Chapter III

Conceptual Multidimensional Models

Riccardo Torlone
Università Roma Tre, Italy

ABSTRACT

A variety of multidimensional data models have recently been proposed by both academic and industry communities, but consensus on formalism or even a common terminology has not yet emerged. In this chapter, we first discuss the requirements that an ideal conceptual multidimensional model should fulfill. These requirements are suggested by general information system modeling principles and the specific characteristics of OLAP applications. Building on these requirements, we then present a general conceptual multidimensional data model and show how it can be used to describe the basic aspects of a business application in a way that is easy to understand and independent of the criteria for actual data organization in the various systems. Starting from the characteristics of the model proposed, we summarize the general features that a multidimensional conceptual model should support. We then survey various multidimensional models proposed and relate their characteristics to these general features. Finally, we discuss the main points raised in the chapter and some problems that remain to be solved in this context.
The ability to represent information in an abstract and implementation-independent way is crucial in the lifecycle of every information system application—not only in its design but also in its operational phase. This is particularly true in the context of data warehousing and OLAP where, because of the level of complexity, application development and management are usually difficult and error-prone tasks.

In spite of this, conceptual data models for data warehousing have received little attention for a long period in the applicative area. Traditionally, multidimensional applications are modeled in a way that strictly depends on the corresponding implementation. One of the most used formalisms for data representation in this context is the relational model, which is clearly well suited in the case of a ROLAP (Relational OLAP) implementation. In general however, using a logical data model has a number of negative consequences. First, a logical representation is conceived to describe, at the appropriate level of abstraction, how data is stored in a specific DBMS, but it is usually not expressive enough to capture in an effective way the essential, multidimensional aspects of a data warehousing application. Second, it is difficult to define a design methodology that includes a general, conceptual step, independent of any specific system but suitable for all. Finally, in specifying aggregations of data, analysts often need to take care of tedious details that refer to the distribution of the information along the various structures used for its storage. For these reasons, data warehouse developers today understand that conceptual data models and methodologies are fundamental ingredients for the realization of good-quality products and for effective employment of their content.

It is now widely accepted that traditional conceptual data models, such as the Entity-Relationship model, are not appropriate for description of the multidimensional and aggregative nature of OLAP applications. For this reason, a variety of multidimensional data models have recently been proposed by both academic and industry communities, although it should be noted that a consensus on formalism or even a common terminology has not yet emerged.

In this chapter, we first discuss the requirements that an ideal conceptual multidimensional model should fulfill. These requirements are suggested by general information system modeling principles and the specific characteristics of OLAP applications. Building on these requirements, we then present a general conceptual multidimensional data model and show how it can be used to describe the basic aspects of a business application in a way that is easy to understand and independent of the criteria for actual data organization in the various systems. Far from being complete, this model aims at capturing the core of the various proposals of multidimensional data models and the conceptual means adopted by OLAP systems for data representation and manipulation. The model relies on a few agreed-upon concepts. The basic notions are the dimension and the data cube. A dimension represents a business perspective under which data analysis is to be performed and is organized in a hierarchy of levels, which correspond to different ways to group its