la sélection multiple, nous proposons une nouvelle approche permettant d'abord de partager l'ensemble des attributs extraits des requêtes entre les deux structures, ensuite sélectionner chaque structure avec un algorithme. Pour réaliser ce partage, nous proposons d'utiliser la méthode K-means. Une étude expérimentale et des tests comparatifs sur un entrepôt de données réel sous le SGBD Oracle 11g sont proposés illustrant l'intérêt de notre approche.

1 Introduction

Dans les bases de données traditionnelles, la tâche d'un administrateur était principalement concentrée sur la gestion des utilisateurs et d'un nombre restreint de structures d'optimisation comme les index mono table ou les différentes implémentations de l'opération de jointure (boucles imbriquées, tri fusion, hash, etc.). Dans les applications décisionnelles, la conception physique est devenue un enjeu important, comme l'indique Surajit Chaudhuri dans son papier intitulé *Self-Tuning Database Systems : A Decade of Progress*, à qui on a décerné le prix de *10 Year Best Paper Award* à la conférence Very Large Databases (VLDB'2007). Cette amplification est due aux caractéristiques suivantes liées aux entrepôts de données : (1) le