Comparative Study of Data warehouse Design Approaches from Security perspective

1V.David Sushil, 2B.Siva NagaRaju, 3K.RaghavaRao
1,2 M.Tech(CSE)Students,
3 Professor, Dept of CSE,
KL University,
Guntur, Andhra Pradesh, India

Abstract:

Data warehouse systems integrate data from heterogeneous sources and are used by decision makers to analyze the status and the development of an organization. Due to the sensitive data contain in the DW it is most important to specify a security, from early stages of DW design so that data must be secured. The process of developing a data warehouse starts with identifying and gathering requirements, designing the dimensional model followed by testing and maintenance. The design phase is the most important activity in the successful building of a data warehouse. In this paper I describe the literature related to work done by various authors in last few years and Comparison of various approaches in data warehouse design from security perspective by using various parameters.

1. INTRODUCTION:

Fundamental concept of a data warehouse is the distinction between data and information. Data is composed of observable and recordable facts that are often found in operational or transactional systems. At Rutgers, these systems include the registrar’s data on students (widely known as the SRDB), human resource and payroll databases, course scheduling data, and data on financial aid. In a data warehouse environment, data only comes to have value to end-users when it is organized and presented as information. Information is an integrated collection of facts and is used as the basis for decisionmaking. For example, an academic unit needs to have diachronic information about its extent of instructional output of its different faculty members to gauge if it is becoming more or less reliant on part-time faculty.
Data Warehouse (DW) systems are used by decision makers to analyze the status and the development of an organization, based on a large amount of data integrated from heterogeneous sources into multidimensional model (MD). Data Warehouses (DWs) manage business’ historical information used to take strategic decisions and usually follow a multidimensional approach in which the information is organized in facts classified per subject called dimensions. In a typical DW architecture, ETL (extraction/transformation/load) processes extract data from heterogeneous data Sources and then transform and load this information into DW repository. Finally, this information is analyzed by Data base management systems (DBMS) and On line Analytical Processing (OLAP) tools. Since data in DWs are crucial for enterprise, it is very important to avoid unauthorized access to information by considering in all layers and operations of the DW, from the early stages of development as a strong requirement to the final implementation in DBMS or OLAP tools (Thuraisingham, Kantarcioglu et al. 2007). Several approaches for DWs modeling through a specific structural characteristic (Facts, Dimensions, Bases, Hierarchies, etc) exist, but only some of them include security aspects in their modeling [1]. However, these contributions deal with security problem in a static manner in which a set of security constraints basically establish what information will be shown to or hidden from the user, depending on his/her security profile.

FIG 1: Data warehouse Architecture

2. BACKGROUND WORK OF DW DESIGN APPROACHES:

In the literature, different data models both conceptual and logical have been proposed for data warehouse design. These approaches are based on their own visual modeling languages or make use of well known graphical notation like ER model or UML, but to the best of our knowledge, there is no standard method or model that allows us to model all aspects of a
Moreover, during our survey we noticed that most of the research efforts in designing and modelling DWs have been focused on the development of MD data models and conceptual design, the interest on the physical design of DWs has been very poor. The pioneer author in the field of data warehouse design is Juan Trujillo. He has made a major contribution. He proposed the use of UML for the design of data warehouse. He defined four UML profiles for modelling different aspects of data warehouse: the UML profile for Multidimensional Modelling, the Data Mapping profile, the ETL profile and Database Deployment profile. In authors propose an approach that provides a theoretical foundation for the use of OO databases and Object relational databases in DW. This approach introduces a set of minimal constraints and extensions to UML for representing multidimensional modeling properties for DW. In authors have proposed a multidimensional profile for the Data warehouse conceptual schema and Client conceptual schema. The author has also shown work in the field of physical schema.

### Fig 2: Development Process of Data Warehouse

#### 3 MODELING SECURITY IN DW:

Information security is a basic requirement for a wide range of applications. In the case of DWs, among the different aspects of security, confidentiality (i.e., ensuring that users can only access the information they have privileges for) is particularly relevant, because business information is a very sensitive and can be discovered by executing a simple query. One of the main concerns in DW design is data security, which is usually seen as a non-functional requirement. The security modeling for DWs comprises several initiatives to include security in the DW design. In the authors describe a prototype model for DW security based on metadata, whose main goal is to reduce user queries to only those data which are to be seen by that user. However, this does not permit the specification of complex restrictions of confidentiality such as
deny allow access to a special user combining groups and security constraints. Rosenthal and Score extend SQL grants and create a mechanism of inferences through which to establish DW security, which derives permission on tables and views of the system, thus establishing easy administration. A further attempt is that of the architecture for both Federated Information Systems (FIS) and DWs which preserves Multi Level security integration between FIS and DWs .The authorization of the DW scheme built takes into account the security policy of the federation itself. Kirkgöze et al. defines a model based on the Discretionary Access Model (DAC) which propose a security concept for OLAP, a role based security model for DWs. According to these security rules, a derived data cube is defined for each role. Essmayr et al. shows how access privileges for DWs and OLAP can be expressed more intuitively than by using SQL’s grant statements. This access control model focuses specifically on expressiveness and usability. These approaches 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 the early stages of the DW design. More elaborate initiatives which propose authorization models for DWs design also exist. For example, “Priebe and Pernul” propose a security design methodology similar to the classical database methodology (requirement analysis, conceptual, logical, and physical design) which covers requirements and concrete implementations in commercial systems.

**CONCLUSION:**

In this paper we have seen that several approach have been proposed by a variety of authors to secure a data warehouse but most of the approach basically start from conceptual level to final implementation of DW. However the approaches that starts from the conceptual level gives the complete information in data warehouse design from security perspectives. But none of the approaches will be start to secure a data warehouse design from requirement level to final implementation.
REFERENCES


