

Modélisation des bases de données multidimensionnelles à agrégations multiples et différenciées

Ali Hassan*, Franck Ravat*, Olivier Teste**, Ronan Tournier*, Gilles Zurfluh*

*Université Toulouse 1 Capitole - IRIT (UMR 5505)

**Université Toulouse 3 Paul Sabatier - IRIT (UMR 5505)

118, Route de Narbonne - 31062 Toulouse cedex 9 (France)

{hassan, ravat, teste, tournier, zurfluh}@irit.fr

Résumé. De nombreux modèles ont été proposés pour représenter les données multidimensionnelles. Ces propositions considèrent généralement une même fonction d'agrégation pour déterminer les valeurs d'une mesure aux différents niveaux de granularité de l'espace multidimensionnel. Nous proposons un nouveau modèle conceptuel plus flexible supportant des agrégations multiples différenciées. L'agrégation multiple permet d'associer à une même mesure, une fonction d'agrégation différente pour chaque dimension. L'agrégation différenciée autorise des agrégations spécifiques à chaque paramètre. Notre modèle repose sur un double formalisme graphique suffisamment expressif pour contrôler la validité des fonctions d'agrégation. Nous étudions également les conséquences de cette modélisation conceptuelle pour la construction efficace des treillis de pré-agrégats dans le contexte R-OLAP.

1 Introduction

Les analyses OLAP consistent à suivre des indicateurs considérés comme des points observés dans un espace défini par différents axes d'analyse. Les données analysées sont hiérarchisées selon différents niveaux de détails et des fonctions d'agrégation sont utilisées pour obtenir une vision synthétique. Les données sont regroupées selon le niveau de détail sélectionné et agrégées avec les fonctions. Les opérations de forage (*roll-up et drill-down*), souvent employées lors des analyses OLAP, font une utilisation intensive de ces fonctions d'agrégation.

Les bases de données multidimensionnelles (BDM) offrent classiquement un cadre adéquat aux analyses décisionnelles, cependant les fonctions d'agrégation fournies par défaut peuvent s'avérer imparfaites. Par exemple, l'analyse de températures d'un territoire peut être réalisée de manière uniforme avec une seule et même fonction d'agrégation : la température maximale annuelle est obtenue à partir des températures maximales mensuelles, elles-mêmes obtenues à partir des températures maximales quotidiennes. Par contre, l'analyse des températures par département repose généralement sur la moyenne des températures par ville, tandis que l'analyse des températures par région peut être effectuée à partir d'une agrégation des températures des départements qui tient compte de la superficie du département (moyenne pondérée). Cette analyse fait donc intervenir différentes fonctions d'agrégation (moyenne ou moyenne pondérée) selon le niveau d'analyse considéré. D'autre part, dans l'analyse classique, l'agrégation à