

This paper appears in the publication, International Journal of Data Warehousing and Mining, Volume 5, Issue 1 edited by David Taniar © 2009, IGI Global

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

Copyright © 2009, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

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

- 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 *mul-tidimensional 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



Copyright © 2009, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

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/surveytemporal-data-warehousing/1820

Related Content

Adoption of Dual Iris and Periocular Recognition for Human Identification

R. Deepika, M. R. Prasad, Srinivas Chetanaand T. C. Manjunath (2016). *Research Advances in the Integration of Big Data and Smart Computing (pp. 250-266).* www.irma-international.org/chapter/adoption-of-dual-iris-and-periocular-recognition-for-human-identification/139406

Deconstructing Smart Cities

Michael Batty (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 1957-1969).*

www.irma-international.org/chapter/deconstructing-smart-cities/150251

Gaussian-Stacking Multiclassifiers for Human Embryo Selection

Dinora A. Morales, Endika Bengoetxeaand Pedro Larrañaga (2009). Data Mining and Medical Knowledge Management: Cases and Applications (pp. 307-331). www.irma-international.org/chapter/gaussian-stacking-multiclassifiers-human-embryo/7539

Empirical Investigation of Decision Tree Ensembles for Monitoring Cardiac Complications of Diabetes

Andrei V. Kelarev, Jemal Abawajy, Andrew Stranieriand Herbert F. Jelinek (2013). International Journal of Data Warehousing and Mining (pp. 1-18). www.irma-international.org/article/empirical-investigation-of-decision-tree-ensembles-for-monitoringcardiac-complications-of-diabetes/105117

Improvement of Data Stream Decision Trees

Sarah Nait Bahloul, Oussama Abderrahim, Aya Ichrak Benhadj Amarand Mohammed Yacine Bouhedadja (2022). *International Journal of Data Warehousing and Mining (pp. 1-17).*

www.irma-international.org/article/improvement-of-data-stream-decision-trees/290889