

Chapter 3

An MDA Approach and QVT Transformations for the Integrated Development of Goal-Oriented Data Warehouses and Data Marts

Jesús Pardillo

University of Alicante, Spain

Jose-Norberto Mazón

University of Alicante, Spain

Juan Trujillo

University of Alicante, Spain

ABSTRACT

To customize a data warehouse, many organizations develop concrete data marts focused on a particular department or business process. However, the integrated development of these data marts is an open problem for many organizations due to the technical and organizational challenges involved during the design of these repositories as a complete solution. In this article, the authors present a design approach that employs user requirements to build both corporate data warehouses and data marts in an integrated manner. The approach links information requirements to specific data marts elicited by using goal-oriented requirement engineering, which are automatically translated into the implementation of corresponding data repositories by means of model-driven engineering techniques. The authors provide two UML profiles that integrate the design of both data warehouses and data marts and a set of QVT transformations with which to automate this process. The advantage of this approach is that user requirements are captured from the early development stages of a data-warehousing project to automatically translate them into the entire data-warehousing platform, considering the different data marts. Finally, the authors provide screenshots of the CASE tools that support the approach, and a case study to show its benefits.

DOI: 10.4018/978-1-4666-2044-5.ch003

INTRODUCTION

A *corporate data warehouse* is a repository that provides decision makers with a large amount of historical data concerning the overall enterprise strategy. A data-warehousing architecture defines a set of data repositories and their relationships to support the decision-making process in a given organization. Several architectural options (Cabibbo & Torlone, 2001; Jarke et al., 1999; Jukic, 2006; Samos et al., 1998; Watson et al., 2001) and methodologies (Bonifati et al., 2001; Giorgini et al., 2008; Luján-Mora & Trujillo, 2006a; Mazón et al., 2007a; Sen & Sinha, 2005) have been proposed to develop these repositories. Specifically, two foundational data-warehousing alternatives have been broadly discussed (Breslin, 2004): the *top-down* approach originally stated by Inmon (2005) and the *bottom-up* approach stated by Kimball and Ross (2002). The basis of these approaches consists of which data repositories should be developed first: a *corporate data warehouse* in which an organization's data are stored and integrated in a single repository (top-down) or *departmental data marts* in which data are aggregated and customized for particular information needs (bottom-up). Although the former is considered to be the most elegant solution from a theoretical point of view, it is usually hard to implement since the project scope involves the whole organization (Watson et al., 2001), and the second approach is thus more suitable for agile developments despite the problems that arise during data-mart integration (Watson et al., 2001; Chaudhuri & Dayal, 1997). Both approaches fail when they attempt to derive the second data repositories (*i.e.*, data marts or corporate data warehouse, respectively) due to the inherent high cost associated to the integration of huge amounts of data (top-down) and to the duplicated integration tasks done by data marts (bottom-up). In order to overcome these limitations, Kimball and Ross (2002) have also proposed a bus architecture articulated by *conformed dimensions*. These dimensions account

for 90 percent of the integration efforts made in order to tie data marts together (Kimball & Ross, 2002). They are obtained through the agreement of the entire organization, thus supporting truly cross-departmental decision-making processes. Despite all this, this solution is designed at the logical level (*i.e.*, by using relational schemata), and does not therefore provide suitable mechanisms to drive complex developments such as methodologies (Bonifati et al., 2001; Giorgini et al., 2008; Luján-Mora & Trujillo, 2006; Mazón et al., 2006; Mazón & Trujillo, 2008) based on conceptual modeling (Abelló et al., 2006; Golfarelli et al., 1998; Hüsemann et al., 2000; Luján-Mora et al., 2006). Furthermore, existing matching methods do not cover the particular problems of integrating data warehouse and data mart schemas (Evermann, 2008).

However, we believe that the surrounding architectural debate (Breslin, 2004) has been overlooked by the current development approaches which are mainly based on conceptual modelling. These approaches have focused on capturing information requirements by means of *multidimensional modelling* (Kimball & Ross, 2002; Chaudhuri & Dayal, 1997) which organizes data in terms of *facts* and *dimensions* of analysis, but does not specify how data repositories (*i.e.*, corporate data warehouse and their dependent data marts) are built from them. For instance, departmental data marts may be built by different development teams in isolation. They therefore lack incorporated conformity issues to solve the integrated development of data marts and corporate data warehouses, in order to assure cross-departmental information needs such as those answered by *drill-across* operations during “on-line analytical processing” (OLAP) (Chaudhuri & Dayal, 1997).

In this article, we present an approach based on *goal-oriented requirement engineering* (Yu & Mylopoulos, 1994) and *model-driven engineering* (Bézivin, 2006) technologies to solve the architectural debate (Breslin, 2004) by supporting Kimball's insights (Kimball & Ross, 2002) at the

26 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/mda-approach-qvt-transformations-integrated/74389

Related Content

Fault-Tolerant Quorum Consensus Scheme for Replication Control in Mobile Distributed Database Systems: FTQC

Siwoo Byun and Songchun Moon (1998). *Journal of Database Management* (pp. 16-24).

www.irma-international.org/article/fault-tolerant-quorum-consensus-scheme/51200

From Chinese Philosophy to Knowledge Discovery in Databases A Case Study: Scientometric Analysis

Pei Liu and Eric Boutin (2009). *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 632-642).

www.irma-international.org/chapter/chinese-philosophy-knowledge-discovery-databases/20749

Database Pointers in Navigational and Object-Oriented Database Management Systems: A Comparison

Mark Gillenson, Raymond D. Frost and Michael G. Kilpatrick (1995). *Journal of Database Management* (pp. 14-23).

www.irma-international.org/article/database-pointers-navigational-object-oriented/51155

Practical Case Study of a Web-Based Tutor Payment System

Tanguy Chateau, Cecile Leroy, Johanna W. Rahayu and David Taniar (2003). *Web-Powered Databases* (pp. 88-120).

www.irma-international.org/chapter/practical-case-study-web-based/31425

Integrity Issues in the Web: Beyond Distributed Databases

Jose F. Aladana Montes, Mariemma I. Yague del Valle and Antonio C. Gomez Lora (2002). *Database Integrity: Challenges and Solutions* (pp. 293-321).

www.irma-international.org/chapter/integrity-issues-web/7885