Chapter II

Multidimensionality in Statistical, OLAP, and Scientific Databases

Arie Shoshani
Lawrence Berkeley National Laboratory, USA

ABSTRACT

The term “multidimensional databases” refers to data that can be viewed conceptually in a multidimensional space, where each dimension represents some attributes of the data. Viewing data in this form is natural for many applications, yet the concepts are not treated in a uniform way in the database literature. In this chapter, we show the commonality of concepts between three database areas: statistical, OLAP, and scientific databases. We show that these domains have two main structural concepts: the cross-product space of the dimensions, and the classification hierarchy structure associated with each dimension. In the first part of this chapter we describe how these structures are used to represent data in statistical and OLAP databases and how summarization operators can be applied to them. Further, we discuss how these structures can be extended to represent related information using federated database concepts. In the second part of the chapter we show that these concepts are common to many scientific database applications. In particular, we discuss the importance of supporting classification structures and the difficulty in representing them.
INTRODUCTION AND BACKGROUND

There is a lot of data that can be viewed as multidimensional data. The term multidimensional databases typically refers to a collection of objects, each represented as a point in a multidimensional space. Even data that is represented in a tabular form, such as relations, can be thought of as multidimensional data, if each row (tuple) is thought of as an object, and the columns (attributes) are thought of as the dimensions. For example, consider the following table: employee (personID, age, sex, salary) shown in Figure 1a. If each person is represented as a point in the multidimensional space of (age, sex, salary), then that table can be represented as in Figure 1b.

The utility of representing data in the multidimensional space is that it is more natural to view certain features of the data in this way. For example, it is natural to view clusters in the multidimensional space. In Figure 1b, one can easily see that there is a small cluster of highly paid people (perhaps representing managers who are generally older) and a larger cluster of lower paid people. We can also see “outliers” as is the case with the younger person with a high salary. Of course, these concepts extends to data in more than three dimensions, but cannot be viewed as easily. The

As tables in relational databases. We also discuss data structures to support multidimensional databases, emphasizing space-time representation, clustering in multidimensional space, indexing in multidimensional space, and supporting classification structures. We conclude by arguing that the concepts of multidimensionality and classification structures as well as the operation over them should be elevated to “first class” object types. These object types should be visible by the application user explicitly in the conceptual schemas as well as exposing them in the user interfaces.
Related Content

Preferred Repairs for Inconsistent Databases
www.irma-international.org/chapter/preferred-repairs-inconsistent-databases/11192/

Shape Matching for Foliage Database Retrieval
Haibin Ling and David W. Jacobs (2009). Semantic Mining Technologies for Multimedia Databases (pp. 100-129).
www.irma-international.org/chapter/shape-matching-foliage-database-retrieval/28830/

On Non-Constrained, Constrained and Mandatory Many-to-Many Relationship Types
www.irma-international.org/article/non-constrained-constrained-mandatory-many/51113/

Factors Influencing the Lifetime of Telecommunication and Information Technology Standards: Results of an Explorative Analysis of the PERINORM Database
www.irma-international.org/chapter/factors-influencing-lifetime-telecommunication-information/7990/

Metrics for Workflow Design: How an Information Processing View on Business Processes Helps to Make Good Designs
www.irma-international.org/chapter/metrics-workflow-design/4355/