Chapter 2
Improving Expressive Power in Modeling Data Warehouse and OLAP Applications

Elzbieta Malinowski
University of Costa Rica, Costa Rica

ABSTRACT

Data warehouse and OLAP systems are widely required during the decision-support process, since they provide integrated data in a form that facilitates the expression of complex queries. In order to exploit both systems to their full capabilities, dimensions with hierarchies must be clearly defined. Dimensions can be of different types and they allow users to see quantified data from different perspectives. Hierarchies are important in analytical applications, since they give users the possibility of representing data at different abstraction levels. However, even though there are different kinds of hierarchies in real-world applications and some of them are already implemented in commercial tools, there is still a lack of a well-accepted conceptual model that allows decision-making users to express their analysis needs. In this chapter, we show how the conceptual multidimensional model can be used to facilitate the representation of complex hierarchies and different kinds of dimensions in comparison to their representation in the relational model and commercial OLAP tools, using as an example Microsoft Analysis Services.

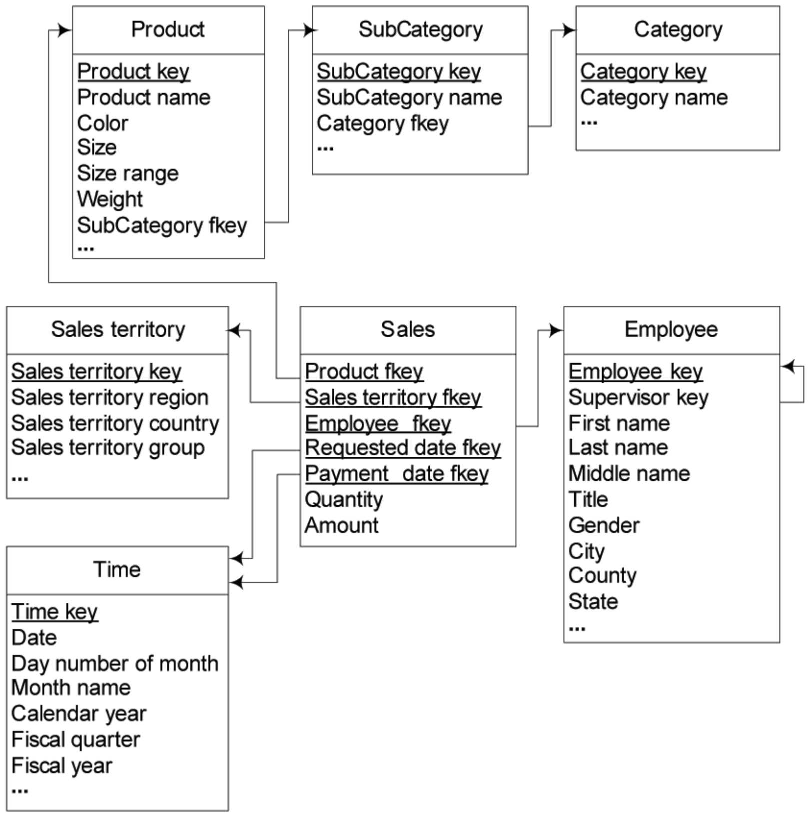
INTRODUCTION

A Data Warehouse (DW) provides users with high quality data organized in a way that facilitates expression of complex queries, ensuring at the same time efficient and accurate responses to such queries. Different systems and tools, such as **online analytical processing** (OLAP) systems, can be used to access and analyze the data contained in DWs. These systems allow users to interactively query and automatically aggregate data using roll-up and drill-down operations. The former operation transforms detailed data into a summarized one, e.g., daily sales into monthly sales, while the latter operation does the contrary.

The data for DW and OLAP systems is usually organized into fact tables related to several dimension tables. A **fact table** (*Sales* in Figure

DOI: 10.4018/978-1-60566-816-1.ch002

Figure 1. Example of a DW for analyzing employees' sales



1) represents the focus of analysis (e.g., analysis of employees' sales) and typically includes attributes called **measures**. These are usually numeric values (e.g., *Quantity* and *Amount* in Figure 1) that facilitate a quantitative evaluation of various aspects of interest. **Dimensions** (e.g., *Sales territory* in Figure 1) are used to see the measures from different perspectives, e.g., according to geographic distribution of a company. Dimensions typically include attributes that form **hierarchies**. When a hierarchy is traversed from finer to coarser levels, measures are aggregated, e.g., moving in a hierarchy from a product to a subcategory will give aggregated values of sales for different products subcategories.

Hierarchies can be included in a flat table (e.g., attributes *City*-*County*-*State* in the *Employee* table in Figure 1) forming the so-called **star schema** or using a normalized structure (e.g., tables *Product*, *SubCategory*, and *Category* in the figure), called the **snowflake schema**.

In order to exploit OLAP systems to their fullest capabilities hierarchies must be clearly defined. Hierarchies are important in analytical applications, since they represent data at different abstraction levels. However, in real-world situations, users must deal with different kinds of hierarchies that either cannot be represented using the current DW and OLAP systems or are represented at the logical level without the possibility of capturing the essential semantics of multidimensional applications. For example, the *Employee* table includes a hierarchy that represents the supervisor-supervisee relationship (the attributes *Employee key* and *Supervisor key*); this hierarchy is difficult to distinguish even though it may be important to consider during the analysis process. Another hierarchy in the same table can be

Related Content