

Conception et optimisation d'un entrepôt de données médicales

Maria Trinidad Serna Encinas, Michel Adiba

LSR-IMAG Laboratory, BP 72 38402 Saint-Martin d'Hères, France
<http://www-lsr.imag.fr/Les.Groupes/STORM/Storm2002/Francais/index.html>.
{*Maria-Trinidad.Serna, Michel.Adiba*}@imag.fr

Résumé. Un entrepôt de données intègre des informations provenant des sources de données internes mais aussi externes à l'entreprise. L'ensemble des données est utilisé pour l'aide à la décision, ainsi la conception du modèle multidimensionnel et la sélection des vues à matérialiser représentent un processus complexe et délicat. Dans cet article, nous proposons la définition d'un modèle multidimensionnel qui se compose de trois classes : cube, dimension et hiérarchie. Nous proposons également un algorithme pour la sélection de l'ensemble optimal des vues à matérialiser. Notre algorithme utilise les paramètres de fréquence d'utilisation, de coût de calcul et de fréquence de mises à jour des relations de base. Nous avons eu l'opportunité de travailler dans le cadre d'un projet médical, ce qui nous a permis de vérifier et de valider notre proposition sur des données réelles.

1 Introduction

Les entrepôts de données sont apparus au début des années 1990 en réponse à la nécessité de rassembler toutes les informations de l'entreprise en une base de données unique destinée aux analystes et aux gestionnaires [Doucet et Gangarski, 2001]. Un processus de transformation et de nettoyage doit être appliqué aux données avant leur stockage dans l'entrepôt. L'ensemble des données, y compris leur historique, est utilisé pour l'aide à la décision [Kimball, 1996, Kimball et Ross, 2003].

La conception et la mise en oeuvre d'un entrepôt de données est une tâche complexe et délicate. Elle se compose de divers processus communément appelés extraction-intégration, organisation et interrogation. Pour l'extraction, nous devons analyser l'ensemble des sources de données internes et externes. Cette analyse sert aussi bien à la sélection de l'ensemble de données à stocker dans l'entrepôt, qu'à la sélection des outils requis pour l'extraction et la transformation de ces données avant leur stockage. Le deuxième processus consiste à organiser ces données à l'intérieur de l'entrepôt. Pour ce faire, nous devons concevoir le modèle multidimensionnel à utiliser ainsi que définir l'ensemble optimal de vues à matérialiser. Finalement, le dernier processus consiste à déterminer les outils nécessaires pour la visualisation de l'ensemble des données.

Nous avons participé au projet ADELEM (Aide à la DEcision Logistique Et Médicale). Ce projet nous a permis de valider notre proposition. Ainsi, dans ce travail, nous présentons une expérimentation sur les données médicales du projet.