

Matérialisation partielle et interrogation d'un hypercube de données dynamiques

Usman Ahmed, Anne Tchounikine, Maryvonne Miquel

Université de Lyon, CNRS
INSA-Lyon, LIRIS, UMR5205, F-69621, France
firstname.lastname@insa-lyon.fr

Résumé. Les entrepôts de données ont généralement une stratégie de chargement des données par bloc et hors ligne ce qui les rendent peu compatibles avec des applications où les performances en temps sont critiques. Dans cet article, nous présentons un modèle multidimensionnel pour entreposer en temps réel les données d'un espace multidimensionnel hiérarchique. Nous proposons une matérialisation partielle de l'hypercube de données dans une structure d'arbre qui regroupe les données multidimensionnelles dans des partitions non ordonnées appelées Minimum Bounding Spaces (MBS). Nous présentons le principe des algorithmes d'insertion d'un nouveau fait et de requêtage. Nous évaluons la performance de notre solution en utilisant le Star Schema Benchmark. L'étude expérimentale montre que notre proposition est particulièrement performante à la fois en temps d'insertion et pour le traitement des requêtes.

1 Introduction

De plus en plus d'applications telles que la gestion des risques environnementaux, la finance ou le suivi de déplacement requièrent des analyses décisionnelles en temps réel. Or les modèles d'entrepôts de données classiques, avec leurs structures de données et leurs stratégies de mise à jour, ne sont plus efficaces dans des environnements dynamiques et temps réel. En effet :

- Du fait des politiques de mises à jour hors-ligne, les données les plus récentes sont intégrées dans l'entrepôt avec un délai qui peut nuire à la prise de décision. Elles ne sont pas prises en compte tant que l'opération de mise à jour (quotidienne, hebdomadaire, mensuelle, ...) n'est pas effectuée.
- Dans les entrepôts de données classiques, l'insertion des faits par blocs (bulk insertion) est privilégiée. Or, pour les applications décisionnelles où il est nécessaire de disposer des données les plus récentes, l'insertion de données dans l'entrepôt devrait être effectuée à la volée, tuple par tuple. Mais l'insertion d'un nouveau fait implique la mise à jour des agrégats, qui est une opération complexe et coûteuse en temps.
- Les applications visées nécessitent d'insérer non seulement de nouveaux faits, mais également de nouveaux membres de dimensions afin de tenir compte de l'évolution de l'environnement, comme par exemple l'installation ou le déplacement d'un capteur.