An MDA Approach for the Evolution of Data Warehouses

Said Taktak, MIRACL Laboratory, University of Sfax, Sfax, Tunisia
Saleh Alshomrani, Faculty of Computing and Information Technology, University of Jeddah, Jeddah, Saudi Arabia
Jamel Feki, MIRACL Laboratory, University of Sfax, Sfax, Tunisia
Gilles Zurfluh, Institut de Recherche en Informatique de Toulouse (IRIT), Université Toulouse I Capitole, Toulouse, France

ABSTRACT

Modeling and data warehousing have been considered, for more than one decade, as a new challenging research topic for which different approaches have been proposed. Nevertheless these proposals have focused on static aspects only. In practice, the evolution of the operational information system can lead to changes in its dependent multidimensional data warehouse (i.e. that this system feeds with data), and therefore may require the evolution of the data warehouse model. In this evolving context, the authors propose a model-driven based approach in order to automate the propagation of the evolutions occurred in the source database towards the multidimensional data warehouse. This approach is based on two evolution models, along with a set of transformation rules formalized in Query/View/Transformation. This paper describes this evolution approach for which we are developing a software prototype called DWE© (Data Warehouse Evolution).

Keywords: Data Warehouse, Evolution Model, MDA, QVT, Transformation Rules

1. INTRODUCTION

A data warehouse (DW) is characterized by a complex architecture where data are extracted from transactional data sources then processed, cleaned and finally loaded into fact tables and dimension tables (Inmon, 2002) in order to be used for decision purposes by On-Line Analytical Processing (OLAP). The DW may undergo some necessary changes that affect its data values or even its data model.

All DW design methods and DW software tools (Oracle warehouse builder, SQL Server…) are based on a rigid assumption that the conceptual model of the DW is time-invariant. However, in practice, this assumption restricts the evolution of reality and does not hold most of the time since, the DW may evolve due to internal and external actions. In fact, whatever is the used design method top-down (Kimball & Ross, 2002), bottom-up (Golfarelli, Rizzi, & Vrdoljak, 2001; Rusu, Rahayu, & Taniar, 2005) or mixed (Nabli, Soussi, Feki, Ben Abdallah, & Gargouri, DOI: 10.4018/ijdsst.2015070104
2010; Phipps & Davis, 2002), it is really difficult to determine completely the DW model at the
design phase; firstly because user requirements are often ad hoc and then could not be thoroughly
determined a priori; and secondly because the dynamic evolution of business processes within
the enterprise can lead to the evolution of the data source model that feeds the DW with data.
Naturally, the frequency of changes differs from one domain to another; for instance, telecommu-
nication data sources change their schemas every 7–13 days, on the average. Banking data
sources are more stable but they change their schemas every 2–4 weeks on the average (Bel-
latreche & Wrembel, 2013).

Therefore, the above reasons and situations lead necessarily to change the DW model after
its implementation; as a consequence, the ETL process (Extract-Transform-Load) also becomes
subject to changes.

Despite the fact that some DW software programs manage slight changes, nevertheless
they consider changes of data values only, such as the problem of slowly changing dimensions
(Santos & Belo, 2011). Hence, we consider that the problem of changes in the DW model has
not been sufficiently addressed yet by the research community.

Furthermore, in the literature of decision support systems, the evolution of DW due to
the evolution of the data source model was not treated in the context of MDA (Model Driven
Architecture). From our point of view, using the MDA paradigm is extremely interesting to ef-
ficiently automate the propagation of changes from the data source model towards the DW model
especially for critical-time changes (telecom providers, banking sector…). The more rapid is
the propagation the more the decision support system is ready and the decisions are taken on-
time. The rapidity and efficiency factors represent very crucial criteria here in order to ensure
the continuity of services provided by the DW.

Actually, our big challenge is how to propagate changes performed on a data source model
towards the DW conceptual/physical and ETL models. This paper is interested in this problematic
and discusses it. Specifically, we focus on studying the impact of changing the relational data
source model on the DW multidimensional model. In order to achieve this objective, we adopt
the MDA paradigm.

This paper is organized as follows. In Section 2, we give an overview of the works related
to the DW evolution problem. Section 3 describes our approach which is based on MDA for the
propagation of the evolution from a relational data source (DS) model towards the multidimen-
sional DW model. Section 4 defines two evolution models which are the DS evolution model
and the DW evolution model. Section 5 formalizes the transformation rules whereas Section 6
describes the prototype supporting our approach. Finally, we conclude the paper and enumerate
some perspectives.

2. RELATED WORKS

The DW evolution problem has been the main topic of several research studies. It was treated
from different viewpoints which we can classify as follows: (i) The evolution of multidimensional
model, (ii) The maintenance of materialized views, and (iii) The adaptation of the ETL process.
We review these works in order to indicate our own position as compared to their contributions
and then, we focus on how to propagate the evolution of the DS model towards the DW model.

2.1. The Evolution of Multidimensional Model

We distinguish two trends of works dealing with the evolution problems of multidimensional
models: Schema evolution and Schema versioning.
Related Content

Structural Similarity Measures in Sources of XML Documents
www.igi-global.com/chapter/structural-similarity-measures-sources-xml/28154?camid=4v1a
The Effects of Prudential Supervision on Bank Resiliency and Profits in a Multi-Agent Setting
www.igi-global.com/chapter/the-effects-of-prudential-supervision-on-bank-resiliency-and-profits-in-a-multi-agent-setting/170896?camid=4v1a

The Role Of Simulation And Modern Business Games
www.igi-global.com/chapter/role-simulation-modern-business-games/29715?camid=4v1a

The HRA-Based Road Crash Data: A Methodology for Crash Investigation and Distribution Characteristics of Driver’s Failure Rate
www.igi-global.com/article/the-hra-based-road-crash-data/120541?camid=4v1a