Chapter 3.7
Constraint-Based Multi-Dimensional Databases

Franck Ravat
Université Toulouse I, France

Olivier Teste
Université Toulouse III, France

Gilles Zurfluh
Université Toulouse I, France

ABSTRACT

This chapter deals with constraint-based multi-dimensional modelling. The model we define integrates a constellation of facts and dimensions. Along each dimension, various hierarchies are possibly defined and the model supports multiple instantiations of dimensions. The main contribution is the definition of intra-dimension constraints between hierarchies of a same dimension as well as inter-dimension constraints of various dimensions. To facilitate data querying, we define a multi-dimensional query algebra, which integrates the main multi-dimensional operators such as rotations, drill down, roll up… These operators support the constraint-based multi-dimensional modelling. Finally, we present two implementations of this algebra. First, OLAP-SQL is a textual language integrating multi-dimensional concepts (fact, dimension, hierarchy), but it is based on classical SQL syntax. This language is dedicated to specialists such as multi-dimensional database administrators. Second, a graphical query language is presented. This language consists in a graphical representation of multi-dimensional databases, and users specify directly their queries over this graph. This approach is dedicated to non-computer scientist users.

INTRODUCTION

OnLine Analytical Processing (OLAP) has emerged to support multi-dimensional data analysis by providing manipulations through aggregations of data drawn from various transactional databases. This approach is often based on multi-dimensional databases. The multi-dimensional modelling (Kimball, 1996) represents data as points in multi-dimensional space. Data are viewed as a subject of analysis (fact) associated to axis of analysis (dimensions). Each dimension contains one or several viewpoints of analysis (hierarchies) representing data granularities. For example, *sale amounts* could be analysed according to *time*, *stores*, and *customers*. Along *store* dimension, a hierarchy could group *individual stores* into *cities*, which are grouped into *states* or *regions*, which are grouped into *countries*.

This approach induces topics of interests for the scientific community (Rafanelli, 2003). The main issues focus on technologies and tools that enable the business intelligence lifecycle from data modelling and acquisition to knowledge extraction. These problems are based on researches, which deal with design methods, multi-dimensional models, OLAP query languages, and tools that facilitate data extraction and data warehousing. Multi-dimensional data are crucial for the decision-making. Nevertheless, only a few researches focus on multi-dimensional data integrity (Hurtado & Mendelzon, 2002).

The confidence in a multi-dimensional database lies in its capacity to supply relevant information. A multi-dimensional model integrating constraints must provide an accurate model of the organisation activities, and it allows valid data restitution (Hurtado & Mendelzon, 2002). This chapter deals with constraint-based multi-dimensional modelling and querying.

The chapter outline is composed of the following sections. The second section gives an overview of related works. The third section defines a constellation model where dimensions support multiple instantiations as well as multiple hierarchies. The fourth section specifies a query algebra. We show the effect of constraints during multi-dimensional manipulations. The fifth section presents a tool named Graphic-OLAPSQL. It supports two languages: a SQL-like language for administrators, and a graphical language dedicated to casual users as decision makers.

BACKGROUND

Efforts to model multi-dimensional databases have followed two directions; some models extend relational approaches to integrate multi-dimensional data, and others approaches model directly and more naturally multi-dimensional data. Multi-dimensional databases are manipulated through data cubes.

- In the relational context, the data cube operator (Gray, Bosworth, Layman, & Pirahesh, 1996) was introduced to expand relational tables by computing the aggregations over all the attribute combinations. Kimball (1996) introduces multi-dimensional models based on dimension tables and fact tables, whereas Li and Wang (1996) represent cubes through dimension relations and functions, which map measures to grouping relations. Barralis, Paraboschi, and Teniente (1997) consider multi-dimensional databases as a set of tables forming de-normalised star schemata.
- To integrate more naturally multi-dimensional data, Agrawal, Gupta, and Sarawagi (1997) introduce a model, which supports a symmetric treatment of dimensions and measures, and it provides a set of operators (manipulation of cubes). Several approaches were presented that support cubes with n-dimensions (Gyssen & Lakshmanan, 1997) and cubes integrating explicitly multiple hierarchies (Agrawal, Gupta, & Sarawagi, 1997; Lehner, 1998; Mendelzon & Vaisman, 2000; Vassiliadis & Sellis, 1999). Cabbibo and Torlone (1997) define a multi-dimensional database through dimensions, which are constructed from hierarchies of dimension levels, and f-tables, which store

Related Content