Vers une modélisation agile des entrepôts de données basée sur la technique Anchor Modeling

Thinhinane Hamitouche*, Omar Boussaid**
Fadila Bentayeb**

* LCSi, École nationale supérieure d'informatique, Alger
ft_hamitouche@esi.dz

**ERIC, Université de Lyon, Lyon2
omar.boussaid@univ-lyon2.fr, Fadila.Bentayeb@univ-lyon2.fr

Résumé. Sous l'influence des données massives (Big Data) le système d'information doit s'adapter et évoluer. Cette évolution passe par une transformation des systèmes décisionnels pour la prise en compte de la multiplication des sources de données et l'émergence de nouveaux besoins d'analyse. Même si la question d'évolution de modèle dans les entrepôts n'est pas nouvelle, concevoir des entrepôts évolutifs reste un défi à relever. Dans cet article nous proposons une approche de modélisation agile des entrepôts basée sur la technique Anchor Modeling qui permet de faire évoluer le modèle d'un entrepôt. Nous définissons des règles de passage d'un modèle multidimensionnel en étoile vers un modèle d'ancrage, puis d'un modèle d'ancrage vers un modèle physique relationnel. Nous validons notre approche par un prototype logiciel et comparons notre modèle multidimensionnel agile avec les modèles d'entrepôts en étoile traditionnels.

1 Introduction

La force des entrepôts de données réside dans leur capacité à intégrer et à agréger un ensemble de sources de données hétérogènes dans un système commun, offrant aux décideurs une vue unifiée des informations générées. Bien que les rafraîchissements des données dans un entrepôt soit une opération bien maîtrisée, sa portée reste limitée à une mise à jour des données sans impact sur la structure initiale de l'entrepôt, ni sur une prise en compte de nouvelles données pouvant faire évoluer les besoins d'analyse initiaux. Un entrepôt de données stocke des données structurées selon un modèle multidimensionnel, généralement représenté par un schéma en étoile. Or, l'environnement d'un entrepôt de données est en perpétuel changement, d'autre part, interfacier un entrepôt de données avec un lac de données, représentant un réel potentiel de données, permettrait à la BI, ou à tout système d'information décisionnel en général, d'exploiter davantage de données pour élargir et enrichir les besoins d'analyse. Transformer les entrepôts de données en structures agiles, pouvant accepter toute évolution, est