Chapter XI
Methodology for Improving Data Warehouse Design using Data Sources Temporal Metadata

Francisco Araque
University of Granada, Spain

Alberto Salguero
University of Granada, Spain

Cecilia Delgado
University of Granada, Spain

ABSTRACT

One of the most complex issues of the integration and transformation interface is the case where there are multiple sources for a single data element in the enterprise Data Warehouse (DW). There are many facets due to the number of variables that are needed in the integration phase. This chapter presents our DW architecture for temporal integration on the basis of the temporal properties of the data and temporal characteristics of the data sources. If we use the data arrival properties of such underlying information sources, the Data Warehouse Administrator (DWA) can derive more appropriate rules and check the consistency of user requirements more accurately. The problem now facing the user is not the fact that the information being sought is unavailable, but rather that it is difficult to extract exactly what is needed from what is available. It would therefore be extremely useful to have an approach which determines whether it would be possible to integrate data from two data sources (with their respective data extraction methods associated). In order to make this decision, we use the temporal properties of the data, the temporal characteristics of the data sources, and their extraction methods. In this chapter, a solution to this problem is proposed.
INTRODUCTION

The ability to integrate data from a wide range of data sources is an important field of research in data engineering. Data integration is a prominent theme in many areas and enables widely distributed, heterogeneous, dynamic collections of information sources to be accessed and handled.

Many information sources have their own information delivery schedules, whereby the data arrival time is either predetermined or predictable. If we use the data arrival properties of such underlying information sources, the Data Warehouse Administrator (DWA) can derive more appropriate rules and check the consistency of user requirements more accurately. The problem now facing the user is not the fact that the information being sought is unavailable, but rather that it is difficult to extract exactly what is needed from what is available.

It would therefore be extremely useful to have an approach which determines whether it would be possible to integrate data from two data sources (with their respective data extraction methods associated). In order to make this decision, we use the temporal properties of the data, the temporal characteristics of the data sources and their extraction methods. Notice that we are not suggesting a methodology, but an architecture. Defining a methodology is absolutely out of the scope of this paper, and the architecture does not impose it.

It should be pointed out that we are not interested in how semantically equivalent data from different data sources will be integrated. Our interest lies in knowing whether the data from different sources (specified by the DWA) can be integrated on the basis of the temporal characteristics (not in how this integration is carried out).

The use of DW and Data Integration has been proposed previously in many fields. In (Haller, Proll, Rettschitzger, Tjoa, & Wagner, 2000) the Integrating Heterogeneous Tourism Information data sources is addressed using three-tier architecture. In (Moura, Pantoquillo, & Viana, 2004) a Real-Time Decision Support System for space missions control is put forward using Data Warehousing technology. In (Oliva & Saltor, A Negotiation Process Approach for Building Federated Databases, 1996) a multi-level security policies integration methodology to endow tightly coupled federated database systems with a multi-level security system is presented. In (Vassiliadis, Quix, Vassiliou, & Jarke, 2001) a framework for quality-oriented DW management is exposed, where special attention is paid to the treatment of metadata. The problem of the little support for automatized tasks in DW is considered in (Thalhammer, Schrefl, & Mohania, 2001), where the DW is used in combination with event/condition/action (ECA) rules to get an active DW. Finally, in (March & Hevner, 2005) an integrated decision support system from the perspective of a DW is exposed. Their authors state that the essence of the data warehousing concept is the integration of data from disparate sources into one coherent repository of information. Nevertheless, none of the previous works encompass the aspects of the integration of the temporal parameters of data.

In this chapter a solution to this problem is proposed. Its main contributions are: a DW architecture for temporal integration on the basis of the temporal properties of the data and temporal characteristics of the sources, a Temporal Integration Processor and a Refreshment Metadata Generator, that will be both used to integrate temporal properties of data and to generate the necessary data for the later DW refreshment.

Firstly, the concept of DW and the temporal concepts used in this work and our previous related works are revised; following our architecture is presented; following section presents whether data from two data sources with their data extraction methods can be integrated. Then we describe the proposed methodology with its corresponding algorithms. Finally, we illustrate the proposed methodology with a working example.
19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage: [www.igi-global.com/chapter/methodology-improving-data-warehouse-design/28169?camid=4v1](www.igi-global.com/chapter/methodology-improving-data-warehouse-design/28169?camid=4v1)


Related Content

Some Issues in Design of Data Warehousing Systems
Ladjel Bellatreche, Kamalakar Karlapalem and Mukesh Mohania (2002). *Data Warehousing and Web Engineering* (pp. 22-76). [www.igi-global.com/chapter/some-issues-design-data-warehousing/7861?camid=4v1a](www.igi-global.com/chapter/some-issues-design-data-warehousing/7861?camid=4v1a)

Analytical Competition for Managing Customer Relations

Graphical Data Mining

Homeland Security Data Mining and Link Analysis