

Une nouvelle approche mixte d'enrichissement de dimensions dans un schéma multidimensionnel en constellation

Lucile Sautot*, Sandro Bimonte**, Ludovic Journaux***, Bruno Faivre*

*Laboratoire Biogéosciences (UMR CNRS-uB 6282)

6 boulevard Gabriel 21 000 Dijon FRANCE

lucile.sautot@u-bourgogne.fr, bruno.faivre@u-bourgogne.fr

**Irstea- Centre de Clermont-Ferrand

9 avenue Blaise Pascal 63178 Aubière FRANCE

sandro.bimonte@irstea.fr

***Laboratoire Électronique, Informatique et Image (UMR CNRS 6306)

allée Alain Savary 21000 Dijon France

ludovic.journaux@agrosupdijon.fr

Résumé. Les entrepôts de données (DW) et les systèmes OLAP sont des technologies d'analyse en ligne pour de grands volumes de données, basés sur les besoins des utilisateurs. Leur succès dépend essentiellement de la phase de conception où les exigences fonctionnelles sont confrontées aux sources de données (méthodologie de conception mixte). Cependant, les méthodes de conception existantes semblent parfois inefficaces, lorsque les décideurs définissent des exigences fonctionnelles qui ne peuvent être déduites à partir des sources de données (approche centrée sur les données), ou lorsque le décideur n'a pas intégré tous ces besoins durant la phase de conception (approche centrée sur l'utilisateur). Cet article propose une nouvelle méthodologie mixte d'enrichissement de schémas en constellation, où l'approche classique de conception est améliorée grâce à la fouille de données dans le but de créer de nouvelles hiérarchies au sein d'une dimension. Un prototype associé est également présenté.

1 Introduction

Les entrepôts de données (DW) et des systèmes OLAP sont des technologies permettant l'analyse en ligne de grands volumes de données. Les données entreposées sont organisées selon un modèle multidimensionnel qui définit les concepts de dimensions et de faits. Les dimensions représentent les axes d'analyse, qui sont organisés en hiérarchies, tandis que les faits, qui sont les sujets d'analyse, sont décrits par des indicateurs numériques appelés mesures. Les données entreposées sont ensuite explorées et agrégées en utilisant les opérateurs OLAP (par exemple, Roll-Up, Slice, etc.) (Kimball, 1996).

Le succès des projets de mise en place d'entrepôts de données dépend essentiellement de la phase de conception durant laquelle les besoins fonctionnels sont confrontés aux sources de données (Phipps et Davis, 2002). Trois principaux types de méthodologie de conception ont été développées : Une approche centrée sur les besoins utilisateurs, une centrée sur les sources