Data Warehouse Security through Conceptual Models

Nitin Anand[*]**, Poornima Sharma**[**]

[*]CSE Deptt, Ambedkar Institute of Advanced Communication Technologies and Research (AIACTR)
[**]CSE Deptt, Shri Venkateswara College

Abstract- A key challenge for data warehouse security is how to manage the entire system coherently – from sources and their export tables, to warehouse stored tables (conventional and cubes) and views defined over the warehouse. Permissions on the warehouse must satisfy the restrictions of the data owners, and be updated quickly as those local concerns evolve. Yet the system cannot demand extensive administrator time, since there are too few people with both technical skills to understand derivation logic, and business skills to balance security versus accessibility. Security aspects should be considered in the design phase of the data warehouse to better match the security requirements and to avoid later fundamental,cost-intensive adaptations. For the identification of security requirements legal, audit, network and other issues have to be considered

Index Terms- Conceptual models, Data warehouse, Metadata,Multidimensional data model, OLAP

I. Introduction

Traditionally, Data warehouse were queried by high level users (executive management , business analyst only). Data warehousing systems enable enterprise managers to acquire and integrate information from heterogeneous sources and to query very large databases efficiently. Data warehousing systems enable enterprise managers to acquire and integrate information from heterogeneous sources and to query very large databases efficiently. Building a data warehouse requires adopting design and implementation techniques completely different from those underlying operational information systems.

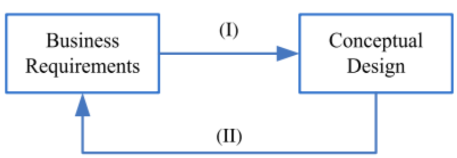

Fig 1 . Early phases of DW design [19]

The complexity of data warehouse environments is rising every day and data volumes are growing at a significant pace. The data warehouse administration and design group should manage enterprise information efficiently, and with high quality results. Approximately 30–50% of the data warehouse design process is spent on analysis, to ensure that source systems are understood and that there is alignment between the involved parties: business and administrators. Usually, such design process involves the representation of the business requirements into a conceptual design (Fig. 1-I) and it is continuously refined through a feedback process (Fig. 1-II). An integrated metadata strategy may reduce the time needed, lower risk, and produce an ongoing record of understanding, useful in the whole data warehouse lifecycle. Hence, it is of great importance to provide a formal, explicit, and well-defined way to represent all the parameters and properties guiding this early stage. Moreover, the proposed solution should be simple and comprehensive enough to be used by all the parties involved, which usually have different technical skills, knowledge background, and communication codes. Data warehouse often store historical information which is extracted from multiple, heterogeneous, autonomous and distributed data sources, thereby; the survival of the organizations depends on the correct management, security and confidentiality of the information. The application of the Model Driven Architecture (MDA) [1] in the secure modeling of DWs allows obtaining the secure logical scheme from the conceptual model.Confidentiality is defined as the absence of unauthorized disclosure of information, integrity as the absence of improper system alterations and availability as readiness for correct service [5] and Dependability is a broader concept that encompasses all primary aspects of security save confidentiality.

MULTIDIMENSIONAL DATA MODELLING

Multidimensional data modeling is an integrated aspect of OLAP . . It involves the analysis of selected facts or measures of the business area. Multidimensional modeling is a prominent factor in interactive analysis of large amount of data for decision making purpose. Basically multidimensional modeling is the foundation of the data warehouses[20].The multidimensional data model is composed of logical cubes, measures, dimensions, hierarchies, levels and attributes. The simplicity of the model is inherent because it defines objects that represent real-world business entities. Analysts know which business measures they are interested in examining, which dimensions and attributes make the data meaningful, and how the dimensions of their business are organized into levels and hierarchies. Multidimensional data cubes, are the basic logical model for OLAP application is shown in fig 2.

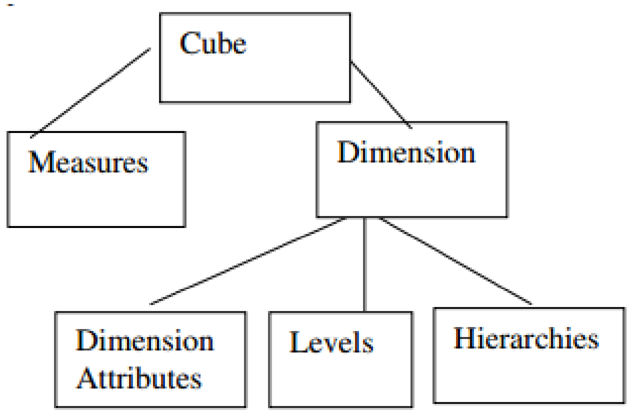

Fig 2 Multidimensional model of Data Warehouse

To define the data warehousing security and audit requirements. [11]

DEFINE ACCESS REQUIREMENTS

We should classify the data and users in the data warehouse. The data can be classified by its sensitivity (e.g., restricted, confidential) or by user role. Users can be classified by organizational unit, by user role, or by individual. Identify restrictions by data and user classification. Define time frames when secure data cannot be published to non-secure staff.

DEFINE AUDIT REQUIREMENTS

Define audit requirements. Consider gathering information on connections, disconnections, data accesses, and data changes.

DEFINE NETWORK REQUIREMENTS

Identify network requirements, such as data encryption/decryption and routing restrictions.

DEFINE DATA REQUIREMENTS

Identify any legal restrictions that apply to data kept on customers, employees, etc. To meet legal restrictions it may be necessary to store summarized data so that individual entities (e.g., employees, companies) cannot be identified.

Identify data privacy laws that must be adhered to (e.g., most countries require that companies that hold data on customers must make this data available to the customer on demand).

While defining data movement security requirements. The following points should be considered:

· Can data be transferred to a diskette?
· Can data be stored on a PC?
· Who might have access to data transferred to a diskette or a PC?
· Do back-ups have to be encrypted?
· Where should back-ups be kept?
· Who might have access to the back-ups?

DEFINE HIGH SECURITY REQUIREMENTS

If high levels of security are required, define the requirements, such as "trusted" versions of database management systems, "trusted" versions of operating systems, and secure facilities.

II. BACKGROUND

Database security was addressed in the 1960s by introducing mandatory access control (MAC), driven mainly by military requirements. Today, role-based access control (RBAC) [6] is the commonly used access control model in commercial databases. RBAC does not define the degree of scalability implemented in a system with respect to the number of roles, number of permissions, size of role hierarchy, or limits on user-role assignments, etc.The security model in the data warehouse restricts the viewership of users to data for only those projects that they have access to. By default, the security is disabled, giving all users the ability to see all data in all projects. [11]

When users have permission to a project, they have permission to access all the data within the project, both in the data mart and the operational data store. Data that does not belong to a specific project can be seen by all users. The relationship of users to projects is defined in a lookup table in the data warehouse. The switching of security is handled through a parameter in the Framework Manager data warehouse model. Security is disabled by default and all users can see all data regardless of the user-to-project mapping in the lookup table, if loaded. If security needs to be switched on, you must load the RESOURCE and PROJECT_RESOURCE_LOOKUP tables to the data warehouse and republish the reporting package. All the tables that have project IDs in data warehouse have a security filter that is controlled by the VW_PROJECT_RESOURCE_LOOKUP table. The switching of security is handled through a parameter in the Framework Manager data warehouse model, and data permission is achieved through query subject filters.

III. RELATED WORK

This section divides the related work according to two main research topics covered by this paper: ETL modeling, and data warehouse modeling. The paper focuses on the logical modeling specifications in these topics. Many data structures in the data warehouse are completely devoid of sensitive individual identities by design and, therefore, do not require protection appropriate for the most private and sensitive data. For example, when data has been aggregated into summaries by brand or region, as is often the case with data warehousing, the data no longer presents the risk of compromising the private identities of individuals. However, the data can still have value as competitive intelligence of market trends, and thus requires careful handling to keep it out of the hands of rival firms. Relaxed security does not mean a lack of commitment to security. The point is that differing levels of security requirements ought to remind us that one-size-fits-all solutions are likely to create trouble [12]. Moreover, a number of security models have been proposed for data warehouse and OLAP. For example, [14] propose a security model based on mandatory access control for OLAP cubes. Another model for data warehouse security based on metadata is presented in [15]. The authors elaborate on requirements of and impacts on the selection of an adequate security model for a data warehouse environment.

Building upon on a similar mindset [16] focuses on ideas that can contribute to warehouse security. The concepts take common operations in to account such as replication control, aggregation and generalization, ex-aggregation and misleading, anonymity, and user profile based security.

[17] suggest to base data warehouse security on view security. This paper aims to provide a theory that permits automated inference of many permissions for the warehouse, in a way that minimize both the learning curve for administrators and the amount of new software that vendors would need to implement. Moving from theory to practical challenges [18] focus mainly on the technical issues such as authorization and access control.

IV. DATA WAREHOUSE AND SECURITY

It empowers end-users to perform data access and analysis. It also gives an organisation certain competitive advantages, such as fostering a culture of information sharing, enabling employees to effectively and efficiently solve dynamic organisational problems, minimizing operating costs and maximising revenue, attracting and maintaining market shares, and minimizing the impact of employee turnovers. The security requirements of the data warehouse environment are similar to those of other distributed computing systems [2]

4.1 SECURITY RESTRICTIONS

A data warehouse is an open, accessible system. The aim of a data warehouse generally is to make large amounts of data easily accessible to users, thereby enabling them to extract information about the business as a whole. Any security restrictions can be seen as obstacles to that goal, and they become constraints on the design of the warehouse. The main problem with data collection is that people might allow companies to use data for specific reasons (such as recommending related products) but do not consent to other uses of the same data. Usage control [4] is a concept that makes it possible to enforce pre- and postconditions when using data. It is similar to a traditional reference monitor, only that the restrictions are enforced during the entire access [7]

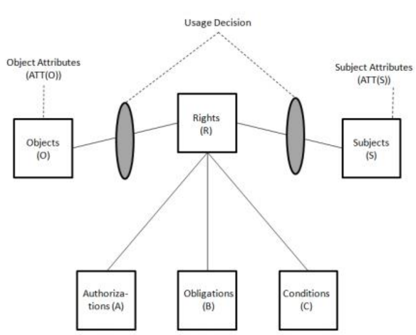

Fig 3 The UCONABC usage control model [4]

4.2 SECURITY REQUIREMENTS

Security requirements describe all security conditions that have to be considered in the data warehouse environment. It is important to determine in an early stage any security requirements that will be enforced in the data warehouse, because they can seriously impair the organisation and design of the warehouse. It is very much important to freeze the security at the beginning of requirement phase. This helps in system design.

4.3 METADATA AND SECURITY MODEL

Data about data is called metadata, Metadata not only describe the contents of the data warehouse but also provide the user with information useful in judging the quality of the content [3]. The role of metadata is rapidly expanding as organisations develop a data warehouse strategy that may result in the creation of operational data stores, integrated data warehouses and multiple data marts. Metadata, which are used by developers in order to manage and control the creation and maintenance of the data warehouse, are kept outside of the data warehouse. The metadata concerning data warehouse users are on the contrary a part of data warehouse. This data are used to control access to and analysis of data. Figure shows a simplified diagram of how a metadata points to all information in a data warehouse.

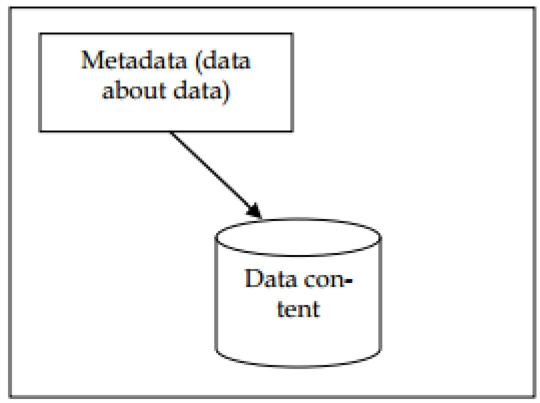

Figure 4: Metadata in the data warehouse [21]

Consequently there are two types of metadata [8]:
• Structural metadata
• Access metadata

4.2.1. STRUCTURAL METADATA

Structural metadata are used for the creation and maintenance of the data warehouse. They completely describe the data warehouse - its structure and its content. The basic element of structural metadata is a model which describes their data subjects, their features and their intermediate relationships. [9]

4.2.2. ACCESS METADATA

Access metadata represent a dynamic relationship between the data warehouse and the end-user applications. They contain the measured values of the enterprise, user-defined names and aliases. These data hold the description of the data warehouse server, databases, tables, detailed data and summed up data with a description of the initial data sources and of the transformations being carried out. Access metadata set up "drill-down" and "roll-up" rules as well as the views over dimensions and available hierarchies (such as products, markets and customers). This data can also contain the rules for user-defined calculations and queries. Such metadata allow the security implementation for an individual user, user groups or the whole enterprise, relating reading, modifying etc. of calculations, summed up data or analyses. [9]

V. CONCLUSION AND FUTUREWORK

In this paper we focus on a basic authorization model for data warehouses and OLAP offering greater expressiveness and highly increased usability with respect to security. Future work will also have to deal with special security requirements [13] with respect to aggregation functions different to SUM, like variance or average, for instance. Likewise the issue of derived authorizations has to be addressed in the context of data warehouses especially while serving a drill-through operation that directly forwards access to the OLTP data sources.

REFERENCES

[1] Object Management Group: MDA Guide 1.0.1, http://www.omg.org/cgi-bin/doc?omg/03-06-01,2003rchitecture (MDA)

[2] Slemo Warigon, Data Warehouse Control and Security, Association of College and University Auditors LEDGER, Vol. 41, No. 2, April 1997; pp. 3

[3] Alex Berson, Stephen j. Smith, Data Warehousing, Data Mining & OLAP, McGraw-Hill Series on Data Warehousing and Data Management, 1997

[4] Park, J., & Sandhu, R. (2004). The uconabc usage control model. ACM Transactions on Information Security, 7(1), 128–174.

[5] Algirdas Avizienis, J.-C., Laprie, Randell, B., & Landwehr, C. (2004). Basic concepts and taxonomy of dependable and secure computing. IEEE Transactions of Dependable and Secure Computing, 1(1), 11–33.

[6] Sandhu, R.S., Ferraiolo, D., & Kuhn, R. (2000, July). The NIST model for role based access control: Towards a unified standard. In Proc. of 5th ACM Workshop on Role-Based Access Control, Berlin, Germany: ACM, ACM Press.

[7] Thuraisingham, B. (2005). Privacy constraint processing in a privacy-enhanced database management system. Data & Knowledge Engineering, 55, 159–188.

[8] A. Perkins, Developing a Data Warehouse, the Enterprise Engineering Approach, Visible Systems Corporation, 1995-96

[9] N.Katic, Ein Metadaten-basiertes Sicherheitsmodell für OLAP Datenbanken, MSC. Thesis, Vienna University of Technology, 1997.

[10] http://it.toolbox.com/blogs/enterprise-solutions/defining-data-security-requirements-36031

[11] http://pic.dhe.ibm.com/infocenter/rentrpt/v1r0m1/index.jsp?topic=%2Fcom.ibm.rational.raer.help.doc%2Ftopics%2Fc_dw_security.html

[12] http://www.information-management.com/issues/20031001/7398-1.html

[13] S. Barker and A. Rosenthal. Flexible Security Policies in SQL. Proc. 15th Annual IFIP WG 11.3 Working Conf. on Database and Application Security, Niagara on the Lake, Ontario, Canada, July 15-18, 2001.

[14] R. Kirkgöze, N. Katic, M. Stolba, and A M. Tjoa. A Security Concept for OLAP. Proceedings of the 8th. International Workshop on Database and Expert Systems Applications (DEXA'97), IEEE Computer Society, 1997.

[15] N. Katic, G. Quirchmayr, J. Schiefer, M. Stolba, and A M. Tjoa. A Prototype Model for Data Warehouse Security based on Metadata. Proc. 9th Int. Conf. on Database and Expert Systems Applications (DEXA'98), Aug. 26-28, 1998, Vienna, Austria. IEEE Computer Society, Vol. 8, pp. 300-308, 1998

[16] B. Bhargava. Security in Data Warehousing (Invited Talk). Proceedings of the 3rd Data Warehousing and Knowledge Discovery (DAWAK'00), 2000.

[17] A. Rosenthal, and E. Sciore. View Security as the Basis for Data Warehouse Security. Proceedings of the International Workshop on Design and Management of Data Warehouse (DMDW'2000), Manfred A. Jeusfeld, Hua Shu, Karlstad Martin Staudt, Gottfried Vossen (Editor), Sweden, June 2000.

[18] T. Priebe, and G. Pernul. Towards OLAP Security Design – Survey and Research Issues. Proc. of the 3rd ACM International Workshop on Data Warehousing and OLAP (DOLAP 2000), Washington, DC, November 10, 2000.

[19] Alkis Simitsis, Dimitrios Skoutas, Malú Castellanos, Representation of conceptual ETL designs in natural language using Semantic Web technologyData & Knowledge Engineering 69 (2010) 96–115

[20] Umashanker Sharma, Anjana Gosain, " Dimensional modeling for Data Warehouse", In: Proc. ISCET, Database. 2011

[21] Dr. S.L. Gupta, Sonali Mathur, Palak Modi, "Data Warehouse Vulnerability and Security" , International Journal of Scientific & Engineering Research Volume 3, Issue 5, May-2012

AUTHORS

First Author – Nitin Anand, CSE Deptt, Ambedkar Institute of Advanced Communication Technologies and Research (AIACTR), Email:proudtobeanindiannitin@gmail.com
Second Author – Poornima Sharma ,CSE Deptt, Shri Venkateswara College, Email: poornimamanchanda16@gmail.com