Chapter IX
A Multidimensional Pattern Based Approach for the Design of Data Marts

Hanene Ben-Abdallah
*University of Sfax, Tunisia*

Jamel Feki
*University of Sfax, Tunisia*

Mounira Ben Abdallah
*University of Sfax, Tunisia*

ABSTRACT

Despite their strategic importance, the wide-spread usage of decision support systems remains limited by both the complexity of their design and the lack of commercial design tools. This chapter addresses the design complexity of these systems. It proposes an approach for data mart design that is practical and that endorses the decision maker involvement in the design process. This approach adapts a development technique well established in the design of various complex systems for the design of data marts (DM): Pattern-based design. In the case of DM, a multidimensional pattern (MP) is a generic specification of analytical requirements within one domain. It is constructed and documented with standard, real-world entities (RWE) that describe information artifacts used or produced by the operational information systems (IS) of several enterprises. This documentation assists a decision maker in understanding the generic analytical solution; in addition, it guides the DM developer during the implementation phase. After over viewing our notion of MP and their construction method, this chapter details a reuse method composed of two adaptation levels: one logical and one physical. The logical level, which is independent of any data source model, allows a decision maker to adapt a given MP to their analytical requirements and to the RWE of their particular enterprise; this produces a DM schema. The physical specific level projects the RWE of the DM over the data source model. That is, the projection identifies the data source elements necessary to define the ETL procedures. We illustrate our approaches of construction and reuse of MP with examples in the medical domain.
INTRODUCTION

Judicious decision making within an enterprise heavily relies nowadays on the ability to analyze large data volumes generated by the enterprise daily activities. To apprehend the difficulties and often impossibility of manual analyses of huge data volumes, decision makers have manifested a growing interest in installing decision support systems (DSS) on top of their computerized information systems (IS) (Kimball R. 1996). This interest triggered the proposition of several methods dealing with various phases of the DSS life cycle. However, two main difficulties impede the wide spread adoption of so far proposed methods. One difficulty stems from the fact that some methods presume that decision makers have a good expertise in IS modeling; this is the case of bottom-up DSS design methods (Golfarelli M., Maio D. & Rizzi S. 1998a), (Golfarelli M., Lechtenbörger J., Rizzi S. & Vossen G. 1998b), (Hüsemann, B., Lechtenbörger, J. & Vossen G. 2000), (Chen Y., Dehne F., Eavis T., & Rau-Chaplin A. 2006), (Cabibbo L. & Torlone R. 2000) and (Moody L. D. & Kortink M. A. R. 2000). The second difficulty is due to the fact that other methods rely on the ability of decision makers to define their analytical needs in a rigorous way that guarantees their loadability from the data in the operational IS; this is the case of top-down DSS design methods (Kimball 2002), (Tsois A., Karayannidis N. & Sellis T. 2001).

Independently of any design method and software tool used during its development, a DSS is typically organized into a data warehouse (DW) gathering all decisional information of the enterprise. In addition, to facilitate the manipulation of a DW, this latter is reorganized into data marts (DM) each of which representing a subject-oriented extract of the DW. Furthermore, a DM uses a multidimensional model that structures information into facts (interesting observations of a business process) and dimensions (the recording and analysis axes of observations). This model enables decision makers to write ad hoc queries and to manipulate/analyze easily the results of their queries (Chrisment C., Pujolle G., Ravat F., Teste O. & Zurfluh G. 2006).

Despite the advantages of this dedicated multidimensional model, the design of the DM schema remains a difficult task. Actually, it is a complex, technical process that requires a high expertise in data warehousing yet, it conditions the success and efficiency of the obtained DM.

The originality of the work presented in this chapter resides in proposing a DM design approach that relies on the reuse of generic OLAP requirement solutions we call multidimensional patterns (MP). In fact, reuse-based development is not a novel technique in itself; it has been applied for several application domains and through various techniques, e.g., design patterns (Gamma E., Helm R., Johnson J. & Vlissides J. 1999), components (Cheesman J. & Daniels J. 2000), and more recently the OMG model driven architecture (MDA) (OMG 2003). However, the application of reuse techniques in the design of DSS has not been well explored.

More specifically, this chapter presents a pattern-based method for the construction of DM schemes. By analogy to a design pattern, which represents a generic solution to a reoccurring problem in a given application domain, we consider a multidimensional pattern as a typical standard, conceptual solution defined as a generic star-schema in one activity domain of the enterprise (Feki J. & Ben-Abdallah H. 2007). This concept of MP can be used in a top-down design approach either to prototype, or to build a DSS directly on top of the enterprise’s operational system. Such a DSS can be either light (a set of independent DMs), or complete (a DW-dependant set of DMs). In the first case, decision makers define their OLAP requirements by adapting/reusing several MPs to derive DM schemes. This MP reuse context is well suited for small enterprises that are generally unable to bear the relatively high cost of a system containing both a DW and several DMs; instead,
Related Content

Graph-Based Data Mining
[www.irma-international.org/chapter/graph-based-data-mining/10656/](www.irma-international.org/chapter/graph-based-data-mining/10656/)

Integration of Data Sources through Data Mining
[www.irma-international.org/chapter/integration-data-sources-through-data/7649/](www.irma-international.org/chapter/integration-data-sources-through-data/7649/)

Re-Sampling Based Data Mining Using Rough Set Theory
[www.irma-international.org/chapter/sampling-based-data-mining-using/7818/](www.irma-international.org/chapter/sampling-based-data-mining-using/7818/)

An Ontology of Data Modelling Languages: A Study Using a Common-Sense Realistic Ontology
[www.irma-international.org/chapter/ontology-data-modelling-languages/7828/](www.irma-international.org/chapter/ontology-data-modelling-languages/7828/)

Data Mining and Knowledge Discovery in Metabolomics Armin
[www.irma-international.org/chapter/data-mining-knowledge-discovery-metabolomics/7730/](www.irma-international.org/chapter/data-mining-knowledge-discovery-metabolomics/7730/)