

Auto-administration des entrepôts de données complexes

Kamel Aouiche*

Jérôme Darmont**, Omar Boussaid**, Fadila Bentayeb**

Laboratoire ERIC, Université Lumière Lyon 2
5 avenue Pierre Mendès-France
69676 Bron Cedex

* kaouiche@eric.univ-lyon2.fr

** {jdarmont,boussaid,bentayeb}@univ-lyon2.fr

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

Résumé. Les requêtes définies sur les entrepôts de données sont souvent compliquées et utilisent plusieurs opérations de jointure qui sont coûteuses en terme de temps de calcul. Dans le cadre de l'entreposage de données complexes, les adaptations apportées aux schémas classiques d'entrepôts induisent des jointures supplémentaires lors des accès aux données. Ce coût devient encore plus important quand les requêtes opèrent sur de très grands volumes de données. Il est donc primordial de réduire ce temps de calcul. Pour cela, les administrateurs d'entrepôts de données utilisent en général des techniques d'indexation comme les index de jointure en étoile ou les index *bitmap* de jointure. Cela demeure néanmoins complexe et fastidieux.

La solution proposée s'inscrit dans une optique d'auto-administration des entrepôts de données. Dans ce cadre, nous proposons une stratégie de sélection automatique d'index. Pour cela, nous avons recouru à une technique de fouille de données, plus particulièrement la recherche de motifs fréquents, pour déterminer un ensemble d'index candidats à partir d'une charge donnée. Nous proposons ensuite des modèles de coût permettant de sélectionner les index engendrant le meilleur profit. Ces modèles de coût évaluent en particulier le temps d'accès aux données à travers des index *bitmap* de jointure, ainsi que le coût de maintenance et de stockage de ces index.

Mots clés : Entrepôts de données, données complexes, auto-administration, sélection d'index, motifs fréquents, modèles de coût, index *bitmap* de jointure.

1 Introduction

Les entrepôts de données classiques 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 les clés (étrangères) des tables dimensions et les mesures. La Figure 1 montre un exemple d'entrepôt de données modélisé en étoile composé de la