

Un cadre conceptuel basé sur UML et Spatial OCL pour la définition des contraintes d'intégrité dans les systèmes SOLAP

Kamal Boulil*, Sandro Bimonte*
François Pinet*

*Irstea, UR TSCF, 24 Avenue des Landais, 63172 AUBIERE, France
kamal.boulil@irstea.fr,
sandro.bimonte@irstea.fr,
francois.pinet@irstea.fr

Résumé. Les entrepôts de données spatiales et les systèmes Spatial OLAP sont des technologies d'aide à la décision permettant l'analyse en ligne des gros jeux de données géo-référencées. Dans un tel type de systèmes, la qualité de l'analyse dépend : de la qualité des données entreposées, de comment les agrégations sont effectuées, et de comment les données entreposées sont explorées. Dans cet article, nous proposons une méthode pour garantir la qualité de ces trois facteurs, fondée sur un profil UML et sur des contraintes d'intégrité définies avec Spatial OCL. Afin de valider notre proposition, nous proposons également une implémentation automatique dans une architecture ROLAP.

1 Introduction

Les Entrepôts de Données Spatiales (EDS) et les systèmes Spatial OLAP (SOLAP) fournissent de nouvelles possibilités d'analyse spatio-multidimensionnelle de l'information géolocalisée (Bédard et al., 2006). Les EDS ont été définis comme "une collection de données spatiales et non spatiales orientées sujet, intégrées, non volatiles et variant dans le temps disponible pour le support du processus de prise de décision" (Stefanovic et al., 2000). Les données spatiales entreposées sont analysées au moyen de systèmes SOLAP qui ont été définis par Yvan Bédard comme des "plates-formes visuelles conçues spécialement pour faciliter une analyse spatio-temporelle rapide et simple, ainsi qu'une exploration des données, en suivant une approche multidimensionnelle basée sur des niveaux d'agrégation visualisés dans des affichages cartographiques ainsi que dans des tableaux et des diagrammes" (Bédard et al., 2006). Comme les EDS intègrent des données issues de plusieurs sources hétérogènes, plusieurs problèmes liés à leur qualité émergent. Afin d'améliorer la qualité des données dans les EDS, certaines approches ont été proposées pour "réparer" les données au moyen de techniques statistiques, de techniques de fouille de données, etc. (Ribeiro et al., 2011). Dans le même temps, les Contraintes d'Intégrité (CI) sont reconnues comme des méthodes efficaces pour exprimer des règles pour contrôler la cohérence logique des EDS (Salehi, 2009). En outre, la qualité des analyses spatio-multidimensionnelles dépend aussi de l'agrégation des mesures. Cette agréga-