

Calcul et Représentation Efficace de Cubes de Données pour une Visualisation Orientée Pixel

Noël Novelli*, David Auber**

*Université de la Méditerranée ; Faculté des Sciences de Luminy
163, av. de Luminy - Case 901 - LIF ; F-13288 Marseille cedex 9 ; France
novelli@lif.univ-mrs.fr
<http://www.lif.univ-mrs.fr>

**Université de Bordeaux I ; Bât A30, LaBRI
351, cours de la Libération ; F-33405 Talence cedex ; France
auber@labri.fr
<http://www.labri.fr/~auber>

Résumé. Les cubes de données fournissent une aide non négligeable lorsqu'il s'agit d'interroger des entrepôts de données. Un cube de données représente un pré-calcul de toutes les requêtes OLAP et ainsi améliore leur temps de réponses. Les approches proposées jusqu'à présent réduisent les temps de calcul et d'entrée sortie mais leur utilisation reste très coûteuse. D'autres travaux de recherche se sont intéressés à la visualisation de données pour les exploiter de façon interactive.

Nous proposons une adaptation de la représentation condensée des cubes de données basée sur le modèle partitionnel. Cette technique nous permet de calculer efficacement un cube de données et de représenter les liens entre les données pour la visualisation. La visualisation proposée dans cet article est basée sur des techniques de visualisation orientée pixel et sur des techniques de diagramme de liens entre nœuds pour offrir à la fois une vision globale et locale pour l'exploitation. Cette nouvelle approche utilise d'une part les calculs efficaces de cubes de données et d'autre part les techniques avancées de visualisation.

Contribution

Notre objectif est de fournir aux analystes un outil de visualisation interactive de cubes de données (Gray et al. (1996)). Pour cela, nous proposons une technique à la fois globale et locale à l'aide de deux représentations. Notre visualisation permet non seulement de visualiser le cube mais aussi les liens entre ses éléments. Pour réduire les besoins mémoire, la visualisation n'utilise pas de mémoire pour les calculs de représentation ou d'interaction. Pour les calculs liés aux cubes de données et à leurs manipulations, nous proposons un algorithme (extension de PCUBE (Casali et al. (2006)) basé sur la notion de partition (Cosmadakis et al. (1986)) pour calculer le cube et retrouver les connexions entre les éléments du cube.

Principe général de l'approche L'architecture logicielle du prototype que nous avons réalisé se décompose en deux parties : la visualisation (application) et les calculs et accès aux données (bibliothèque). Nous appliquons la méthode de *Ray Tracing* afin de ne calculer que les parties du cube qui seront effectivement exploitées et visualisées. Pour cela, l'application demande à la bibliothèque les informations dont elle a besoin pour la visualisation (valeur agrégée d'un élément, les prédécesseurs et successeurs d'un élément...). Tous les calculs de correspondance des coordonnées d'un pixel vers l'élément correspondant dans le cube sont réalisés en temps constant (de même pour l'opération inverse lors de sélections par exemple) grâce à des fonctions de plongement bijectives en utilisant des techniques similaires à Keim (2000) et Ware (2000). La technique de partition utilisée pour la représentation interne du cube de données nous permet de répondre rapidement aux demandes de l'interface. En effet, les calculs d'agrégation sont linéaires aux nombres de valeurs à agréger et les informations de voisinages sont calculées à la demande (linéaire aux nombres de voisins pour les prédécesseurs et au nombre de tuples pour les successeurs).

Références

- Casali, A., R. Cicchetti, L. Lakhal, et N. Novelli (2006). Lossless reduction of datacubes. In *DEXA'2006*, Volume 4080 of *Lecture Notes in Computer Science*, pp. 409–419.
- Cosmadakis, S., P. Kanellakis, et N. Spyrtos (1986). Partition Semantics for Relations. *Journal of Computer and System Sciences* 33(2), 203–233.
- Gray, J., A. Bosworth, A. Layman, et H. Pirahesh (1996). Data Cube : A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Total. In *ICDE'96, New Orleans, Louisiana*, pp. 152–159.
- Keim, D. A. (2000). Designing pixel-oriented visualization techniques : Theory and applications. *IEEE Transactions on Visualization and Computer Graphics* 6(1), 59–78.
- Ware, C. (2000). *Information Visualization : Perception for design*. Interactive Technologies. Moragn Kaufmann.

Summary

Datacubes are especially useful for answering efficiently queries on data warehouses. A datacube is a pre-computing of the Olap queries to optimize the answer times. The recent approaches reduce the computing times and the I/O costs but the use of these representations can be very expensive. Other research works were interested in visualization of data in order to exploit them interactively.

We proposed an adaptation of the condensed representation of datacubes, based on the partition model. The condensed representation allows to effectively compute the datacube and to represent the relations between data for the visualization. The visualization proposed in the paper is based on the pixel oriented visualization techniques and a link node diagram technique that enables both global and local exploration. The approach is a new approach that uses the effective compute of the datacubes on the one hand and the advanced techniques of visualization on the other hand.