

Chapter XXII

Empowering the OLAP Technology to Support Complex Dimension Hierarchies

Svetlana Mansmann

University of Konstanz, Germany

Marc H. Scholl

University of Konstanz, Germany

ABSTRACT

Comprehensive data analysis has become indispensable in a variety of domains. OLAP (On-Line Analytical Processing) systems tend to perform poorly or even fail when applied to complex data scenarios. The restriction of the underlying multidimensional data model to admit only homogeneous and balanced dimension hierarchies is too rigid for many real-world applications and, therefore, has to be overcome in order to provide adequate OLAP support. We present a framework for classifying and modeling complex multidimensional data, with the major effort at the conceptual level as to transform irregular hierarchies to make

them navigable in a uniform manner. The properties of various hierarchy types are formalized and a two-phase normalization approach is proposed: heterogeneous dimensions are reshaped into a set of well-behaved homogeneous subdimensions, followed by the enforcement of summarizability in each dimension's data hierarchy. Mapping the data to a visual data browser relies solely on metadata, which captures the properties of facts, dimensions, and relationships within the dimensions. The navigation is schema-based, that is, users interact with dimensional levels with on-demand data display. The power of our approach is exemplified using a real-world study from the domain of academic administration.

INTRODUCTION

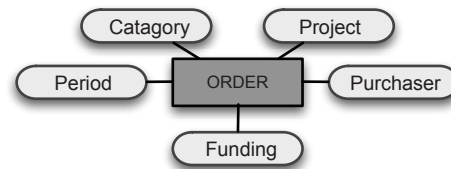
Data warehouse technology introduced in the early 90s to support data analysis in business environments has recently reached out to non-business domains such as medicine, education, research, government, etc. End-users interact with data using advanced visual interfaces that enable intuitive navigation to the desired data subset and granularity and provide a visually enhanced presentation using a variety of visualization techniques.

Data warehouse systems adopt a *multidimensional data model* tackling the challenges of the online analytical processing (OLAP) (Codd, Codd, & Salley, 1993) via efficient execution of queries that aggregate over large amount of detailed data. The analysis is preceded by a highly complex ETL (extract, transform, load) process of integrating the data from multiple systems and bringing it into a consistent state.

In relational OLAP systems, multidimensional views of data, or *data cubes*, are structured using a *star* or a *snowflake* schema consisting of fact tables and dimension hierarchies. *Fact tables* contain data records (*facts*) such as transactions or events, which represent the focus of the analysis. Facts are composed of two types of attributes: (1) *measures* (i.e., the actual elements of the analysis), and 2) *dimensions*, which uniquely determine the measures and serve as exploration axes for aggregation. Members of a dimension are typically organized in a containment type hierarchy to support multiple granularities. In the dimension table, the attributes that form the hierarchy are called *dimension levels*, or *categories*. Other descriptive attributes belonging to a particular category are known as *property* attributes. Dimension levels along with parent/child relationships between them are referred to as the dimension's *intension*, or *schema*, whereas the hierarchy of its members forms its *extension*.

Figure 1 shows the star schema view of a data cube storing the administrative expenditures of

Figure 1. Star schema view of data



a university: the facts in the fact table ORDER are determined by five dimensions. In the star schema, the whole dimension hierarchy is placed into a single table, whereas the snowflake schema enforces the hierarchy to be decomposed into separate tables, one table per dimension level.

The two logical design options are illustrated in Figure 2 at the example of the dimension Period. The star schema produces a single table period with all dimension levels and property attributes. Obviously, in such denormalized view it is impossible to explicitly recognize the hierarchical relationships. In the snowflake schema, however, each dimension category with its property attributes is placed into a separate table referencing its parent. The arrows correspond to the foreign keys (i.e., the roll-up relationships between the levels). The resulting schema is rather complex, but it offers the advantage of automatic extraction of the hierarchy schema with all valid aggregation paths from the foreign key constraints. Notice that reoccurring intervals such as weeks, months, quarters, etc. are presented by a two-category lattice (e.g., months → month) in order to be able to roll-up single instances to the instance's type. For example, months instances "January 1997" and "January 1998" rollup to month instance "January."

Summarizability and Homogeneity

The rigidity of the standard OLAP technology is caused primarily by the enforcement of summarizability for all dimensional hierarchies. The concept of summarizability, coined by Rafanelli

19 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/empowering-olap-technology-support-complex/28566

Related Content

Framework for a Geographic Districting DSS using an Intelligent Object-oriented Model

Ramesh Subramanian and Minnie Yen (1997). *Journal of Database Management* (pp. 3-15).

www.irma-international.org/article/framework-geographic-districting-dss-using/51180

Techniques for Sampling Online Text-Based Data Sets

Lynne M. Webb and Yuanxin Wang (2014). *Big Data Management, Technologies, and Applications* (pp. 95-114).

www.irma-international.org/chapter/techniques-for-sampling-online-text-based-data-sets/85452

Materialized Views in Multidimensional Databases

Stefano Paraboschi, Giuseppe Sindoni, Elena Baralis and Ernst Teniente (2003). *Multidimensional Databases: Problems and Solutions* (pp. 222-251).

www.irma-international.org/chapter/materialized-views-multidimensional-databases/26970

It's Not My Fault: The Transfer of Information Security Breach Information

Tawei Wang, Yen-Yao Wang and Ju-Chun Yen (2019). *Journal of Database Management* (pp. 18-37).

www.irma-international.org/article/its-not-my-fault/234276

Graph Representation and Anonymization in Large Survey Rating Data

Xiaoxun Sun and Min Li (2012). *Graph Data Management: Techniques and Applications* (pp. 315-333).

www.irma-international.org/chapter/graph-representation-anonymization-large-survey/58617