Computer Standards & Interfaces 30 (2008) 341–350

Contents lists available at ScienceDirect

Computer Standards & Interfaces

journal homepage: www.elsevier.com/locate/csi

Building a secure star schema in data warehouses by an extension of the relational package from CWM

Emilio Soler [a,*], Juan Trujillo [b], Eduardo Fernández-Medina [c], Mario Piattini [c]

[a] Department of Computer Science, University of Matanzas, Autopista de Varadero km 3, Matanzas, Cuba
[b] Department of Software and Computing Systems, University of Alicante, C/San Vicente S/N 03690 Alicante, Spain
[c] ALARCOS Research Group, Information Systems and Technologies Department, UCLM-Soluziona Research and Development Institute, University of Castilla-La Mancha, Paseo de la Universidad 4, 13071 Ciudad Real, Spain

ARTICLE INFO

Available online 10 March 2008

Keywords:
Security
Star scheme
CWM
Data Warehouses

ABSTRACT

Data Warehouses (DWs) are widely accepted as the core of current decision support systems. Therefore, it is vital to incorporate security requirements from the early stages of the DWs projects and enforce them in the further design phases. Very few approaches specify security and audit measures in the conceptual modeling of DWs. Furthermore, these security measures are specified in the final implementation on top of commercial systems as there is not a standard relational representation of security measures for DWs (i.e. the well-known star schema does not allow us to specify security and audit measures on its multidimensional representation of data; instead, they must be specified on top of the implemented relational tables). On the other hand, the Common Warehouse Metamodel (CWM) has been accepted as the standard for the exchange and the interoperability of metadata. Nevertheless, it does not allow us to specify security measures for DWs. In this paper, we make use of the own extension mechanisms provided by the CWM to extend the relational package in order to build a star schema that represents the security and audit rules captured during the conceptual modeling phase of DWs. Finally, in order to show the benefits of our extension, we apply it to a case study related to the management of the pharmacy consortium business.

1. Introduction

Data Warehouses (DWs) play a central role in current decision support systems because they provide crucial business information through which to improve strategic decision-making processes [8]. Organizations have begun to adopt more and more computerized information systems, which rely upon databases and DWs that require more security, because the very survival of the organization depends on the appropriate manipulation, security and confidentiality of information [4]. On the other hand, it is widely accepted that the design of DWs is based on multidimensional (MD) modeling which structures the information into facts and dimensions.

As for traditional databases, we follow the three design phases proposed by ANSI/SPARC [2] in the design of data warehouses from user requirements to the final implementation, i.e. the business (requirement analysis), conceptual, logical and physical levels.

Following this procedure, we align the design of DWs with the Model Driven Architecture (MDA) [21]. MDA allows us to achieve the portability and interoperability of software systems by means of transformations between models. To help us with these transformations, the Query/View/Transformation (QVT) Language [22] is proposed in order to allow us to accomplish automatic transformations between models. MDA proposes several models at different levels: Computation Independent Model (CIM), Platform Independent Model (PIM), Platform Specific Model (PSM) and Code. In our context, the CIM corresponds with the work [15], which is based on an extension of the i* framework considered [36]. The proposal will be extended to consider security at the business level (see 7: Future work section). The PIM corresponds with an extension of the Unified Modeling Language (UML) profile [11] presented in [34] which allows us to consider the main properties of secure MD modeling at this stage. The PSM corresponds with our extension of the Common Warehouse Metamodel (CWM) at the logical level and the Code with implementation at the physical level, i.e. with a Database Management System (DBMS). Fig. 1 shows the relationship between MDA and the DWs lifecycle in detail. In this paper, we focus on the logical level. It is out of the scope of this paper to consider the business and physical levels. See, the 7: Future work section.

As we will describe in the related work section, security has hardly been contemplated in literature. Normally in DWs projects, security aspects are implemented in the final stages of the design. However, information security is a serious requirement which must be given careful thought, not as an isolated aspect, but as an element which is

* Corresponding author.
E-mail addresses: emilio.soler@umcc.cu (E. Soler), jtrujillo@dlsi.ua.es (J. Trujillo), eduardo.fdezmedina@uclm.es (E. Fernández-Medina), mario.piattini@uclm.es (M. Piattini).

0920-5489/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csi.2008.03.002

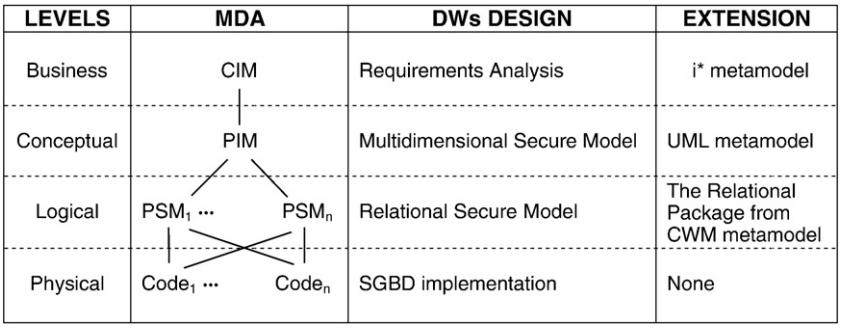

Fig. 1. Aligning the design of DWs with MDA.

in all development lifecycle stages, from requirement analysis to implementation and maintenance [3].

The works [6,7,34] extend the proposal presented in [11] to incorporate security aspects to the DWs design at the conceptual level. In order to align our architecture with MDA we need to build a secure PSM that allows us to represent security and auditing aspects. On the other hand, MDA does not provide security specific modeling facilities and its general modeling facilities fail to satisfy one or more of the security protocol modeling aspects [16].

The previous work presented in [14] employs MDA for DWs development, choosing the relational metamodel from CWM [19]. The relational package of CWM enables mediated interchange between relational databases from the majority of relational commercial systems [26]. CWM also offers the On-Line Analytical Processing (OLAP) package for the essential OLAP concepts common to most OLAP systems. However, the OLAP package is a general metamodel with which to exchange metadata information, and therefore, it only considers general aspects of multidimensional modeling. Thus, we do not consider it to be appropriate for the conceptual modeling of complex and real case DWs. For the conceptual modeling phase, we have based our proposal on UML [11], which offers a greater level of expressiveness of MD modeling at the conceptual level. The CWM offers facilities through which to access and exchange data warehouse metadata. However, security and audit measures cannot be modeled in the CWM because it does not provide the modeling constructors through which to represent data security related to issues such as access rights, users or roles [18]. Most data access control approaches are based on the proprietary metadata structures of specific software products [25]. Thus, integrating security related to metadata into the CWM improves the security support and facilitates the establishment of a standardized access control mechanism for data warehouses [18]. According to the MDA we do not need the metadata of a DBMS; we need a metamodel that allows us to represent security and audit measures at the logical level, i.e., a PSM which in our case corresponds with a relational platform. In order to achieve this goal we have extended the relational package from the CWM.

Hence, in this paper we present an extension of the relational metamodel from CWM by using its own extension mechanisms. By using QVT, we automatically transform all the security and audit measures captured from the conceptual modeling phase of the DWs design at the logical level. Our main contribution is that the proposed extension allows us to represent all the requirements of security and audit captured during the conceptual modeling phase at the logical level.

The remainder of this paper is structured as follows. The works related to our proposal are discussed in Section 2. Secure MD modeling is introduced in Section 3. Section 4 shows an overview of the CWM. Section 5 presents our extension of the relational metamodel from CWM. Then, in Section 6, we develop a case study in order to build a secure star scheme using our extension in the design of secure DWs. Finally, Section 7 draws the main conclusions and outlines our immediate future work.

2. Related work

Relevant literature on this subject comprises several initiatives to include security in the DW design. In [9] the authors describe a prototype model for DWs security based on metadata, which enables the definition of views of data for each group of users. However, this does not permit the specification of complex restrictions of confidentiality. Rosenthal and Sciore [27] extend SQL grants and create a mechanism of inferences through which to establish security. Another attempt is the architecture for both Federated Information Systems (FIS) and DWs that preserve MultiLevel security integration between FIS and DWs [28]. These approaches [9,27,28] are attractive but only focus on practical issues such as acquisition, storage and access control on the OLAP side. None of them examine the representation of security at both a conceptual and a logical stage.

On the other hand, there are more elaborate initiatives that propose authorization models for the design of DWs. For instance, in [10], the authors propose a security concept for OLAP, which is a role based security model for data warehouses. Priebe and Pernul [25] propose a security design methodology similar to that of the classical database design (requirement analysis, conceptual, logical, and physical design) which covers requirements and concrete implementations in commercial systems. The same authors extend the ADAPTed UML model for the previous conceptual phase [24], specifying a methodology and a MD security constraint language for the conceptual modeling of OLAP security. In [5], the authors show that access privileges for DWs and OLAP can be expressed more intuitively than by using SQLs grant statements. Their access control model focuses specifically on expressiveness and usability. These proposals [10,25,24] offer security models at the conceptual level by means of security constraints, but basically deal with OLAP operations. These proposals [25,24] are one of the best references in this area. As a summary, these works implement the security rules considered in their conceptual approach to commercial database systems. On the other hand, we base our approach on the works of [6,7,34] in which the authors call for the design of the security rules at all stages of the DWs design, from requirement analysis to final implementation. Therefore, in this paper, we formally extend the CWM in order to allow us to automatically transform the security rules considered at the conceptual level in the logical representation of the DW.

Some proposals related to the access and preservation of the confidential information in the analysis at the data cube level have recently appeared. In [35] the authors proposed a method for protecting sensitive data in OLAP cubes from unauthorized access and malicious inferences of sensitive values. Other proposals for preserving data privacy and range queries in data cubes in DWs are the zero-sum and the cubic-wise balance methods (respectively)