Available online at www.sciencedirect.com
ScienceDirect
Data & Knowledge Engineering 59 (2006) 348–377
DATA & KNOWLEDGE ENGINEERING
www.elsevier.com/locate/datak

Hierarchies in a multidimensional model: From conceptual modeling to logical representation

E. Malinowski *, E. Zimányi

Department of Computer and Network Engineering, Université Libre de Bruxelles, 1050 Brussels, Belgium

Received 26 August 2005; received in revised form 26 August 2005; accepted 13 September 2005
Available online 10 October 2005

Abstract

Hierarchies are used in data warehouses (DWs) and on-line analytical processing (OLAP) systems to see data at different levels of detail. However, many kinds of hierarchies arising in real-world situations are not addressed by current OLAP systems. Further, there is still no agreement on a conceptual model for DW and OLAP design that offers both a graphical representation and a formal definition.

In this paper, we formally define the MultiDimER model, a conceptual multidimensional model that allows to represent facts with measures as well as the different kinds of hierarchies already classified in our previous work [E. Malinowski, E. Zimányi, OLAP hierarchies: a conceptual perspective, in: Proceedings of the 16th International Conference on Advanced Information Systems Engineering, 2004, pp. 477–491]. We also present the mapping of such hierarchies to the relational model, as well as their implementation in commercial DW products.

Keywords: Data warehouses; OLAP hierarchies; Logical design; Data modeling; Conceptual multidimensional model

1. Introduction

The importance of data analysis has grown significantly in recent years as businesses in all sectors have discovered the competitive advantage that just-in-time information can give for the decision-making process. A *data warehouse* (DW) is typically used for this process, since it is a large data repository with integrated historical data organized specifically for analytical purposes.

Bill Inmon defined a data warehouse as a collection of subject-oriented, integrated, non-volatile, and time-variant data to support management's decisions [18]. Subject orientation means that the development of DWs is done according to the analytical necessities of managers on different levels of the decision-making process. Integration represents the complex effort to join data from different operational and external systems and to solve the problems of the definition and content of data, such as differences in data format, data codification, among others. Non-volatility ensures data durability, thus expanding data utility for a longer period of time

* Corresponding author.
E-mail addresses: emalinow@ulb.ac.be (E. Malinowski), ezimanyi@ulb.ac.be (E. Zimányi).

0169-023X/$ - see front matter
doi:10.1016/j.datak.2005.08.003

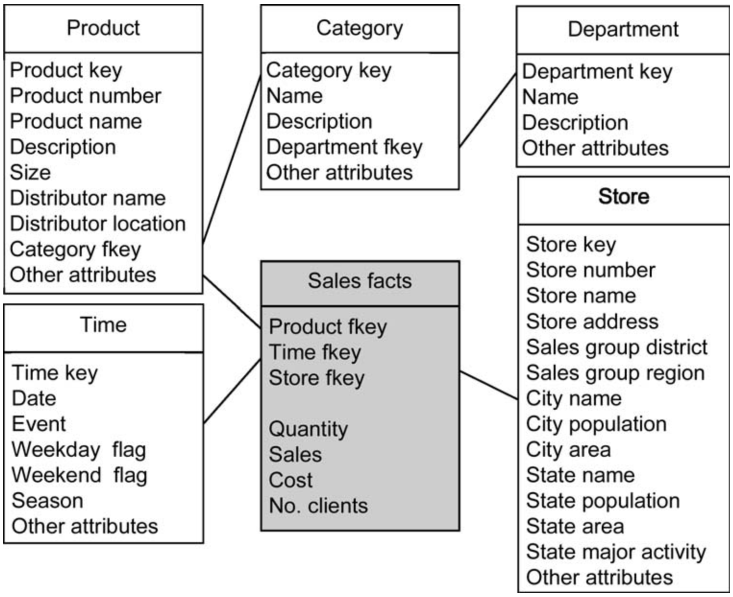

Fig. 1. Example of a snowflake schema.

than operational systems can usually offer. Time-variation indicates the possibility to keep different values of the same object according to its changes in time.

The structure of a DW is usually represented using a *star* or a *snowflake schema*, based on a multidimensional view of data, consisting of fact tables, dimension tables, and hierarchies. An example is given in Fig. 1 [21]. A *fact table* (Sales facts in Fig. 1) represents the subject orientation and the focus of analysis. It typically contains *measures* that are attributes representing the specific elements of analysis (such as quantity sold or sales). A *dimension* contains attributes that allow to explore measures from different perspectives. These attributes can either form a hierarchy (Product–Category–Department in Product dimension in Fig. 1) or be descriptive (Store number in Store dimension in Fig. 1).

On the other hand, OLAP systems allow decision-making users to dynamically manipulate the data contained in a DW. Regardless which kind of storage is used, i.e., relational or multidimensional, OLAP systems also require the specification of facts, dimensions, and hierarchies in order to create a cube. In particular, hierarchies allow the user to start from a general view of data and obtain a detailed view with the *drill-down* operation. Alternatively, the *roll-up* operation allows to transform detailed measures into summarized data.

Nevertheless, the types of hierarchies managed by current OLAP tools are very restrictive when modeling real-world situations. Usually, these tools only cope with hierarchies that ensure summarizability[1] [23]. However, many real-world hierarchies do not satisfy these summarizability conditions. For example, while the geographical division of a country may consist in the hierarchy of City–County–State, some states may not have counties. Therefore, measures aggregated for the State level will be taken from two different levels, the County and the City levels.

In this paper, we present MultiDimER, a conceptual multidimensional model based on the ER model that includes constructs for DW and OLAP modeling. Conceptual models are closer to the way users perceive an application domain than logical models. Nevertheless, the domain of conceptual design for multidimensional modeling is still at a research stage. The analysis presented in [37] shows the little interest of the research community in conceptual multidimensional modeling. The proposed models either provide a graphical representation based on the ER model or UML notations with little or no formal definition or only provide a formal definition without any graphical support. Moreover, some of the models (e.g., [5,10,41]) do not cope with

[1] A correct aggregation of measures in a higher hierarchy level taking into account existing aggregations in a lower hierarchy level.