Chapter I

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 analyzers, who are used to the spreadsheet metaphor; therefore it helps users understand data. On the other hand, it supports performance improvement 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 require-