IDEA GROUP PUBLISHING **ITB12241**

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

This paper appears in the publication, Enterprise Information Systems Assurance and Systems Security
edited by Merril Warkentin © 2006, Idea Group Inc.

Chapter XIX

Designing Secure Data Warehouses

Rodolfo Villarroel, Universidad Católica del Maule, Chile

Eduardo Fernández-Medina, Universidad de Castilla-La Mancha, Spain

Juan Trujillo, Universidad de Alicante, Spain

Mario Piattini, Universidad de Castilla-La Mancha, Spain

Abstract

Organizations depend increasingly on information systems, which rely upon databases and data warehouses *(DWs), which need increasingly more quality and security. Generally, we have to deal with sensitive information such as the diagnosis made on a patient or even personal beliefs or other sensitive data. Therefore, a final DW solution should consider the final users that can have access to certain specific information. Unfortunately, methodologies that incorporate security are based on an operational environment and not on an analytical one. Therefore, they do not include security into the multidimensional approaches to work with DWs. In this chapter, we present a comparison of six secure-systems design methodologies. Next, an extension of the UML that allows us to specify main security aspects in the multidimensional conceptual modeling is proposed, thereby allowing us to design secure DWs. Finally, we present how the conceptual model can be implemented with* Oracle Label Security *(OLS10g).*

Introduction

The goal of information confidentiality is to ensure that users can only access information that they are allowed. In the case of *multidimensional* (MD) models, confidentiality is crucial, because very sensitive business information can be discovered by executing a simple query. Sometimes, MD databases and *data warehouses* (DWs) also store information regarding private or personal aspects of individuals; in such cases, confidentiality is redefined as privacy. Ensuring appropriate information privacy is a pressing concern for many businesses today, given privacy legislation such as the United States' HIPAA that regulates the privacy of personal health care information, Gramm-Leach-Bliley Act, Sarbanes-Oxley Act, and the *European Union*'s (EU) Safe Harbour Law.

Generally, information systems security is taken into consideration once the system has been built, is in operation, and security problems have already arisen. This kind of approach — called "penetrate and patch" — is being replaced by methodologies that introduce security in the systems development process. This is an important advance but, unfortunately, methodologies that incorporate security are based on an operational environment and not on an analytical one. If we tried to use the operational environment to process consistent, integrated, well-defined and time-dependent information for purposes of analysis and decision making, we would notice that data available from operational systems do not fulfil these requirements. To solve this problem, we must work in an analytical environment strongly supported by the use of multidimensional models to design a DW (Inmon, 2002).

Several papers deal with the importance of security in the software development process. Ghosh, Howell, and Whittaker (2002) state that security must influence all aspects of design, implementation and software tests. Hall and Chapman (2002) put forward ideas about how to build correct systems that fulfil not only normal requirements but also those pertaining to security. These ideas are based on the use of several formal techniques of requirement representation and a strong correction analysis of each stage. Nevertheless, security in databases and data warehouses is usually focused on secure data storage and not on their design. Thus, a methodology of data warehouse design based on the *Unified Modeling Language* (UML), with the addition of security aspects, would allow us to design DWs with the syntax and power of UML and with new security characteristics ready to be used whenever the application has security requirements that demand them. We present an extension of the UML (*profile*) that allows us to represent the main security information of the data and their constraints in the MD modeling at the conceptual level. The proposed extension is based on the *profile* presented by Luján-Mora, Trujillo, and Song (2002) for the conceptual MD modeling, because it allows us to consider main MD-modeling properties and it is based on the UML. We consider the multilevel security model but focus on considering aspects regarding *read* operations, because this is the most common operation for final user applications. This model allows us to classify both information and users into security classes and enforce mandatory access control. This approach makes it possible to implement the secure MD models with any of the *database management systems* (DBMS) that are able to implement multilevel databases, such as Oracle Label Security (Levinger, 2003) and DB2 Universal Database, UDB (Cota, 2004).