

Chapter 1.16

Conceptual Modeling Solutions for the Data Warehouse

Stefano Rizzi

DEIS – University of Bologna, Italy

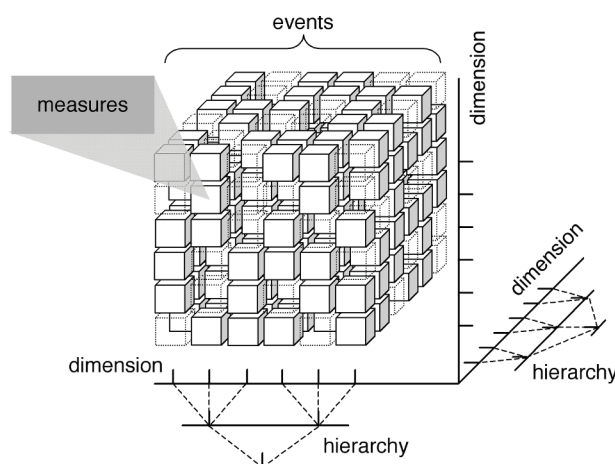
ABSTRACT

In the context of data warehouse design, a basic role is played by conceptual modeling, that provides a higher level of abstraction in describing the warehousing process and architecture in all its aspects, aimed at achieving independence of implementation issues. This chapter focuses on a conceptual model called the DFM that suits the variety of modeling situations that may be encountered in real projects of small to large complexity. The aim of the chapter is to propose a comprehensive set of solutions for conceptual modeling according to the DFM and to give the designer a practical guide for applying them in the context of a design methodology. Besides the basic concepts of multidimensional modeling, the other issues discussed are descriptive and cross-dimension attributes; convergences; shared, incomplete, recursive, and dynamic hierarchies; multiple and optional arcs; and additivity.

INTRODUCTION

Operational databases are focused on recording transactions, thus they are prevalently characterized by an OLTP (online transaction processing) workload. Conversely, data warehouses (DWs) allow complex analysis of data aimed at decision support; the workload they support has completely different characteristics, and is widely known as OLAP (online analytical processing). Traditionally, OLAP applications are based on *multidimensional modeling* that intuitively represents data under the metaphor of a cube whose cells correspond to events that occurred in the business domain (Figure 1). Each event is quantified by a set of measures; each edge of the cube corresponds to a relevant dimension for analysis, typically associated to a hierarchy of attributes that further describe it. The multidimensional model has a twofold benefit. On the one hand, it is close to the way of thinking of data analysts, who are used to the spreadsheet metaphor; therefore it helps users understand data. On the other hand, it supports performance improvement

Figure 1. The cube metaphor for multidimensional modeling



as its simple structure allows designers to predict the user intentions.

Multidimensional modeling and OLAP workloads require specialized design techniques. In the context of design, a basic role is played by *conceptual modeling* that provides a higher level of abstraction in describing the warehousing process and architecture in all its aspects, aimed at achieving independence of implementation issues. Conceptual modeling is widely recognized to be the necessary foundation for building a database that is well-documented and fully satisfies the user requirements; usually, it relies on a graphical notation that facilitates writing, understanding, and managing conceptual schemata by both designers and users.

Unfortunately, in the field of data warehousing there still is no consensus about a formalism for conceptual modeling (Sen & Sinha, 2005). The entity/relationship (E/R) model is widespread in the enterprises as a conceptual formalism to provide standard documentation for relational information systems, and a great deal of effort has been made to use E/R schemata as the input for designing nonrelational databases as well (Fahrner

& Vossen, 1995); nevertheless, as E/R is oriented to support queries that navigate associations between data rather than synthesize them, it is not well suited for data warehousing (Kimball, 1996). Actually, the E/R model has enough expressivity to represent most concepts necessary for modeling a DW; on the other hand, in its basic form, it is not able to properly emphasize the key aspects of the multidimensional model, so that its usage for DWs is expensive from the point of view of the graphical notation and not intuitive (Golfarelli, Maio, & Rizzi, 1998).

Some designers claim to use star schemata for conceptual modeling. A *star schema* is the standard implementation of the multidimensional model on relational platforms; it is just a (denormalized) relational schema, so it merely defines a set of relations and integrity constraints. Using the star schema for conceptual modeling is like starting to build a complex software by writing the code, without the support of and static, functional, or dynamic model, which typically leads to very poor results from the points of view of adherence to user requirements, of maintenance, and of reuse.

18 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/conceptual-modeling-solutions-data-warehouse/7642

Related Content

Sensor Field Resource Management for Sensor Network Data Mining

David J. Yates and Jennifer Xu (2010). *Intelligent Techniques for Warehousing and Mining Sensor Network Data* (pp. 280-304).

www.irma-international.org/chapter/sensor-field-resource-management-sensor/39550

A Framework for Efficient Association Rule Mining in XML Data

Ji Zhang, Han Liu, Tok Wang Ling, Robert M. Bruckner and A. Min Tjoa (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (pp. 509-529).

www.irma-international.org/chapter/framework-efficient-association-rule-mining/7662

Efficient Computation of Data Cubes and Aggregate Views

Leonardo Tininini (2005). *Encyclopedia of Data Warehousing and Mining* (pp. 421-426).

www.irma-international.org/chapter/efficient-computation-data-cubes-aggregate/10634

Analyses and Evaluation of Responses to Slowly Changing Dimensions in Data Warehouses

Lars Frank and Christian Frank (2009). *Progressive Methods in Data Warehousing and Business Intelligence: Concepts and Competitive Analytics* (pp. 324-337).

www.irma-international.org/chapter/analyses-evaluation-responses-slowly-changing/28174

Metric Methods in Data Mining

Dan A. Simovici (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (pp. 849-879).

www.irma-international.org/chapter/metric-methods-data-mining/7678