

Chapter 1.17

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.

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.

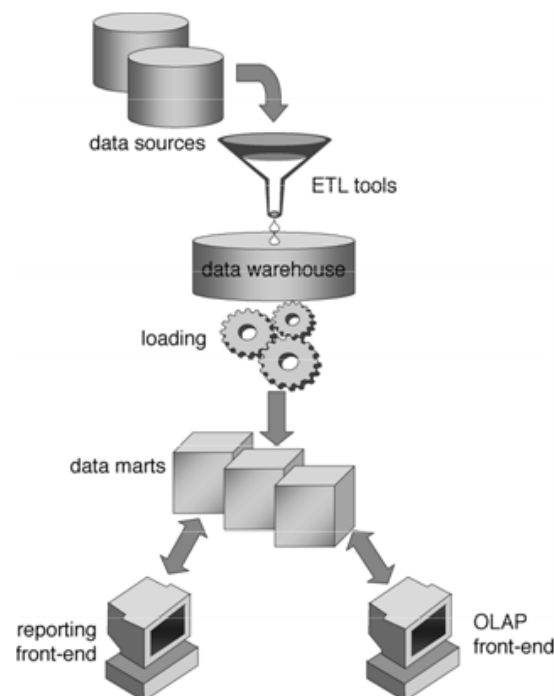
In the remainder of the paper, for the sake of terminological consistency, we will refer to a classic architecture for data warehousing systems, illustrated in Figure 1, that relies on three levels:

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

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/chapter/survey-temporal-data-warehousing/7914

Related Content

A Space-Efficient Protocol for Consistency of External View Maintenance on Data Warehouse Systems: A Proxy Approach

Shi-Ming Huang, David C. Yen and Hsiang-Yuan Hsueh (2007). *Journal of Database Management* (pp. 21-47).

www.irma-international.org/article/space-efficient-protocol-consistency-external/3373

Categorizing Post-Deployment IT Changes: An Empirical Investigation

David Kang (2007). *Journal of Database Management* (pp. 1-24).

www.irma-international.org/article/categorizing-post-deployment-changes/3368

High Speed Optical Higher Order Neural Networks for Discovering Data Trends and Patterns in Very Large Databases

David R. Selviah (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 1084-1107).

www.irma-international.org/chapter/high-speed-optical-higher-order/7960

Towards Code Reuse and Refactoring as a Practice within Extreme Programming

Vijayan Sugumaran and Gerald DeHondt (2009). *Advanced Principles for Improving Database Design, Systems Modeling, and Software Development* (pp. 63-78).

www.irma-international.org/chapter/towards-code-reuse-refactoring-practice/4292

Ubiquitous Computing and Databases

George Roussos and Michael Zoumboulakis (2005). *Encyclopedia of Database Technologies and Applications* (pp. 714-719).

www.irma-international.org/chapter/ubiquitous-computing-databases/11229