

Chapter VIII

Requirement Specification and Conceptual Modeling for Data Warehouses

Elzbieta Malinowski

Universidad de Costa Rica, Costa Rica

INTRODUCTION

Data warehouses (DWs) integrate data from different source systems in order to provide historical information that supports the decision-making process. The design of a DW is a complex and costly task since the inclusion of different data items in a DW depends on both users' needs and data availability in source systems.

Currently, there is still a lack of a methodological framework that guides developers through the different stages of the DW design process. On the one hand, there are several proposals that informally describe the phases used for developing DWs based on the authors' experience in building

such systems (Inmon, 2002; Kimball, Reeves, Ross, & Thornthwaite, 1998). On the other hand, the scientific community proposes a variety of approaches for developing DWs, discussed in the next section. Nevertheless, they either include features that are meant for the specific conceptual model used by the authors, or they are very complex. This situation has occurred since the need to build DW systems that fulfill user expectations was ahead of methodological and formal approaches for DW development, just like the one we had for operational databases.

Specifically, the approaches for requirements specifications and conceptual modeling differ significantly because some of them rely mainly on user requirements, while others take into con-

sideration the underlying operational databases instead of user needs. Such diversity of approaches may overwhelm designers who could find it difficult to identify the one approach that better fits to particularities of a DW project.

In this chapter we refer to requirements specification and conceptual modeling phases for DW design. Our proposal unifies the already existing approaches by giving an overall perspective of different alternatives available to designers when developing a DW.

BACKGROUND

The requirements specification phase is one of the earliest steps in system development and it has a major impact on the success of DW projects (Winter & Strauch, 2003). This phase will help to identify the essential elements of a multidimensional schema, i.e., facts with associated measures, dimensions, and hierarchies, required to facilitate future data manipulations and calculations. These elements should be clearly and concisely represented in a conceptual schema in a later stage. This schema will serve as basis for analysis tasks performed by the users and will be used by the designers during future evolutions of the DW.

There are different approaches for requirements specification and conceptual modeling of DWs. The so-called **user-driven approach**¹ takes into account the fact that users play a fundamental role during requirement analysis and must get actively involved in the elucidation of relevant facts and dimensions (Freitas, Laender, & Campos, 2002; Luján-Mora & Trujillo, 2003). The **business-driven approach**² bases derivation of DW structures on the analysis of either business requirements or business processes (Giorgini, Rizzi, & Garzetti, 2005; List, Schiefer, & Min Tjoa, 2000). On the other hand, the **source-driven approach**³ analyzes the underlying source systems in order to obtain the DW schema (Böehnlein

& Ulbrich-vom Ende, 1999; Cabibbo & Torlone, 1998; Golfarelli & Rizzi, 1998; Moody & Kortink, 2000). Finally, the **combined approach**⁴ puts together the business- or user- driven and data-driven approaches representing what the business or user demands are and what the source systems can provide (Bonifati, Cattaneo, Ceri, Fuggetta, & Paraboschi, 2001; Winter & Strauch, 2003).

Nevertheless, the proposed approaches are difficult to use. On the one hand, different authors use techniques or models that require specific technical knowledge. On the other hand, the variety of existing approaches may overwhelm even experienced designers who tend to concentrate more on technical issues, e.g., physical modeling or query performance, and therefore, many DW projects skip the requirements specification and conceptual modeling phases.

Presenting the variety of approaches in a coherent whole may facilitate their understanding without losing their semantic differences. It could also facilitate the developer's team to choose one approach that better fits the particular needs of a DW project. Furthermore, professionals who are highly skilled in the development of operational databases but inexperienced in the development of DWs could better understand the different aspects that must be considered during the DW design process.

MAIN FOCUS

In this section we present three different approaches for requirements specifications and conceptual modeling. Depending on whether users (business) or source systems are the driving force for requirements specifications, we propose, respectively, analysis-driven and source-driven approaches. We also present what we call the analysis/source-driven approach, which combines both previously-mentioned approaches. Although we separate the requirements specifications and the conceptual modeling phases for readability purposes, in reality these phases often overlap.

7 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/requirement-specification-conceptual-modeling-data/20689

Related Content

Database Administration at the Crossroads: The Era of End-User-Oriented, Decentralized Data Processing

Mark L. Gillenson (1991). *Journal of Database Administration* (pp. 1-11).

www.irma-international.org/article/database-administration-crossroads/51094

Population of a Method for Developing the Semantic Web Using Ontologies

Adolfo Lozano-Tello and Asunción Gómez-Pérez (2005). *Advanced Topics in Database Research, Volume 4* (pp. 159-177).

www.irma-international.org/chapter/population-method-developing-semantic-web/4373

Visualization of Predictive Modeling for Big Data Using Various Approaches When There Are Rare Events at Differing Levels

Alan Olinsky, John Thomas Quinn and Phyllis A. Schumacher (2018). *Handbook of Research on Big Data Storage and Visualization Techniques* (pp. 604-631).

www.irma-international.org/chapter/visualization-of-predictive-modeling-for-big-data-using-various-approaches-when-there-are-rare-events-at-differing-levels/198779

Creativity of Participants in Crowdsourcing Communities: The Effects of Promotion Focus and Extrinsic Motivation

Lingfei Zou, Shaobo Wei, Weiling Ke and Kwok Kee Wei (2020). *Journal of Database Management* (pp. 40-66).

www.irma-international.org/article/creativity-of-participants-in-crowdsourcing-communities/256847

Enabling Resource Access Visibility for Automated Enterprise Services

Kaushik Dutta and Debra VanderMeer (2014). *Journal of Database Management* (pp. 1-28).

www.irma-international.org/article/enabling-resource-access-visibility-for-automated-enterprise-services/117742