

Requêtes Skyline Hiérarchiques

Tassadit Bouadi*, Marie-Odile Cordier*, René Quiniou**

*IRISA - Université de Rennes 1, Campus de Beaulieu, 35042 RENNES, France
tassadit.bouadi@irisa.fr, marie-odile.cordier@irisa.fr

**IRISA - INRIA Rennes, Campus de Beaulieu, 35042 RENNES, France
rene.quiniou@inria.fr

Résumé. Les requêtes skyline constituent un puissant outil d'analyse de données multidimensionnelles et d'aide à la décision. Lorsque les dimensions sont conflictuelles, les requêtes skyline retournent les meilleurs compromis associés à ces dimensions. De nombreux travaux se sont intéressés à l'extraction de points skyline dans le contexte de bases de données multidimensionnelles, mais, à notre connaissance, aucun de ces derniers n'a traité la problématique des skyline associés à des données agrégées lorsque les dimensions sont multiples et hiérarchiques. Cet article présente une méthode qui étend la recherche des meilleurs compromis aux dimensions hiérarchiques. Notre proposition *HSky* (Hierarchical Skyline Queries) se focalise sur la définition de la sémantique et du calcul efficace de points skyline sur des dimensions présentant plusieurs niveaux hiérarchiques.

1 Introduction

Les requêtes skyline constituent un puissant outil d'analyse de données multidimensionnelles et d'aide à la décision. De nombreux travaux tels Raïssi et al. (2010); Wong et al. (2009); Tao et al. (2008); Jin et al. (2007b); Huang et al. (2008); Borzsonyi et al. (2001) se sont intéressés à l'extraction de points skyline dans le contexte de bases de données multidimensionnelles, mais aucun de ces derniers n'a traité la problématique des skyline associés à des données agrégées lorsque les dimensions sont multiples et hiérarchiques comme dans les entrepôts de données. Nous souhaitons coupler l'analyse skyline à l'analyse en ligne afin de permettre à l'utilisateur de naviguer vers les faits les plus intéressants de l'entrepôt de données. Dans cet objectif, il faut calculer les skyline en présence de dimensions hiérarchiques. Pour résoudre ce problème, plusieurs verrous scientifiques et techniques sont à lever. Peut-on les dériver à partir des skyline de niveau supérieur ou inférieur ? Peut-on étendre les algorithmes existants de calculs de skyline en tirant parti des dimensions hiérarchiques ? Doit-on recalculer tous les skyline pour chaque niveau hiérarchique ?

Dans la littérature, quelques récents travaux comme Antony et al. (2009, 2008); Jin et al. (2007a) se sont intéressés à l'application des requêtes skyline sur des données agrégées. Ces travaux sont principalement axés sur l'optimisation de requêtes sollicitant les deux opérateurs *Skyline* et *Group-By*, ce dernier étant un opérateur d'agrégation dans les bases de données. Cependant, ces approches se sont contentées d'exécuter séquentiellement ces deux opérateurs