

Résumé

Toute entreprise possède actuellement d'importants volumes de données, stockées le plus souvent dans des différents médias (bases de données, documents papiers...) et a besoin d'outils permettant une exploitation efficace et performante de ces données pour l'aider dans ses prises de décision.

Les entrepôts de données (ou Data Warehouses) apportent des solutions à cette problématique. Il permet de stocker des données nécessaires à la prise de décision, il est alimenté et mis à jour via des extractions de données provenant des bases de données pouvant être hétérogènes.

Cette pluralité des sources et l'hétérogénéité des informations qu'elles peuvent contenir exige un modèle unifié qui permet à la fois de décrire les structures de données et de les intégrer au sein de l'entrepôt de données. Plusieurs chercheurs préconisent l'utilisation de XML comme format pivot de description et d'échange de données dans le but d'intégrer différentes sources de données compatibles avec le schéma de l'entrepôt. Car dans la plupart des cas, les structures des données sources et leurs sémantiques peuvent être différentes de la structure et la sémantique du Data Warehouse.

Les avantages du langage XML ont fait de lui le meilleur candidat pour l'intégration de données, car en effet, il est riche, clair, simple, puissant, extensible, préserve la sémantique des données, indépendant de toute plate-forme et de tout logiciel.

L'objectif de notre travail est d'apporter une solution pour régler ce problème d'intégration, notre proposition est une méthode exhaustive, simple et automatique qui permet de traiter tous les cas de conflits générés lors de l'intégration de données externes dans un Data Warehouse.

Mots clés : Entrepôt de données, XML, XML Schémas, Bases de Données XML natives, Intégration, Hétérogénéité.

Abstract

Every firm has nowadays an important amount of information, stored generally in different medias (data bases, files...) and needs tools in order to operate efficiently on these data to facilitate taking decision.

The Data Warehouses bring solutions to his problem. They allow stocking all the necessary data to take decision, they are alimented and updated via data extractions coming from different data bases which can be totally heterogeneous.

This heterogeneity needs a unified model which provides the description of the information contained in the different data bases and their integration into the Data Warehouse. A lot of studies show that XML is the best language to use as a unified model. In fact, XML can represent the semantic and the syntax of different data bases.

XML has also different advantages, it is rich, clear, simple, extensible, preserves the semantic in the files and it is independent from any plat-form or software.

The goal of our approach is to provide a way to solve the problems faced when integrating heterogeneous information into Data Warehouse. Our proposition is an exhaustive, simple and automatic method which handles all conflicts cases generated during the data integration into a Data Warehouse.

Key words: Data Warehouses, XML, XML Schemas, XML Native Data Bases, Integration, Heterogeneity.