Chapter 3.19
A Survey of Multidimensional Modeling Methodologies

Oscar Romero
Universitat Politècnica de Catalunya, Spain

Alberto Abelló
Universitat Politècnica de Catalunya, Spain

ABSTRACT

Many methodologies have been presented to support the multidimensional design of the data warehouse. First methodologies introduced were requirement-driven but the semantics of a data warehouse require to also consider data sources along the design process. In the following years, data sources gained relevance in multidimensional modeling and gave rise to several data-driven methodologies that automate the data warehouse design process from relational 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 methodologies have been introduced proposing to combine data-driven and requirement-driven approaches. On the other hand, new approaches focus on considering other kind of structured data sources that have gained relevance in the last years such as ontologies or XML. In this article we present the most relevant methodologies introduced in the literature and a detailed comparison showing main features of each approach.

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's 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 (O*n-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 (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 give rise to what is known as the *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 methodologies and approaches have been developed and presented in the literature to support the multidimensional design of a data warehouse. However, since we lack a standard multidimensional terminology, terms used among methodologies to describe the multidimensional concepts may vary. To avoid misunderstandings, in this section we detail a specific terminology to establish a common framework where map and compare current multidimensional design methodologies.

Multidimensionality is based on the fact/dimension dichotomy. Dimensional concepts give rise to the multidimensional space where the fact is placed. By dimensional concepts we refer to any concept likely to be used as a perspective of analysis. Traditionally, they have been classified as dimensions, levels and descriptors. Thus, we consider a dimension to contain a hierarchy of levels representing different granularities (or levels of detail) to study data, and a level to contain descriptors. On the other hand, a fact contains measures of analysis. One fact and several dimensions to analyze it give rise to a multidimensional schema. Finally, we denote by base a *minimal* set of levels functionally determining a fact. Thus, two different instances of data cannot be placed in the same point of the multidimensional space.

Let us consider now the example depicted in Figure 1. There, one fact (sales) containing two measures (price and discount) is depicted. This fact has four different dimensions of analysis (buyer, seller, time and item sold). Two of these dimensions contain just one level of detail and two other have an aggregation hierarchy with more than one level. For instance, the time dimension has three levels of detail which contain, at the same

Related Content

E-Readiness in IT/IS Implementation: A Benchmarking Analysis Based on 100 Case Studies

Ayman Altameemand Mohamed Zairi (2012). *Cases on E-Readiness and Information Systems Management in Organizations: Tools for Maximizing Strategic Alignment (pp. 222-254).*

www.irma-international.org/chapter/readiness-implementation-benchmarking-analysis-based/61103

Making the Business Process Execution Language (BPEL) Flexible

Daniela Wolffand Nishant Singh (2011). *E-Strategies for Resource Management Systems: Planning and Implementation (pp. 1-20).*

www.irma-international.org/chapter/making-business-process-execution-language/45095

Dynamics of E-Business Infrastructure

Petter Gottschalk (2007). *Business Dynamics in Information Technology (pp. 62-102).*

www.irma-international.org/chapter/dynamics-business-infrastructure/6055

Web Services as XML Data Sources in Enterprise Information Integration

Ákos Hajnal, Tamás Kifor, Gergely Lukácsyand László Z. Varga (2009). *Services and Business Computing Solutions with XML: Applications for Quality Management and Best Processes (pp. 82-97).*

www.irma-international.org/chapter/web-services-xml-data-sources/28969

The e-ERP Transformation Matrix

Colin G. Ashand Janice M. Burn (2005). *Managing Business with SAP: Planning Implementation and Evaluation (pp. 158-178).*

www.irma-international.org/chapter/erp-transformation-matrix/25723