

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

définition (une règle doit se baser sur des données disponibles dans l'entrepôt); (2) validité temporelle (une règle a une durée de validité bien définie); et (3) cohérence des règles (deux règles ne doivent pas se contredire).

Notre modèle *R-DW* présente plusieurs avantages par rapport aux modèles d'entrepôt existants. D'une part, il permet de faire évoluer les contextes d'analyses en permettant de créer des hiérarchies de dimensions de façon dynamique. D'autre part, il renforce l'interaction entre l'utilisateur et le système d'aide à la décision en permettant à celui-ci d'intégrer ses propres connaissances.

Pour valider notre approche, nous avons implémenté un prototype du modèle *R-DW* prenant en compte les connaissances exprimées par l'utilisateur. Nous avons développé une plateforme Web (HTML/PHP) qui interface le SGBD Oracle. La table de faits et les tables de dimensions sont définies dans Oracle. Une table supplémentaire regroupe l'ensemble des règles d'agrégation. La plateforme Web permet à l'utilisateur de visualiser et de définir les règles qui génèrent des axes d'analyses, et d'obtenir des analyses. Notre approche a été appliquée au cas réel de la banque LCL¹.

Ce travail ouvre différentes perspectives. Nous voulons mesurer la performance de notre approche. Puis, nous voulons greffer la partie évolutive sur un schéma en flocon de neige ou en constellation. Par ailleurs, après avoir considéré une connaissance liée aux dimensions, nous souhaitons prendre en compte celle portant sur les mesures.

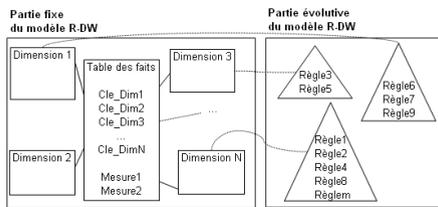


FIG. 1 – *Modèle R-DW*.

Références

- Hurtado, C. A., A. O. Mendelzon, et A. A. Vaisman (1999). Updating OLAP Dimensions. In *DOLAP'99 : 2nd ACM International Workshop on Data Warehousing and OLAP*.
- Morzy, T. et R. Wrembel (2004). On querying versions of multiversion data warehouse. In *DOLAP '04 : Proceedings of the 7th ACM international workshop on Data warehousing and OLAP*, New York, NY, USA, pp. 92–101. ACM Press.

Summary

To take into account users' knowledge for evolving data warehouse schema, we propose a Rule-based Data Warehouse (*R-DW*) model, where rules express the knowledge. Our model is composed of a fixed part, which is a star schema and an evolving part, defined with rules that dynamically create dimension hierarchies.

¹Collaboration avec la Direction d'Exploitation Rhône-Alpes Auvergne de LCL-Le Crédit Lyonnais dans le cadre d'une Convention Industrielle de Formation par la Recherche (CIFRE)