

Modélisation et interrogation de données XML multidimensionnelles

Serge Boucher*, Boris Verhaegen*, Esteban Zimányi*

* Université Libre de Bruxelles (U.L.B.)
Service CoDE - CP165/15 - 1050 Bruxelles - Belgique
prénom.nom@ulb.ac.be

Résumé. XML étant devenu omniprésent et ses techniques de stockage et d'interrogation de plus en plus efficaces, le nombre de cas d'utilisations de ces technologies augmente tous les jours. Un sujet prometteur est l'intégration d'XML et des entrepôts de données, dans laquelle une base de données XML native stocke les données multidimensionnelles et exécute des requêtes OLAP écrites à l'aide du langage d'interrogation XML XQuery. Ce papier explore les questions qui peuvent survenir lors de l'implémentation d'un tel entrepôt de données XML.

1 Introduction

Les données qu'une entreprise rassemble d'années en années deviennent son atout le plus précieux. Les technologies des entrepôts de données aident à tirer le meilleur parti de ces données via le stockage et l'organisation des données opérationnelles d'une manière adaptée pour faciliter l'analyse et la décision. Un entrepôt de données est généralement modélisé de façon multidimensionnelle : l'ensemble des données est considéré comme un ensemble de *faits*. Ces faits contiennent des *mesures* qui peuvent être analysées suivant différentes *dimensions* ou perspectives. Chaque dimension peut être organisée en *hiérarchie* composée de plusieurs niveaux permettant d'analyser les mesures suivant différentes *granularités*. Un tel modèle multidimensionnel se prête à de nombreux types de requêtes qui peuvent avoir une valeur commerciale considérable. Leur traitement est appelé analyse multidimensionnelle ou *Online Analytical Processing (OLAP)*.

L'augmentation de l'échange de données entre applications a incité la création de standards tel que XML, aujourd'hui omniprésent. D'énormes quantités de données sont maintenant disponibles au format XML et les outils permettant d'utiliser ces données s'améliorent chaque jour. Plus particulièrement, les bases de données XML natives et le langage d'interrogation XQuery, sont aujourd'hui suffisamment avancés pour être utilisés dans un environnement de production. L'approche traditionnelle pour l'entreposage de données XML est de les convertir en données relationnelles. Cependant, une piste de recherche intéressante est de mettre en place un entrepôt de données utilisant uniquement les technologies XML : les données peuvent être modélisées en tant que documents XML stockés dans une base de données XML native et analysés à l'aide de requêtes XQuery.

La principale contribution de ce papier est une discussion sur les problèmes soulevés lors de l'implémentation d'un entrepôt de données utilisant uniquement des technologies XML.