MDBE: Automatic Multidimensional Modeling

Oscar Romero and Alberto Abelló

Universitat Politècnica de Catalunya
Dept. Llenguatges i Sistemes Informàtics
{oromero,aabello}@lsi.upc.edu

Abstract. The goal of this demonstration is to present MDBE, a tool implementing our methodology for automatically deriving multidimensional schemas from relational sources, bearing in mind the end-user requirements. Our approach starts gathering the end-user information requirements that will be mapped over the data sources as SQL queries. Based on the constraints that a query must preserve to make multidimensional sense, MDBE automatically derives multidimensional schemas which agree with both the input requirements and the data sources.

Keywords: Multidimensional Design, Design by Examples, DW.

1 Introduction

Traditionally, the design of the *multidimensional* (MD) conceptual schema of a data warehouse (DW) has been performed manually, but automating this process is essential to not depend on the expert's ability to apply the methodology chosen, and to avoid the tedious task of analyzing the data sources. Nowadays, some methodologies to derive the MD conceptual schema from the data sources have been presented, but most of them must be carried out manually and just a few of them automate the process. Automatable methods always rely on a thorough analysis of the relational sources, and they mainly share three limitations: end-user requirements are not considered, design patterns used to identify potential subjects of analysis are based on weak heuristics and they demand data source schemas normalized up to third normal form.

Our methodology [1] was conceived to overcome these limitations. The *MDBE* tool automatically derives MD conceptual schemas from relational sources bearing in mind the end-user requirements. Thus, being able to compare information requirements with actual information availability. Our approach starts gathering the end-user *information requirements* since a DW must give support to the information necessities of a decision maker. These requirements, properly formalized, are mapped over the data sources, and based on the constraints they must preserve to make MD sense, MDBE automatically derives conceptual schemas fulfilling the input requirements. Moreover, MDBE is able to identify implicit MD knowledge according to how the relational concepts are related in the logical schema. In short, MDBE properly tag factual and dimensional data with formal rules derived from the requirements and because of this, it is also able to cope with denormalized schemas.

Q. Li et al. (Eds.): ER 2008, LNCS 5231, pp. 534–535, 2008.

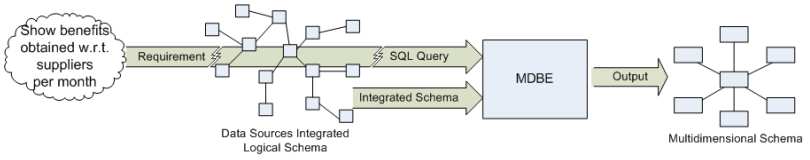

Fig. 1. MDBE overview

2 A MDBE Overview

The MDBE tool demands the MD requirements to lead the whole process. To do so, it is compulsory to translate them to a formal language understandable by computers. In our approach, requirements are translated into SQL queries over the relational data sources of the organization. Each SQL query, altogether with the data sources logical schema, would be the input of MDBE (see Fig. 1). As output, MDBE presents a MD schema derived from the data sources, which allows to retrieve data demanded in the input information requirement.

First step decomposes the input query and creates a *MD graph* corresponding to the query (i.e., relational tables and attributes as well as those relationships among them stated in the query). Our objective in this first step is to identify MD roles (i.e. *facts, measures, dimensions and levels*), whereas the second step aims to analyze the relationships among concepts to infer if indeed, this graph (i.e. the input query) makes MD sense. We say a query makes MD sense if it retrieves data derived from a valid sequence of MD operators over a MD schema. For this purpose, we carried out a study to identify which constraints should be guaranteed by a query in order to represent a combination of MD operators. These constraints may be summarized as follows: data retrieve should be (1) free of data summarizability anomalies, and (2) able to be placed in a MD space. If these constraints are guaranteed then, we may find a set of MD operators retrieving that data from the schema represented by the current graph. Eventually, the output MD schema is directly derived from the graph created along the process. Notice, however, that each query gives rise to a potential MD schema and the last step (not yet implemented) would embrace to conciliate those results in a minimum set of conceptual schemas meeting all the requirements.

Acknowledgments

This work has been partly supported by the Ministerio de Educación y Ciencia under project TIN 2005-05406.

Reference

1. Romero, O., Abelló, A.: Multidimensional Design by Examples. In: Tjoa, A.M., Trujillo, J. (eds.) DaWaK 2006. LNCS, vol. 4081, pp. 85–94. Springer, Heidelberg (2006)