An Automatic Data Warehouse Conceptual Design Approach

Jamel Feki
Mir@cl Laboratory, Université de Sfax, Tunisia

Ahlem Nabli
Mir@cl Laboratory, Université de Sfax, Tunisia

Hanêne Ben-Abdallah
Mir@cl Laboratory, Université de Sfax, Tunisia

Faïez Gargouri
Mir@cl Laboratory, Université de Sfax, Tunisia

INTRODUCTION

Within today's competitive economic context, information acquisition, analysis and exploitation became strategic and unavoidable requirements for every enterprise. Moreover, in order to guarantee their persistence and growth, enterprises are forced, henceforth, to capitalize expertise in this domain.

Data warehouses (DW) emerged as a potential solution answering the needs of storage and analysis of large data volumes. In fact, a DW is a database system specialized in the storage of data used for decisional ends. This type of systems was proposed to overcome the incapacities of OLTP (On-Line Transaction Processing) systems in offering analysis functionalities. It offers integrated, consolidated and temporal data to perform decisional analyses. However, the different objectives and functionalities between OLTP and DW systems created a need for a development method appropriate for DW.

Indeed, data warehouses still deploy considerable efforts and interests of a large community of both software editors of decision support systems (DSS) and researchers (Kimball, 1996; Inmon, 2002). Current software tools for DW focus on meeting end-user needs. OLAP (On-Line Analytical Processing) tools are dedicated to multidimensional analyses and graphical visualization of results (*e.g.*, Oracle Discoverer®); some products permit the description of DW and Data Mart (DM) schemes (*e.g.*, Oracle Warehouse Builder®). One major limit of these tools is that the schemes must be built beforehand and, in most cases, manually. However, such a task can be tedious, error-prone and time-consuming, especially with heterogeneous data sources.

On the other hand, the majority of research efforts focuses on particular aspects in DW development, *cf.*, multidimensional modeling, physical design (materialized views (Moody & Kortnik, 2000), index selection (Golfarelli, Rizzi, & Saltarelli 2002), schema partitioning (Bellatreche & Boukhalfa, 2005)) and more recently applying data mining for a better data interpretation (Mikolaj, 2006; Zubcoff, Pardillo & Trujillo, 2007). While these practical issues determine the performance of a DW, other just as important, conceptual issues (*e.g.,* requirements specification and DW schema design) still require further investigations. In fact, few propositions were put forward to assist in and/or to automate the design process of DW, *cf.*, (Bonifati, Cattaneo, Ceri, Fuggetta & Paraboschi, 2001; Hahn, Sapia & Blaschka, 2000; Phipps & Davis 2002; Peralta, Marotta & Ruggia, 2003).

This chapter has a twofold objective. First, it proposes an intuitive, tabular format to assist decision maker in formulating their OLAP requirements. It proposes an automatic approach for the conceptual design of DW/DM schemes, starting from specified OLAP requirements. Our automatic approach is composed of four steps: *Acquisition of OLAP requirements*, *Generation of star/constellation schemes*, *DW schema generation,* and *Mapping the DM/DW onto data sources*. In addition, it relies on an algorithm that transforms tabular OLAP requirements into DM modelled either as a star or a constellation schema. Furthermore, it applies a set of mapping rules between the data sources and the DM schemes. Finally, it uses a set of unification rules

that merge the generated DM schemes and construct the DW schema.

BACKGROUND

There are several proposals to automate certain tasks of the DW design process (Hahn, Sapia & Blaschka, 2000). In (Peralta, Marotta & Ruggia, 2003), the authors propose a rule-based mechanism to automate the construction of the DW logical schema. This mechanism accepts the DW conceptual schema and the source databases. That is, it supposes that the DW conceptual schema already exists. In addition, being a bottom-up approach, this mechanism lacks a conceptual design methodology that takes into account the user requirements which are crucial in the DW design.

In (Golfarelli, Maio & Rizzi, 1998), the authors propose how to derive a DW conceptual schema from Entity-Relationship (E/R) schemes. The conceptual schema is represented by a Dimensional-Fact Model (DFM). In addition, the translation process is left to the designer, with only interesting strategies and cost models presented. Other proposals, similar to (Marotta & Ruggia 2002; Hahn, Sapia & Blaschka, 2000) also generate star schemes and suppose that the data sources are E/R schemes. Although the design steps are based on the operational data sources, the end-users' requirements are neglected. Furthermore, the majority of these works use a graphical model for the Data Source (DS) from which they generate the DM schema; that is, they neither describe clearly how to obtain the conceptual

graphical models from the DS, nor how to generate the multidimensional schemes.

Other works relevant to automated DW design mainly focus on the conceptual design, *e.g.*, (Hüsemann, Lechtenbörger & Vossen 2000) and (Phipps & Davis 2002) who generate the conceptual schema from an E/R model. However, these works do not focus on a conceptual design methodology based on users' requirements and are, in addition, limited to the E/R DS model.

MAIN FOCUS

We propose an automatic approach to design DW/DM schemes from precisely specified OLAP requirements. This approach (Figure 1) is composed of four steps: *i) Acquisition of OLAP requirements* specified as two/n-dimensional fact sheets producing "semi-structured" OLAP requirements, *ii) Generation of star/constellation schemes* by merging the semi-structured OLAP requirements, *iii) DW generation schema* by fusion of DM schemes, and *iv) Mapping the DM/DW* to the data sources.

OLAP Requirement Acquisition

Decisional requirements can be formulated in various manners, but most generally they are expressed in natural language sentences. In our approach, which aims at a computer aided design, we propose to collect these requirements in a format familiar to the decision

Figure 1. DW design starting from OLAP requirements
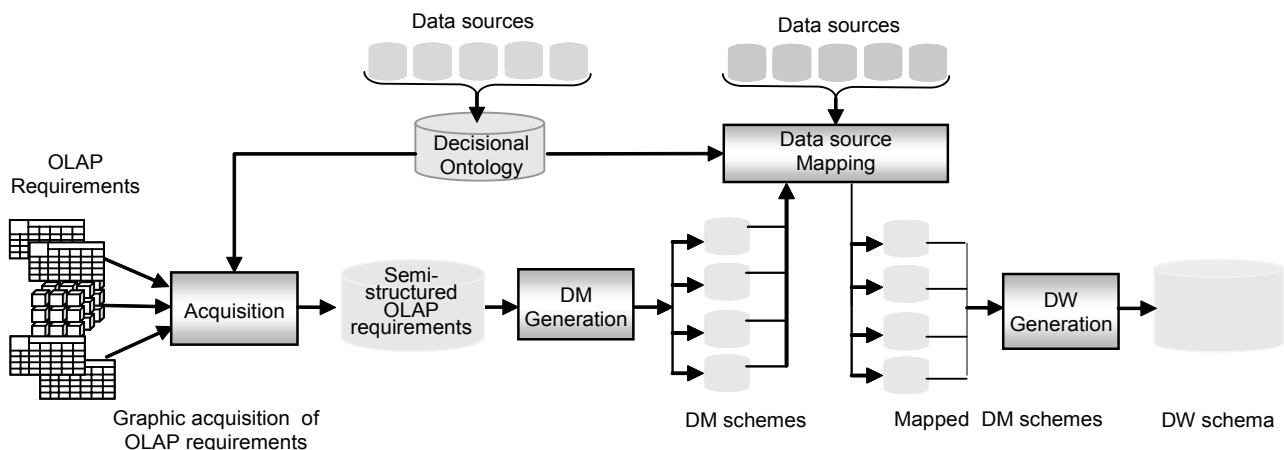
8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/automatic-data-warehouse-conceptual-design/10807

Related Content

Leveraging Unlabeled Data for Classification

Yinghui Yangand Balaji Padmanabhan (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1164-1169).*

www.irma-international.org/chapter/leveraging-unlabeled-data-classification/10969

Video Data Mining

JungHwan Oh (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 2042-2047).*

www.irma-international.org/chapter/video-data-mining/11100

Cluster Analysis with General Latent Class Model

Dingxi Qiuand Edward C. Malthouse (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 225-230).*

www.irma-international.org/chapter/cluster-analysis-general-latent-class/10825

Intelligent Query Answering

Zbigniew W. Rasand Agnieszka Dardzinska (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1073-1078).*

www.irma-international.org/chapter/intelligent-query-answering/10954

Association Bundle Identification

Wenxue Huang, Milorad Krneta, Limin Linand Jianhong Wu (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 66-70).*

www.irma-international.org/chapter/association-bundle-identification/10799