Handling Structural Heterogeneity in OLAP

Carlos A. Hurtado
Universidad de Chile, Chile

Claudio Gutierrez
Universidad de Chile, Chile

Abstract

Structural heterogeneous OLAP data arise when several OLAP dimensions with different structures are mixed into a single OLAP dimension. In this chapter, we examine the problems encountered when handling structural heterogeneity in OLAP and survey techniques that have been proposed to solve them. We show how to incorporate structural heterogeneity in the design of OLAP models. We explain why structural heterogeneity weakens aggregate navigation, the framework that guides users to formulate correct OLAP operations and systems to efficiently process them. We survey different techniques to deal with heterogeneity, including the modeling of heterogeneity by unbalanced dimensions, the solution proposed by Kimball, and the use of null elements to fix heterogeneity. Finally, we present a class of integrity constraints to model structural heterogeneity, called dimension constraints, introduced in previous work of the authors. We show the practical application of dimension constraints to support aggregate navigation and some of the aforementioned techniques for dealing with the problem.
Introduction

Much of the success of OLAP can be attributed to the intuitive approach to data visualization provided by the multidimensional data model. Nowadays, the notions of facts and dimensions have been largely disseminated among database practitioners and researchers and have been proved to be useful metaphors to support querying data for decision support. The simplicity of the multidimensional model, however, stands on some assumptions about the regularity of data which are unnatural in many applications. In this chapter, we study the implications of relaxing one of the cores of such assumptions, namely the homogeneity of the structure of OLAP dimensions. Structurally heterogeneous OLAP data have been reported in the OLAP literature almost since the origins of the term OLAP itself and have concentrated significant research work since then (Hurtado, Gutierrez, & Mendelzon, 2005; Huseman, Lechtenborger, & Vossen, 2000; Jagadish, Lakshmanan, & Srivastava, 1999; Kimball, 1996; Lehner, Albrecht, & Wedekind, 1998; Malinowski & Zimanyi, 2004; Pedersen, Jensen, & Dyreson, 2001).

Motivation

In the multidimensional data model, dimensions represent the perspectives upon which data is viewed, and facts represent events that associate points of such dimensions to measures. For example, a sale of a particular product in a particular store of a retail chain can be viewed as a fact, which may be represented as a point in a space whose dimensions are products, stores, and time, and can be associated with one or more measures such as price or profit.

The phenomenon we study in this chapter is related to OLAP dimensions and, more precisely, to their structure. The structure of a dimension is modeled as a hierarchy of categories. Each category represents a level of abstraction upon which facts are aggregated. For example, in a dimension that models the products of a retailer, shown in Figure 1, we have a category Product which rolls up to a Brand category, which in turn rolls up to the top category All. The elements of the dimensions are grouped into the categories and connected by a child/parent relationship, which yields a hierarchy of elements which parallels the hierarchy of categories. Following terminology from Jagadish et al. (1999) and from Hurtado et al. (2005), we refer to the hierarchies of categories and elements respectively as hierarchy schema and hierarchy domain.

Each element of a dimension can be viewed as having a structure on its own. This structure of an element is the subgraph of the hierarchy schema induced by the ancestors of that element and their child/parent relationship. In our example of Figure 1, the Product element $p_1$ has the entire hierarchy schema as structure, and
Related Content

Techniques for Weighted Clustering Ensembles
www.igi-global.com/chapter/techniques-weighted-clustering-ensembles/11081?camid=4v1a

Temporal Association Rule Mining in Event Sequences
www.igi-global.com/chapter/temporal-association-rule-mining-event/10760?camid=4v1a

Action Rules Mining
www.igi-global.com/chapter/action-rules-mining/10789?camid=4v1a
Mass Informatics in Differential Proteomics
www.igi-global.com/chapter/mass-informatics-differential-proteomics/10971?camid=4v1a