

# GAMM : un modèle multidimensionnel agile à base de graphes pour des entrepôts multi-versions

Redha Benhissen<sup>\*,\*\*</sup>, Fadila Bentayeb<sup>\*,\*\*</sup>  
Omar Boussaid<sup>\*,\*\*</sup>

\* Laboratoire ERIC, Université Lyon 2, Bron 69500, France  
\*\* {redha.benhissen,fadila.bentayeb,omar.boussaid}@univ-lyon2.fr

**Résumé.** Dans le contexte des big data où le volume de données croît continuellement, nous proposons dans cet article une approche agile d'évolution de schéma dans les entrepôts de données qui permet aux concepteurs d'intégrer de nouvelles sources de données et de tenir compte des nouveaux besoins d'utilisateurs afin d'enrichir les possibilités d'analyse. Notre approche est fondée sur un modèle évolutif en multi-versions basé sur la modélisation ensembliste sous forme de graphe. Un méta-modèle permettra la gestion des versions du schéma de l'entrepôt. Nous proposons également des fonctions d'évolution au niveau schéma (l'évolution des instances n'est pas traitée ici). Nous validons notre approche par un prototype logiciel qui gère à la fois les versions et l'évolution de schéma dans le modèle.

## 1 Introduction

L'approche d'entreposage de données Inmon (1992); Kimball (1996) constitue un champ de recherche important dans lequel de nombreux problèmes restent à résoudre. Les entrepôts de données (ED) centralisent des données provenant de différentes sources pour répondre aux besoins d'analyses des utilisateurs. Un des points clés de la réussite du processus d'entreposage de données réside dans la définition du modèle de l'entrepôt en fonction des sources de données et des besoins d'analyse. Avec l'avènement des big data nous assistons à une prolifération des sources de données amenant de nouveaux besoins d'analyse. Les big data offrent ainsi de nouvelles opportunités d'analyse engendrées par la diversification des sources de données. Lorsqu'un entrepôt de données est conçu, son schéma restera figé ; si les besoins d'analyse sont amenés à évoluer, leur prise en compte peut s'avérer coûteuse. Le modèle multidimensionnel classique, basé sur le modèle en étoile, présente des limites aux possibilités de changement, et son évolution est complexe que ce soit au niveau schéma ou au niveau données. Ces limites sont liées au modèle en étoile figé : car il est créé pour des besoins d'analyse connus à l'avance. C'est dans ce contexte que nous nous intéressons à l'évolution des entrepôts de données. Même si cette problématique n'est pas nouvelle car de nombreux travaux existent dans la littérature Faisal et al. (2017), elle continue tout de même à poser des questions de recherche notamment en termes d'évolution de schémas et de données, de gestion de ces évolutions ainsi que d'inté-