SURVEY ARTICLE

A Survey on Temporal Data Warehousing

Matteo Golfarelli, DEIS - University of Bologna, Italy

Stefano Rizzi, DEIS - University of Bologna, Italy

ABSTRACT

Data warehouses are information repositories specialized in supporting decision making. Since the decisional process typically requires an analysis of historical trends, time and its management acquire a huge importance. In this paper we consider the variety of issues, often grouped under term temporal data warehousing, implied by the need for accurately describing how information changes over time in data warehousing systems. We recognize that, with reference to a three-levels architecture, these issues can be classified into some topics, namely: handling data/schema changes in the data warehouse, handling data/schema changes in the data mart, querying temporal data, and designing temporal data warehouses. After introducing the main concepts and terminology of temporal databases, we separately survey these topics. Finally, we discuss the open research issues also in connection with their implementation on commercial tools.

Keywords: business intelligence; data mart; data warehouse; multidimensional database design; spatiotemporal database

INTRODUCTION

At the core of most business intelligence applications, *data warehousing systems* are specialized in supporting decision making. They have been rapidly spreading within the industrial world over the last decade, due to their undeniable contribution to increasing the effectiveness and efficiency of the decisional processes within business and scientific domains. This wide diffusion was supported by remarkable research results aimed at improving querying performance, at refining the quality of data, and

at outlining the design process, as well as by the quick advancement of commercial tools.

In the remainder of the paper, for the sake of terminological consistency, we will refer to a classic architecture for data warehousing systems, illustrated in Figure 1, that relies on three levels:

1. The *data sources*, that store the data used for feeding the data warehousing systems. They are mainly corporate operational databases, hosted by either relational or legacy platforms, but in some cases they

may also include external web data, flat files, spreadsheet files, etc.

2. The *data warehouse* (also called *reconciled data level*, *operational data store* or *enterprise data warehouse*), a normalized operational database that stores detailed, integrated, clean and consistent data extracted from data sources and properly processed by means of ETL tools.

3. The *data marts*, where data taken from the data warehouse are summarized into relevant information for decision making, in the form of *multidimensional cubes*, to be typically queried by OLAP and reporting front-ends.

Cubes are structured according to the *multidimensional model*, whose key concepts are fact, measure and dimension. A *fact* is a focus of interest for the decisional process; its occurrences correspond to *events* that dynamically occur within the business world. Each event is quantitatively described by a set of numerical *measures*. In the multidimensional model, events are arranged within an n-dimensional space whose axes, called *dimensions* of analysis, define different perspectives for their identification. Dimensions commonly are discrete, alphanumerical attributes that determine the minimum granularity for analyzing facts. Each dimension is the root of a *(roll-up) hierarchy* that includes a set of *levels*, each providing a way of selecting and aggregating events. Each level can be described by a set of *properties*.

As a consequence of the fact that the decisional process typically relies on computing historical trends and on comparing snapshots of the enterprise taken at different moments, one of the main characterizations of data warehousing systems is that of storing historical, non volatile data. Thus, time and its management acquire a huge importance. In this paper we discuss the variety of issues, often grouped under term *temporal data warehousing*, implied by the need for

Figure 1. Three-levels architecture for a data warehousing system
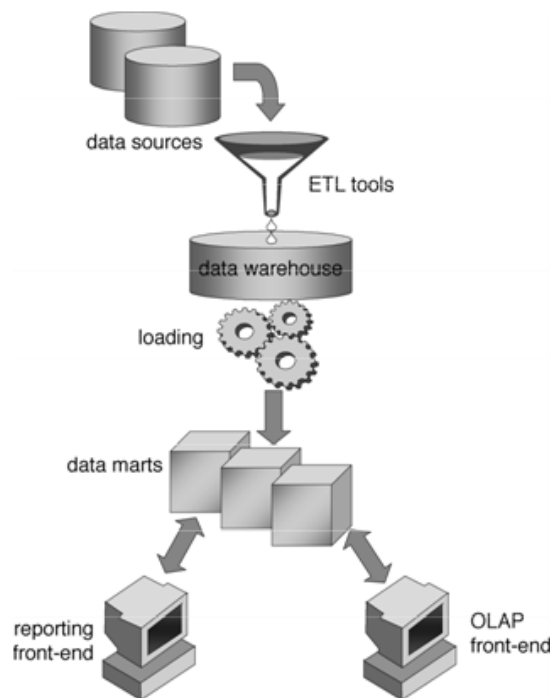
15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/survey-temporal-data-warehousing/1820

Related Content

A Boosting-Aided Adaptive Cluster-Based Undersampling Approach for Treatment of Class Imbalance Problem
Debashree Devi, Suyel Namasudraand Seifedine Kadry (2020). *International Journal of Data Warehousing and Mining (pp. 60-86).*
www.irma-international.org/article/a-boosting-aided-adaptive-cluster-based-undersampling-approach-for-treatment-of-class-imbalance-problem/256163

Recent Developments on Security and Reliability in Large-Scale Data Processing with MapReduce
Christian Espositoand Massimo Ficco (2016). *International Journal of Data Warehousing and Mining (pp. 49-68).*
www.irma-international.org/article/recent-developments-on-security-and-reliability-in-large-scale-data-processing-with-mapreduce/143715

Big Data and Data Modelling for Manufacturing Information Systems
Norman Gwangwava, Khumbulani Mpofuand Samson Mhlanga (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 116-138).*
www.irma-international.org/chapter/big-data-and-data-modelling-for-manufacturing-information-systems/150162

A Cached-Based Stream-Relation Join Operator for Semi-Stream Data Processing
M. Asif Naeem, Imran Sarwar Bajwaand Noreen Jamil (2016). *International Journal of Data Warehousing and Mining (pp. 14-31).*
www.irma-international.org/article/a-cached-based-stream-relation-join-operator-for-semi-stream-data-processing/168484

PSSRC: A Web Service Registration Cloud Based on Structured P2P and Semantics
Qian He, Baokang Zhao, Liang Chang, Jinshu Suand Ilsun You (2016). *International Journal of Data Warehousing and Mining (pp. 21-38).*
www.irma-international.org/article/pssrc/146851