Chapter 5
Multidimensional Design Methods for Data Warehousing

Oscar Romero
Universitat Politècnica de Catalunya, Spain

Alberto Abelló
Universitat Politècnica de Catalunya, Spain

ABSTRACT
In the last years, data warehousing systems have gained relevance to support decision making within organizations. The core component of these systems is the data warehouse and nowadays it is widely assumed that the data warehouse design must follow the multidimensional paradigm. Thus, many methods have been presented to support the multidimensional design of the data warehouse. The first methods introduced were requirement-driven but the semantics of the data warehouse (since the data warehouse is the result of homogenizing and integrating relevant data of the organization in a single, detailed view of the organization business) require to also consider the data sources during the design process. Considering the data sources gave rise to several data-driven methods that automate the data warehouse design process, mainly, from relational data sources. Currently, research on multidimensional modeling is still a hot topic and we have two main research lines. On the one hand, new hybrid automatic methods have been introduced proposing to combine data-driven and requirement-driven approaches. These methods focus on automating the whole process and improving the feedback retrieved by each approach to produce better results. On the other hand, some new approaches focus on considering alternative scenarios than relational sources. These methods also consider (semi)-structured data sources, such as ontologies or XML, that have gained relevance in the last years. Thus, they introduce innovative solutions for overcoming the heterogeneity of the data sources. All in all, we discuss the current scenario of multidimensional modeling by carrying out a survey of multidimensional design methods. We present the most relevant methods introduced in the literature and a detailed comparison showing the main features of each approach.

DOI: 10.4018/978-1-60960-537-7.ch005
**INTRODUCTION**

Data warehousing systems were conceived to support decision making within organizations. These systems homogenize and integrate data of organizations in a huge repository of data (the data warehouse) in order to exploit this single and detailed representation of the organization and extract relevant knowledge for the organization decision making. The data warehouse is a huge repository of data that does not tell us much by itself; like in the operational databases, we need auxiliary tools to query and analyze data stored. Without the appropriate exploitation tools, we will not be able to extract valuable knowledge of the organization from the data warehouse, and the whole system will fail in its aim of providing information for giving support to decision making. OLAP (On-line Analytical Processing) tools were introduced to ease information analysis and navigation all through the data warehouse in order to extract relevant knowledge of the organization. This term was coined by E.F. Codd in (Codd, 1993), but it was more precisely defined by means of the FASMI test that stands for fast analysis of shared business information from a multidimensional point of view. This last feature is the most important one since OLAP tools are conceived to exploit the data warehouse for analysis tasks based on multidimensionality.

The multidimensional conceptual view of data is distinguished by the fact/dimension dichotomy, and it is characterized by representing data as if placed in an n-dimensional space, allowing us to easily understand and analyze data in terms of facts (the subjects of analysis) and dimensions showing the different points of view where a subject can be analyzed from. One fact and several dimensions to analyze it produce what is known as data cube. Multidimensionality provides a friendly, easy-to-understand and intuitive visualization of data for non-expert end-users. These characteristics are desirable since OLAP tools are aimed to enable analysts, managers, executives, and in general those people involved in decision making, to gain insight into data through fast queries and analytical tasks, allowing them to make better decisions.

Developing a data warehousing system is never an easy job, and raises up some interesting challenges. One of these challenges focus on modeling multidimensionality. Nowadays, despite we still lack a standard multidimensional model, it is widely assumed that the data warehouse design must follow the multidimensional paradigm and it must be derived from the data sources, since a data warehouse is the result of homogenizing and integrating relevant data of the organization in a single and detailed view.

**Terminology and Notation**

Lots of efforts have been devoted to multidimensional modeling, and several models and methods have been developed and presented in the literature to support the multidimensional design of the data warehouse. However, since we lack a standard multidimensional terminology, terms used to describe the multidimensional concepts may vary among current design methods. To avoid misunderstandings, in this section we detail a specific terminology to establish a common framework where to map and compare current multidimensional design methods.

Multidimensionality is based on the fact/dimension dichotomy. Dimensional concepts produce the multidimensional space in which the fact is placed. Dimensional concepts are those concepts likely to be used as an analytical perspective, which have traditionally been classified as dimensions, levels and descriptors. Thus, we consider that a dimension consists of a hierarchy of levels representing different granularities (or levels of detail) for studying data, and a level containing descriptors (i.e., level attributes). In contrast, a fact contains measures of analysis. One fact and several dimensions for its analysis produce a multidimensional schema. Finally, we denote by base a minimal set of levels function-