

Sélection simultanée d'index et de vues matérialisées

Nora Maiz, Kamel Aouiche et Jérôme Darmont

Laboratoire ERIC, Université Lumière Lyon 2

5 avenue Pierre Mendès-France

69676 Bron Cedex

{nora.maiz, kamel.aouiche, jerome.darmont}@eric.univ-lyon2.fr

<http://eric.univ-lyon2.fr>

Résumé. Les index et les vues matérialisées sont des structures physiques qui accélèrent l'accès aux données d'un entrepôt. Ces structures engendrent cependant une surcharge de maintenance. Par ailleurs, elles partagent le même espace disque. Les travaux existants dans le domaine de la sélection d'index et de vues matérialisées traitent ces deux structures de manière isolée. Dans cet article, nous couplons au contraire la sélection d'index et de vues matérialisées de façon à prendre en compte les interactions entre ces structures de données et à permettre un partage efficace de l'espace de stockage commun qui leur est alloué. Pour cela, nous avons développé des modèles de coût qui évaluent le bénéfice de la matérialisation de vue et de l'indexation. Ces modèles de coût nous permettent, grâce à un algorithme glouton, de sélectionner une configuration pertinente d'index et de vues matérialisées. Nos expérimentations montrent que notre stratégie se révèle meilleure que celles qui opèrent une sélection isolée des index et des vues matérialisées.

1 Introduction

Les entrepôts de données sont généralement modélisés selon un schéma en étoile contenant une table de faits centrale volumineuse et un certain nombre de tables dimensions représentant les descripteurs des faits Inmon (2002); Kimball et Ross (2002). La table de faits contient des clés étrangères vers les clés primaires des tables dimensions, ainsi que des mesures numériques. Avec ce type de modèle, une requête décisionnelle nécessite une ou plusieurs jointures entre la table de faits et les tables dimensions. De plus, le schéma de l'entrepôt peut comporter des hiérarchies au niveau des dimensions (schéma en flocon de neige), ce qui entraîne des jointures additionnelles. Ces jointures sont très coûteuses en terme de temps de calcul. Ce coût devient prohibitif lorsque les jointures opèrent sur de très grands volumes de données. Il est alors crucial de le réduire.

Les vues matérialisées et les index sont des structures physiques qui permettent de réduire le temps d'exécution des requêtes en précalculant les jointures et en offrant un accès direct aux données. Cependant, lors du rafraîchissement de l'entrepôt de données, ces structures doivent également être mises à jour, ce qui engendre une surcharge pour le système. Par ailleurs, index