

Personnalisation dans les entrepôts de données : bilan et perspectives

Fadila Bentayeb*, Omar Boussaid*, Cécile Favre*
Franck Ravat**, Olivier Teste**

* Université de Lyon, ERIC Lyon 2
5 av. Pierre Mendès-France - 69676 Bron Cedex
bentayeb@eric.univ-lyon2.fr, {cecile.favrelomar.boussaid}@univ-lyon2.fr
<http://eric.univ-lyon2.fr>

** Université de Toulouse, IRIT
118, Route de Narbonne - F-31062 Toulouse cedex 9
{ravatlteste}@irit.fr
<http://www.irit.fr>

Résumé. Nous étudions dans cet article les enjeux et les opportunités relevant de la prise en compte des utilisateurs au sein des entrepôts de données. Pour ce faire, nous présentons tout d'abord un panorama des travaux sur la personnalisation dans les domaines connexes des bases de données et de la recherche d'information. Nous présentons ensuite les quelques travaux qui émergent dans le domaine des entrepôts de données et menons une étude comparative selon différents angles d'approche. Ceci nous permet de faire émerger des perspectives de recherche sur la personnalisation dans les entrepôts de données.

1 Introduction

L'approche d'entreposage de données (Kimball, 1996) (« data warehousing ») constitue un champ de recherche important dans lequel de nombreux problèmes restent à résoudre. Les entrepôts de données sont généralement intégrés dans un système d'aide à la prise de décision où l'on distingue deux espaces de stockage : l'entrepôt de données et les magasins de données (Figure 1). Un entrepôt de données est vu comme un espace de stockage centralisé homogène regroupant des données utiles aux analyses et à la prise de décision, extraites de sources et pouvant faire l'objet d'historisation (Inmon, 1996). Un magasin de données est un extrait de l'entrepôt, où les données sont préparées de manière spécifique pour faciliter leur analyse et leur exploitation par des utilisateurs appelés décideurs. Les données au sein d'un magasin sont souvent organisées de manière multidimensionnelle (Lehner, 1998) : les données analysées sont modélisées comme des points dans un espace à plusieurs dimensions formant un cube de données (Gray et al., 1996). Le cube de données peut alors être exploité par différents outils d'analyse. Il est en particulier possible, pour le décideur, de naviguer dans les données grâce à la technologie OLAP (On Line Analytical Processing).

Les travaux menés ces quinze dernières années ont essentiellement porté sur des aspects techniques tels que les processus ETL (« Extract-Transform-Load ») (Vassiliadis et al., 2002 ; Trujillo et Luján-Mora, 2003), les techniques d'optimisation et de sélection des vues matérialisées (Widom, 1995 ; Kotidis et Roussopoulos, 1999). Bien que les architectures décisionnelles soient considérées centrées utilisateurs, la prise en compte de ces derniers dans les systèmes décisionnels a finalement été peu étudiée. Des propositions ont été faites dans les méthodes de conception, notamment dans les méthodes descendantes, pour prendre