

Vers la définition des contraintes d'intégrité d'entrepôts de données spatiales avec OCL

Kamal Boulil, Sandro Bimonte, Hadj Maboubi, François Pinet

Cemagref, UR TSCF, 24 Avenue des Landais,
63172 Aubière, France
kamal.boulil@cemagref.fr

Résumé. Les Entrepôts de Données Spatiales (EDS) et les systèmes SOLAP représentent une solution efficace pour l'analyse spatiale de phénomènes géographiques. Cependant, la qualité de cette analyse dépend fortement de la qualité des données stockées. Dans ce contexte, quelques travaux se sont intéressés à la définition et la spécification des contraintes d'intégrité spécifiques aux EDS. Dans cet article, motivé par le manque d'implémentation basée sur une approche MDA (Model Driven Approach), nous proposons deux classifications de contraintes d'EDS ainsi que leurs spécifications et implémentations en utilisant le standard OMG OCL (Object Constraint Language).

1 Introduction

Les entrepôts de données associés aux outils d'analyse OLAP représentent une solution efficace pour l'informatique décisionnelle (Inmon, 1996). Par ailleurs, les nouvelles technologies de l'information et de la communication permettent de récolter de très gros volumes de données spatiales. Ces données peuvent par exemple provenir de systèmes de télédétection. Des solutions, connues sous le terme d'OLAP Spatial (Yvan Bédard, 1997) visent à intégrer la donnée spatiale dans les systèmes OLAP. Le SOLAP enrichit les capacités d'analyse des systèmes OLAP classiques, en combinant des analyses multidimensionnelles avec des visualisations cartographiques. Ceci permet par exemple de comprendre la distribution géographique d'un phénomène et de comparer divers phénomènes à diverses échelles géographiques.

L'intégration des données (alphanumériques et notamment géographiques) dans l'analyse OLAP constitue un enjeu majeur car la qualité de cet analyse dépend de la qualité des données stockées dans l'Entrepôt de Données Spatiales (EDS). En effet, contrairement aux Bases de Données (BD), un Entrepôt de Données (ED) permet d'intégrer et d'historiser des données provenant de multiple sources. Cette intégration peut facilement poser des problèmes liés à l'imprécision, l'incertitude ou la sémantique des données. Dans ce contexte, quelques travaux ont été menés concernant la définition des Contraintes d'Intégrité (CI) dans les EDS (Salehi, 2009; Malinowsky et Zimanyi, 2008). Les CI sont des assertions qui identifient les données erronées et incohérentes et/ou empêchent leurs insertions. Dans le contexte spatial, les CI peuvent indiquer les relations spatiales (topologiques, d'ordre, métriques) autorisées entre objets géométriques.