

Structure Réutilisable pour le Calcul et la Manipulation des Cubes de Données

Hassani Hachim, Noël Novelli

LIF-CNRS UMR 6166 – Case 901

Université de la Méditerranée

Faculté des Sciences de Luminy ; 163, Avenue de Luminy

13288, Marseille Cedex 9, FRANCE

{hachim, novelli}@lif.univ-mrs.fr

Résumé. Les cubes de données sont de plus en plus utilisés pour le pré-calcul de requêtes OLAP afin de permettre essentiellement à des analystes de trouver des tendances ou des anomalies dans de grandes quantités de données. Il se révèle que tout problème lié aux cubes de données est coûteux, que ce soit pour la construction, la matérialisation, la manipulation ou la mise à jour. Dans cet article, nous introduisons la notion de pré-calcul de cubes de données et la caractérisation associée qui est basée sur le modèle partitionnel. A notre connaissance, aucune des approches actuelles ne s'est intéressée à la réutilisation de pré-calcul des cubes de données. Pourtant cette dernière permet de calculer et de manipuler efficacement des cubes de données dans plusieurs contextes comme les applications météorologiques, le calcul de requêtes à la volée ou encore le calcul de plusieurs cubes de données en réseau ou en local.

1 Introduction

Conceptuellement, l'opérateur GROUP BY CUBE introduit dans Gray et al. (1996) représente toutes les combinaisons des valeurs des dimensions auxquelles sont associées des mesures (valeurs numériques). C'est un outil analytique inestimable pour les applications nécessitant une analyse de quantités volumineuses de données comme on en trouve dans les systèmes de support de décision, le *business intelligence* et le datamining. De telles applications nécessitent des réponses rapides aux requêtes, ad-hoc pour la plupart, sur une base de données ou un entrepôt de données. Le cube de données constitue l'une des représentations possibles permettant d'optimiser ce type de requêtes car il correspond au pré-calcul de toutes les requêtes possibles pour une fonction agrégative donnée.

Cet opérateur exécute les GROUP BY correspondant à toutes les combinaisons possibles de l'ensemble de dimensions. La figure 1 montre l'équivalence sémantique entre l'opérateur GROUP BY CUBE et l'union de toutes les combinaisons de GROUP BY pour la table *UneTable* avec $Dim = \{A, B, C\}$ l'ensemble des dimensions, \mathcal{M} la mesure et f la fonction agrégative. Cette requête représente $2^{|Dim|}$ GROUP BY et chaque résultat de GROUP BY est un cuboïde.