
Designing Multidimensional Cubes from Warehoused Data and Linked Open Data

Conception de Cubes Multidimensionnels à partir de Données Entrepasées et de Données Ouvertes Liées

Franck Ravat¹, Jiefu Song¹, Olivier Teste²

1. IRIT - Université de Toulouse, Université Toulouse I Capitole
2 Rue du Doyen Gabriel Marty F-31042 Toulouse Cedex 09
{ravat/song}@irit.fr

2. IRIT - Université de Toulouse, Université Toulouse II Jean Jaurès
1 Place Georges Brassens F-31703 Blagnac Cedex
teste@irit.fr

Article accepté et présenté à la conférence internationale RCIS, co-localisée avec INFORSID 2016 à Grenoble. La version longue de l'article, en anglais, est disponible dans les actes de RCIS.

ABSTRACT. A Data Warehouse (DW) is widely used as a consistent and integrated data repository in Business Intelligence systems. Warehoused data provide only a partial view over the activities of an organization. In today's highly dynamic business context, Linked Open Data (LOD) should also be included in analyses to offer multiple perspectives to decision-makers. However, differences between DW and LOD models make it hard to analyze all useful data in a unified way, which complicates the analysis tasks. Moreover, the dispersion of warehoused data and LOD in different schemas leads to repetitive searches for relevant information among various data sources, which reduces the efficiency of analysis.

In this paper, we provide a new conceptual multidimensional model, named Unified Cube, which offers as much useful information as possible for analyses. The definition of Unified Cube is generic enough to represent both warehoused data and LOD according to an analysis subject and a set of analysis axes. Besides the concepts in multidimensional models related to one source, a Unified Cube supports some advanced modeling features depicting inter-source equivalent analysis granularities and aggregation paths spanning different sources.

We describe a two-stage process to build a Unified Cube from DWs and multidimensional LOD. As a first step, schemas published with specific modeling languages are transformed into a common conceptual representation named exportation cube. The second step consists of associating together related data scattered in different exportation cube. An algebraic linking operator is proposed to enable non-expert users to associate relevant data together according to their analytical needs. We propose an algorithm to automate the execution of the linking operator while guaranteeing the overall validity of a Unified Cube.

To demonstrate the feasibility of the proposed concepts, we develop a prototype implementing a Unified Cube built from a R-OLAP DW and two LOD datasets. We illustrate the advantages of our proposal by showing how analyses of both warehoused data and LOD are carried out through the implemented Unified Cube.

RESUME. De nos jours, les systèmes d'aide à la décision intègrent couramment un entrepôt de données (ED) intégrant le plus souvent les données internes d'une l'organisation. Les données entreposées ne contiennent pas toutes les informations nécessaires aux prises de décision. De ce fait, des données externes comme les Données Ouvertes Liées (DOL) pourraient être ajoutées afin d'offrir aux décideurs de nouvelles perspectives d'analyse. Néanmoins, analyser de manière unifiée les données entreposées et les DOL constitue une tâche complexe pour les non-spécialistes. Cette difficulté est due à (i) une modélisation différente des données d'ED et de DOL, (ii) une dispersion sur plusieurs sources des données relatives à un même sujet d'analyse. Pour effectuer une analyse décisionnelle, un décideur est donc obligé d'effectuer une recherche sur différentes sources de données.

Face à ces problématiques, nous proposons, dans cet article, une nouvelle modélisation conceptuelle, appelée Cube Unifié qui permet de regrouper toutes les informations utiles pour des analyses décisionnelles. Un Cube Unifié est suffisamment générique pour représenter aussi bien des données entreposées que des données ouvertes liées. Un cube unifié représente multi-dimensionnellement toutes ces données à l'aide d'un sujet d'analyse et d'un ensemble d'axes d'analyse composés de hiérarchies. L'intégration de ces différentes sources dans un modèle unifié nécessite d'étendre le concept de niveau de granularité d'analyse en y associant différents attributs provenant de différentes sources et en proposant des chemins d'agrégation supplémentaires reposant sur plusieurs sources.

En complément de cette modélisation, nous proposons un processus de construction d'un Cube Unifié en deux étapes. La première étape consiste à transformer différents schémas exprimés avec des langages de modélisation spécifiques en une représentation conceptuelle commune, appelée cube d'exportation. La deuxième étape permet d'associer et d'unifier les données provenant de différents cubes d'exportation. Un opérateur algébrique est proposé afin de traduire les besoins des concepteurs en des relations inter-données. Cet opérateur est implanté à l'aide d'un algorithme pour automatiser l'établissement des relations entre les données inter-sources et garantir la validité globale du Cube Unifié obtenu.

Afin de prouver la faisabilité de notre solution, nous avons développé un prototype implantant un Cube Unifié basé sur un ED R-OLAP (Relational On-Line Analytical Processing) et deux collections de données ouvertes liées. Nous montrons les avantages de notre proposition en identifiant et explicitant des analyses définies sur un cube unifié construit à partir de données entreposées et de DOL.

KEYWORDS: *Multidimensional Modeling; Linked Open Data; Unified Business Analyses*

MOTS-CLES: *Modélisation Multidimensionnelle; Données Ouvertes Liées; Analyses Décisionnelles Unifiées*
