

# Évolution de schémas dans les entrepôts de données : modèle à base de règles

Cécile Favre, Fadila Bentayeb et Omar Boussaïd

Laboratoire ERIC – Université Lumière Lyon 2

Bât. L - 5 av. Pierre Mendès-France

69676 Bron Cedex

{cfavre,bentayeb}@eric.univ-lyon2.fr

omar.boussaid@univ-lyon2.fr

## 1 Contexte et objectif

La mise en œuvre d'un entrepôt de données nécessite un important travail d'étude de l'existant et de recueil de données pour bien traiter les besoins d'analyses. Hors, il est impossible de définir de façon exhaustive ces besoins pour l'ensemble des utilisateurs. La prise en compte de nouveaux besoins d'analyses nécessite l'évolution du schéma de l'entrepôt de données. Deux alternatives existent : la première propose la mise à jour du schéma comme dans Hurtado et al. (1999), la seconde consiste à gérer différentes versions du schéma comme dans Morzy et Wrembel (2004). Ces deux alternatives apportent une réponse à l'émergence de nouveaux besoins d'analyses qui sont engendrés par l'évolution des données, mais pas lorsqu'ils sont engendrés par des connaissances du domaine dont disposent les utilisateurs. En effet, dans les modèles existants, seules les données sont utilisées pour atteindre les objectifs d'analyses, les connaissances étant exclues du processus d'analyse. Notre objectif est donc de pouvoir intégrer les connaissances du domaine dans le processus d'entreposage afin d'apporter au modèle une flexibilité en termes d'évolution des contextes d'analyses. Pour atteindre cet objectif, nous proposons un nouveau modèle conceptuel d'entrepôt de données à base de règles, baptisé *R-DW* (*Rule-based Data Warehouse*), dans lequel les règles permettent d'intégrer les connaissances du domaine, exprimées par l'utilisateur.

## 2 Modélisation à base de règles

Le modèle *R-DW* est composé d'une partie fixe et d'une autre évolutive (Figure 1). La partie fixe est un schéma en étoile, elle comprend une table de faits et les dimensions de premier niveau (qui ont un lien direct avec la table de faits). La partie évolutive comprend un ensemble de règles qui génèrent des hiérarchies de dimensions. Ces règles définissent le lien d'agrégation entre deux niveaux de granularité dans une hiérarchie de dimension. Ces règles d'agrégation traduisent les connaissances du domaine exprimées par l'utilisateur lui-même ou extraites à l'aide d'un processus d'apprentissage. Ce sont des règles de type "*si-alors*". Dans la clause "*alors*" figure la définition du niveau de granularité supérieur, en fonction de conditions exprimées dans la clause "*si*" qui portent sur les niveaux de granularité inférieurs.

Afin d'assurer la cohérence du modèle induit par les règles, et par conséquent la cohérence des analyses, nous définissons des contraintes sur les règles : (1) validité du domaine de