

Intégration des connaissances utilisateurs pour des analyses personnalisées dans les entrepôts de données évolutifs

Cécile Favre, Fadila Bentayeb, Omar Boussaïd

ERIC, Université Lumière Lyon 2
5 avenue Pierre Mendès-France
69676 Bron Cedex

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

Résumé. Dans cet article, nous proposons une approche d'évolution de schéma dans les entrepôts de données qui permet aux utilisateurs d'intégrer leurs propres connaissances du domaine afin d'enrichir les possibilités d'analyse de l'entrepôt. Nous représentons cette connaissance sous la forme de règles de type «si-alors». Ces règles sont utilisées pour créer de nouveaux axes d'analyse en générant de nouveaux niveaux de granularité dans les hiérarchies de dimension. Notre approche est fondée sur un modèle formel d'entrepôts de données évolutif qui permet de gérer la mise à jour des hiérarchies de dimension.

1 Introduction

Les entrepôts de données centralisent des données provenant de différentes sources pour répondre aux besoins d'analyse des utilisateurs. Le schéma de l'entrepôt est défini avec l'objectif d'analyser des *mesures* qui caractérisent des *faits*, en fonction de *dimensions* qui peuvent être organisées sous forme de *hiérarchies*, composées de différents *niveaux de granularité*, déterminant la manière selon laquelle sont agrégées les données.

Pour concevoir le schéma d'un entrepôt, nous distinguons dans la littérature différents types d'approches : celles guidées par les sources de données (Golfarelli et al., 1998), celles guidées par les besoins d'analyse (Kimball, 1996) et les approches mixtes qui combinent les deux approches précédentes, mettant en adéquation des schémas candidats générés à partir des sources de données avec les besoins d'analyse exprimés par les utilisateurs (Nabli et al., 2005).

Cependant, en pratique, les sources de données, tout comme les besoins d'analyse sont amenés à évoluer. Dans la littérature, il existe deux alternatives qui permettent l'évolution de schéma nécessaire suite à ces modifications. D'une part la mise à jour de schéma qui est réalisée grâce à des opérateurs qui font évoluer un schéma donné (Hurtado et al., 1999). D'autre part, la modélisation temporelle qui consiste à garder la trace de ces évolutions en utilisant des labels de validité temporelle. Ces labels sont apposés soit au niveau des instances (Bliujute et al., 1998), soit au niveau des liens d'agrégation (Mendelzon et Vaisman, 2000), ou encore au niveau des versions du schéma (Morzy et Wrembel, 2004). L'inconvénient de ce type de solutions est la nécessité d'une réimplémentation des outils d'analyse, de chargement, ... afin de gérer les particularités de ces modèles.

Les deux alternatives sont intéressantes pour répondre au problème de l'évolution de schéma suite à une modification dans les sources de données, puisque ce sont des solutions techniques