

# Enrichissement de schéma multidimensionnel en constellation grâce à la Classification Ascendante Hiérarchique

Lucile Sautot\*, Sandro Bimonte\*\*,  
Ludovic Journaux\*\*\*, Arnaud Larrère\*\*\*\*,  
Kevin Saint-Paul\*\*\*\*, Bruno Faivre\*\*\*\*

\*LIRIS, INSA de Lyon, Bât. Blaise Pascal, Campus de La Doua,  
69622 Villeurbanne Cedex, lucile.sautot@insa-lyon.fr,

\*\*IRSTEA Centre de Clermont-Ferrand, 9 Avenue Blaise Pascal,  
63170 Aubière, sandro.bimonte@irstea.fr

\*\*\*LE2I, Université de Bourgogne, Allée Alain Savary, 21000 Dijon  
ludovic.journaux@agrosupdijon.fr

\*\*\*\*Laboratoire Biogéosciences, Université de Bourgogne, 6 Boulevard Gabriel  
21000 Dijon, bruno.faivre@u-bourgogne.fr

**Résumé.** Les hiérarchies sont des structures cruciales dans un entrepôt de données puisqu'elles permettent l'agrégation de mesures dans le but de proposer une vue analytique plus ou moins globale sur les données entreposées, selon le niveau hiérarchique auquel on se place. Cependant, peu de travaux s'intéressent à la construction de hiérarchies, via un algorithme de fouille de données, prenant en compte le contexte multidimensionnel de la dimension concernée. Dans cet article, nous proposons donc un algorithme, implémenté sur une architecture ROLAP, permettant d'enrichir une dimension avec des données factuelles.

## 1 Introduction

Les entrepôts de données (DW) et les 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 (Kimball, 1996).

Les hiérarchies sont des structures importantes dans un entrepôt de données car elles permettent d'agréger plus ou moins finement les mesures, selon le niveau hiérarchique auquel on se place. C'est pourquoi plusieurs articles se sont intéressés à la construction de hiérarchies grâce à des algorithmes de fouille de données (Favre et al., 2006; Sautot et al., 2015). Cependant, les méthodes proposées prennent en compte uniquement les membres d'une dimension, et les faits et les autres dimensions du modèle en constellation ne sont pas impactés.