

Index de Jointure Binaires: Stratégies de Sélection & Étude de Performances

Kamel Boukhalfa *, Ladjel Bellatreche** Benameur Ziani ***

* USTHB - Bp 32 El-Alia Alger, Algérie
boukhalk@ensma.fr

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

*** Université de Laghouat - Algérie
bziani@mail.lagh-univ.dz

Résumé. La conception physique des entrepôts de données relationnels est basée essentiellement sur la sélection d'un ensemble d'index afin de réduire le coût d'exécution des requêtes OLAP complexes. Ces entrepôts sont généralement modélisés par un schéma en étoile caractérisé par une table de faits volumineuse et un ensemble de tables de dimension liées à la table des faits par leurs clés étrangères. Les requêtes définies sur ce schéma (appelées requêtes de jointure en étoile) comportent plusieurs jointures entre la table des faits et les tables de dimension ce qui rend leur coût d'exécution considérable. Les index de jointure binaires sont très adaptés pour réduire le coût d'exécution de ces jointures. Ils sont définis sur la table de faits en utilisant un ou plusieurs attributs de dimension. Sélectionner une configuration d'index pour réduire le coût d'exécution d'un ensemble de requêtes est reconnu comme un problème NP-Complet. Dans ce papier, nous présentons d'abord le problème de sélection des index de jointure binaires et les principaux travaux effectués dans ce domaine. Nous présentons par la suite notre approche de sélection et les algorithmes que nous proposons. Nous effectuons des expériences pour comparer les différentes stratégies de sélection. Enfin, nous effectuons une validation réelle des différents algorithmes sous Oracle en utilisant les données issues du banc d'essai APB1.

1 Introduction

Les entrepôts de données stockent un volume de données très important essentiellement dans des modèles logiques relationnels comme les schémas en étoile ou flocon de neige (Kimball et Strehlo, 1995). Ces derniers sont accédés par des requêtes décisionnelles complexes caractérisées par de multiples opérations de sélection, de jointure et d'agrégation. Les requêtes définies sur un schéma en étoile sont appelées *requêtes de jointure en étoile* comportant un nombre de prédicats de sélection définis sur des tables de dimension et des prédicats de jointures entre la table des faits et les tables de dimension. Pour optimiser les sélections et les