Volume 3, Issue 11, November 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

# Review on Security Levels of Data Warehouse

**Ephraim Mbaka Timothy**
*Lovely Professional University*
*India*

**Kavisha Duggal**
*Assistant Professor*
*Lovely Professional University, India*

***Abstract*— Database integrity has a major goal of maintaining the internal consistency of the data in the database. The database system is to maintain internal consistency volume, variety and velocity of data is observed by many users as the principal database security concern. These can be considered to be the minimal requirement for the database system to be used. Data are generated every minute from different source. These data need to be organized and secured for accurate and easy knowledge discovery. So this paper focuses on various security challenges of data in 3 – tier architecture in respect to data- warehouse. The paper throw a light on various factors of security issues in 3- tier architecture while fetching data from various sources, issues related storing these data and retrieving these data. The security problem not only exists in just OLAP that is while storing the data in data warehouse but its concern starts from very beginning when the fetching of data takes place .***

***Keywords*— Data mart, decision support system, Authentication***

## I. INTRODUCTION

Data today, has become one vital products in a business world, due to analysis for productive and growth. Both small and large organization has attached a great deal of value to it. The data of today is the most valuable factor. Loss of data is more concern than the loss of money. In fact today, to income money the one thing is needed is data. Data warehouse is the biggest example of it. Organizations realize that, for data to have any relevant value, it really needs to be discoverable, useful and accessible. To analyse the data is the biggest challenge for today. Therefore it is said that the actual treatment of medicines could be done not by doctors. It is done by researchers. In analysing the data, it needs to be accurate, complete and timely, if any organization need to use it to support analysis from a data warehouse. As a result of these accurate data needed to be secure a lot. Data warehouse is a like a store for storing heterogeneous data. And data mart is the subset of data warehouse where we fetch some of the data from data warehouse whereas, data mining means to extract the meaningful or the data which user need for analysis. The data is extracted from one or more production databases to produce decision support. These data provide decision support system (DSS), it assist the organization management in taking decision for the effective running of the organization.

## II. WHY IS SECURITY IMPORTANT IN DATA WARE-HOUSE?

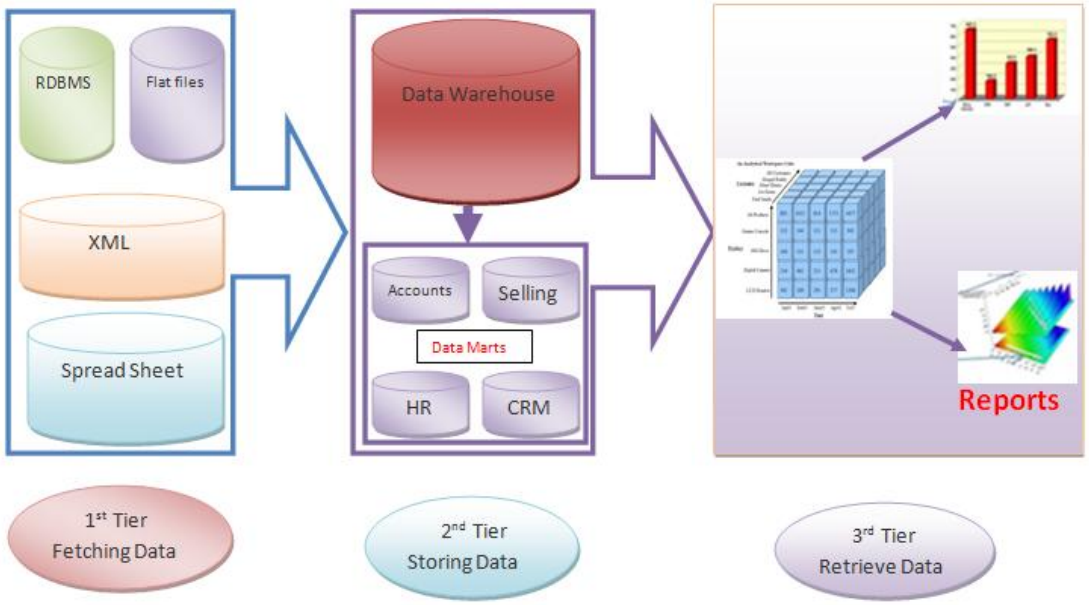
In every system, accuracy of a data is very important for any decision to be reached, for example; if a company wants to take decision on the marketing of their product to a specific area, they can only deliberate and produce a good marketing strategy if they have the right data. For knowledge discovery of the data, its privacy and security must be assured.

## III. FETCHING THE DATA

Operational systems are system in which we store our data but the major difference is in operational System is used to store small database while the data warehouse is used to store a huge amount of data. Operational database is an example of computer hard disk or external hard disk where the data can be stored. But the problem arises when data became huge and it is being stored in hundreds of hard disk then there is a need of data warehouses. Application packages like MS Access, Excel and Peachtree etc. These applications have a common function which is accessing files directly. It means that, the file in which work is going on must be access on a shared network or remotely on a local drive. It is the simplest and easiest of the tier and least secure, since most users have direct access to the data files, they could accidentally or intentionally modify or change its location or worse could be they destroy the data. Data warehouse as it is know is a collection of different types of data from different sources, the data collected here are very important, due to the fact that it's the same data that are transfer into the data warehouse. Internal users are granted excessive privilege and over reliance on network and parameter are major causes of the fail in security here. Organization, generally depend on the traditional style of password security not minding the perimeter security and network base security, these only cannot mitigate attacks from hackers.

Today's threat is often very persistent, organization need to implement more effective security measures on their systems, which will be able to sort out data which are incorrect, and the level of authentication should be rationalized. Noisy data is also one of the main security threat towards the analysis if the wrong input is created it will automatically generate a wrong output. To overcome this problem we can take the help of various software's as rapid miner is one of them in which we can remove the missing values as well duplicate values by Appling various constraints.




## IV. STORING THE DATA

Data ware house must ensure that every data in it should not be compromise; as a result every data which is collected should satisfy ETL which is extraction, transformation and loading. These data should be cleaned before loading. Data from the operational system are transported to the data ware house, using two major components, which is the client-server; the client that runs the operational system and the server that handles the data ware housing. When the client starts it communicate with the server, and keeps the communications between it and the server while processing its information. The client cannot see the server directly there by restricting users from having unwanted access to the server holding the data warehouse there by making it difficult for the user to destroy the data.

Secure data ware house must ensure data does not fall into wrong hands; this is paramount when data is integrated in a big data ware house. Once a user has been created and granted privilege to access a data ware house, his activities would need to be controlled, there should be a system log of what he has done.

According to Oracle, "*Organizations struggle with the implementation of this type of strict, granular access control by building application code in each of the front-end applications. Maintaining this type of complex access control code is not only costly, but also risk-prone*." They risk here is a user can be able to pass through using SQL-PLUS due to the application logic. To solve this, organization should implement security in the whole data and provide access control on the data coming from the sources. Transported data for ware housing should be non-volatile which is one of the characteristic of data warehouse; it should be in read-only format. This will protect the data from being modified by the user. A user should be vigorously authenticated before he is granted access to any data in a warehouse, after granted the access he should be restricted to specific roles and data that are relevant to him or her. An example is in a banking system a human resource employee does not have anything to do with the account of staffs or customers, so he should not be granted permission to access the data for accounts. A data mart is a departmental data warehouse that stores only relevant data. These data are extracted from the whole data ware house. The risk of making decision using inconsistent data is considerable a huge problem in business. One of the troubles of data marts is; different organization view different sets of data and tends to give different answers to same question. Consistent security is the watch word in a data warehousing, when u have several data marts on different database server, maintaining and implementing security will be almost impossible. We would need to have the data marts on the same server, to be able to control and make it secure.

Expending a centralized management base will help to protect the data marts from unauthorized users. It will be much more expensive for an organization to try to implement security across multiple data marts, but once it is made centralized it can be easily handled. Since data mart is for individual department, the data should be assign as a responsibility of different user, so they can be held responsible for any lost or damage to the data As the data is stored in the form of cubes the various operations like summarization of cube, projecting front values in form of average ,partitioning the cubes in horizontal and vertical manner and the most commonly used method encryption of data will further enhance the security of data in 2-tier of data ware house.

## V. RETRIEVING THE DATA

Security consists of three major parts; authentication, authorization and auditor. Authentication asks the question: Is this user who he says he is? If he is, grant him specific roles else deny access, and save a log for every activity. In this 3rd tier we are going to look at the security issues which exist in the discovery of these data. Basically data ware house is mainly used by business analysts and executive management only, but with the growth and access of the warehouse the range of potential user increases. The goal of data warehouse is to provide necessary data which is accessible and easy; these data are mostly very sensitive. Therefore the issues about security cover a broad base like; ethical issues, social issues, moral issues, and legal issues.

Independent data sources have to be associated over a network. As the data may be highly sensitive, it is essential to protect it from snooping and similar confidentiality threats. In the communication between the front-end applications and the OLAP server (or the data warehouse in 2-tier environments) as we discussed earlier, usually a client/server connection will be utilized, possibly to remote sites. Even though the information on this architecture is aggregated and less complete, it might have security issues very critical. The use of the Internet or other possibly insecure networks for the above mentioned connections makes suitable security measures necessary. As only some tools support encrypted communication on application level, virtual private network (VPN) technology might be appropriate. Authentication and verification has to be integrated in the data warehouse system, for a comprehensive user identity verification which is needed to grant relevant access and roles to authorized users after verification and deny access to unauthorized user.

The data in a data ware house are kept basically for knowledge discovery (data mining). Therefore these data should be non-volatile and should be in form of reports. The major frontend tools for data warehouses are *OLAP* applications, providing interactive ad-hoc analysis of multidimensional structured data.

A data warehouse is mainly built as an open system. The objective is to make all necessary data accessible as easy as possible. Data from various sources are consolidated and the users of operational systems are not same user with data warehouse. However, the main problem arises from the relational model which is predominating in operational systems while OLAP systems make use of the non-traditional multidimensional model. Accessing control schemes do not map easily due to its non-traditional multidimensional model. Protection is not defined in terms of tables in a data warehouse, but dimensions, graphs, hierarchical paths, granularity levels. The need for proper OLAP security design will help to secure the 3rd tier architecture of data warehouse.

## VI. CONCLUSION

Security and privacy are major issue when designing a data warehouse. Security in any system means guiding the system against unauthorized users, unwanted moving of the data, modifying the data and destroying the data.

**We observed that:**

- Data should be a clean after collection to make the data secure and accurate because any false data that goes into the warehouse will produce wrong analysis from the data mart.
- There should be proper authentication before a user can be granted access into a data warehouse server.
- A log should be taken on every activity on the ware house to check errors and security vulnerability
- Data marts should be on a single big data warehouse for easy security protection and they should be centralized in other to cut cost of protection.
- The need for proper OLAP security design will help to secure the 3rd tier architecture of data warehouse.

## REFERENCES

1. Kimmo Palletvuori, Security of Data Warehousing Server, Helsinki University of Technology
2. John D. Porter and John J. Rome, Lessons from a Successful Data Warehouse Implementation, Arizona State University
3. Security and the Data Warehouse, *An Oracle White Paper, April 2005*
4. Arnon Rosenthal and Edward Sciore, View Security as the Basis for Data Warehouse Security
5. Arnon Rosenthal and Edward Sciore, View Security as the Basis for Data Warehouse Security, Proceedings of the International Workshop on Design and Management of Data Warehouse (DMDW'2000), Sweden, June, 2000.
6. Alan R. Downing, Ira B. Greenberg, and Teresa F. Lunt, ISSUES IN DISTRIBUTED DATABASE SECURITY
7. WANG Baohua, MA Xinqiang , LI Danning A Formal Multilevel Database Security Model
8. Sohail IMRAN, Dr. Irfan Hyder, Security Issues in Databases
9. http://www.information-management-architect.com/data-warehouse-security.html
10. http://blog.simcrest.com/what-is-3-tier-architecture-and-why-do-you-need-it/