Chapter 14

Using OCL to Model Constraints in Data Warehouses

François Pinet
Cemagref - Clermont Ferrand, France

Myoung-Ah Kang
LIMOS, France

Kamal Boulil
Cemagref - Clermont Ferrand, France

Sandro Bimonte
Cemagref - Clermont Ferrand, France

Gil De Sousa
Cemagref - Clermont Ferrand, France

Catherine Roussey
Cemagref - Clermont Ferrand, France

Michel Schneider
LIMOS, France

Jean-Pierre Chanet
Cemagref - Clermont Ferrand, France

ABSTRACT

Recent research works propose using Object-Oriented (OO) approaches, such as UML to model data warehouses. This paper overviews these recent OO techniques, describing the facts and different analysis dimensions of the data. The authors propose a tutorial of the Object Constraint Language (OCL) and show how this language can be used to specify constraints in OO-based models of data warehouses. Previously, OCL has been only applied to describe constraints in software applications and transactional databases. As such, the authors demonstrate in this paper how to use OCL to represent the different types of data warehouse constraints. This paper helps researchers working in the fields of business intelligence and decision support systems, who wish to learn about the major possibilities that OCL offer in the context of data warehouses. The authors also provide general information about the possible types of implementation of multi-dimensional models and their constraints.

1. INTRODUCTION

Along with the development of new information and communication technologies, we have seen an increase in data. In order to analyze the data issued from different sources, it is needed to use tools to integrate this information. Data warehouses are a specific type of database that serves to integrate and analyze data from various sources (Cali, Lembo, Lenzerini, & Rosati, 2003). Information stored in different databases can be group together in a data warehouse for combined analysis. De-
Using OCL to Model Constraints in Data Warehouses

pending on their requirements, one can load data every week, every month, every year or even less frequently. These data are usually organized in a form that speeds up calculation of indicators. The indicators are made up of aggregated information obtained by aggregation functions such as sum, average, variance, etc. Using data warehouses is therefore important within a decision-making context. For example, a data warehouse containing economic, urban and environmental information will help decision-makers find the best place to establish a new infrastructure. The concept of the data warehouse has great potential for assessing the impact of actions, practices, scenarios and programs (Bimonte, 2010; Mahboubi, Bimonte, Faure, & Pinet, 2010; Nilakanta, Scheibe, & Rai, 2008; Schneider, 2008; Schulze, Spilke, & Lehner, 2007).

We present an example issued from (Trujillo, Palomar, Gomez, & Song, 2001). The facts of a data warehouse are the information to analyze (Malinowski & Zimanyi, 2008a). In the example, we consider the facts of the data warehouse to be the product sales of a company in Euros. Each of the company’s stores provides these data. In a data warehouse, an analysis results from the use of an aggregation operation (e.g., sum or average) on the facts. In the example, the value of the facts (i.e., the sales) to analyze is the amount of sales. This value is called the measure. A possible analysis is the sum of sales calculated by category of product, by store and by month. The result of this analysis can be represented in a cube (Trujillo et al., 2001) (Figure 1). Each dimension of the cube corresponds to a criterion of analysis: type of products, store and month. The cells of the cube are the indicators. They store the sums of sales for each tuple <type of products, store, month>. For instance, in Figure 1, the sum of sales for the tuple <Water, Store 1, December> is 4. In data warehouses, the criteria of analysis are structured in hierarchies called dimensions. Figure 2 shows the three dimensions presented by Trujillo et al. (2001). A data warehouse can produce many analyses by combining different levels of dimensions. For example, other cubes could be calculated:

- Sums of sales by city,
- Sums of sales by brand, by city, by year,
- Sums of sales by type, by state, by season, etc.

Note that data warehouses generally support $n$-dimensional cubes. Data can be combined to provide previously unknown causal links. To do so, users can visualize cubes from the data warehouse using tools like OLAP (On-line Analytical Pro-

Figure 1. Example of a data warehouse - cube storing sales by category of products, by store and by month