Data Warehouse Security Threats and Issues

Akash[1], Nihal Rafeeq[2], Tushith Shukla[3], T. K Koushik Chinnappa[4], Dr. Manjunath Kothari[5]
Students, BE (Appearing), Department of Computer Science and Engineering[1,2,3,4]
HOD, Department of Computer Science and Engineering[5]
Alva's Institute of Engineering and Technology, Mangalore, India

Abstract: *In this review paper we analyzed the ever increasing trend for expansion of information technology which is undisputedly attributed to data warehouses, as its development and redefinition spanned over decades of research and tests. The major constituents of the data warehouses are its humongous amount of data drawn from multiple varied sources which are then feed into decisions makers for further analysis and development of these data. Data warehouses are very much relied upon for making strategic decision making processes. When huge amount of valuable data are stored using data warehouses security must be prioritized at all costs. The brief analysis posed for the security in terms of importance, various approaches and drawbacks regarding them are provided in this review paper.*

Keywords: Data Warehouse (DWH), Data Marts, Security Techniques, Confidentiality, Integrity, Availability, Virtual Data Warehouse

I. INTRODUCTION

Data warehouse is heavily relied upon as it acts as a large stock pile of both current data and historic data as well. It is also considered as the process being most useful is the assembly and the management to get a single detailed view of the required information. The most important task of DWH is to provide an atmosphere consisting of large number of information in a flexible and strategically important data in the most convenient format for the user for further processing, analysis, storage etc. These strategic [1] information range from resource, planning, railways, healthcare sector, forecasting, hotel management [2][3] etc.

In the ever changing world of Data Warehouse the main personality who is still widely regarded as the father of DWH, Bill Inmon said that Data Warehouse is subject-Oriented, integrated, time-variant as well as non-volatile storage of data, it gathered the data from dissimilar types of sources and makes the analyzer able to take the strategic decisions better and faster [4], [5].

The authors are of the view that in the Data Warehousing System the Data Warehouse being the main component is actually a facility that provides an acceptable and consolidated integration of data collection for all the end users reporting and analysis [6].The process of construction of a Data Warehouse is pretty much difficult and also requires assistance and guarantee from business analyst working for an organization also the information technology department [7].
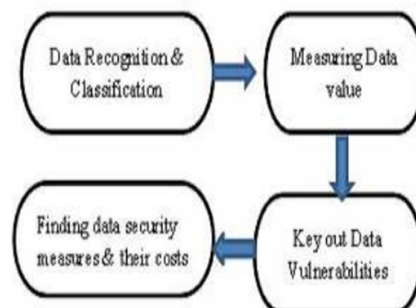

Fig. 1. DWH Security a basic approach

II. BASIC APPROACH FOR DWH SECURITY

There are many approaches in terms of consideration towards data warehouse security as in the business standpoint compromise in DWH is unaffordable. One such being the oracle strategy wherein there are some key considerations are

need made to secure the data from DWH .The first being the end to end security. As there are few steps that includes firstly the collection of data from various sources then it is placed into the DWH after that, those data are distributed to the data marts and also to the analytical servers as well and ultimately these data are to be used by the end users. Data warehouse being a single unit is pretty much easier to secure compare to the security necessary for data marts as data marts are complicated and will be in many numbers also increases the cost exponentially, on the contrast as mentioned before consolidated data warehouse is very much easier to manage as it is low cost mode, It can be equipped with higher level of security. The last important element to be considered for data warehouse is data which being the fundamental component of the data warehouse .So as to protect the precious data a multi layered security model is very much important that can provide reliable security and protection. It is also noted by many data analysts that consolidation is the most vital step of data warehouse .many organization relied on securing data marts but upon realizing the limitations they used single consolidated repository called the Data Warehouse.

There are several steps that are to be applied to data warehouse designing phase [10] which are illustrated in Figure 1.

- **Classification and Recognition of Data:** Listing up of all the information currently accessible from the data warehouse for the end users and making a segregation of this data along with its types. In this step classification is highly prioritized to provide higher levels of security.
- **Data value Measurement:** It is very hard to estimate the average cost of the security, this step however handles the cost estimation for recovery in times of security breaches due to corrupted/loss data or loss of confidentiality.
- **Plotting of data vulnerabilities:** The user doings, account management and technical platform are identified and pointed out in this process.
- **Identification of data security measures and their costs**: Bases on the threats identified the most likely remedies and their costs are then defined for those threats ,then finding out the most cost effective measure for addressing these threats.

These are all the essential steps for the planning up of data warehouse .we can conclude that security provisioning comes with the responsibility of handling the complexity of the system itself .i.e. more complex is the system then more complex it is to manage and thus secure it .Which states the hardship in securing multiple data marts compared to easier protection for single data warehouse.

III. TECHNIQUES USED IN DWH SECURITY

The important issue for the warehousing of data is the proper provisioning of security for all the Data Warehouse [11] which being: Oracles efficient mechanism for the protection of sensitive data of all the DWH by the usage of cryptographic encryption techniques as no further modifications are necessary in the source code [8]. One such efficient method for security being the Advanced Encryption Standard as provided by SQL version 5,but the problem being its efficiency is constrained only to the smaller data bases but fails miserably for all the large databases [12].Then there is the intrusion detection technique as a part of data security enhancement technique which is centric focused on two main elements being hackers attack detection and misuse detection [13].It is also harder to detect malicious behavior to that of normal behavior. Data masking on the other hand successfully was used to segregate the above mentioned two issues. Masking used to change only data values but not the entire format of that data .The main objective in masking of data is to prevent unauthorized access or manipulation towards the sensitive data whatever might the techniques be used.[12].

Data warehouse security includes some layers and all of these must be secured carefully to make the security level more protected. For this security purpose, some aspects are needs to be considered [10]. These aspects are mainly categorized into four areas which are described in Figure 3.

3.1 Integrity and Validation of Data

This process means to ensure that all the data which has been fed into the warehousing system is actually valid data and also accurate in this regard. The filtering process needs to be done including error correction and redundancy in noise in the process of combining information from various sources. It also ensures privacy and confidentiality.
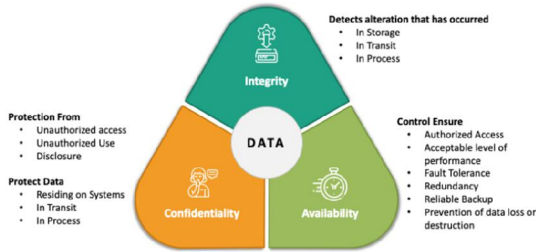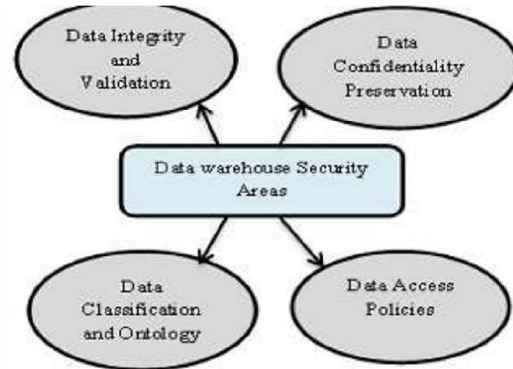
Fig. 2. Issues concerning DWH security

Fig. 3. Data Warehouse security areas

3.2 Access Policy and Restrictions

In access policy methods data protection is done by some access limitations rather than exposing further details.

3.3 Semantics and Data Classification

In order to maintain the proper security level of the system classification is done by understanding the proper nature of the system.

IV. ISSUES IN DATA SECURITY

The important objective used in the encryption methodology is to provide strong privacy for the data with not much attention to the conceptualization and also to maintain high level of performance [15]. Data security focusses on three main issues which are:

- **Confidentiality:** It means protection of the information from all the unauthorized persons
- **Integrity:** It means the assurance which denotes that stored data is accurate and completely trustworthy.
- **Availability:** It ensures the data can be accessed by authorized users and unauthorized users cannot access the precious information.

Some issues regarding these three key fields regarding risks, control and primary focus are described further in the below comparison table1:

Table 1: Issues Comparison Table

CIA Issues	Primary Focus	Risks	Control
Confidentiality	Sensitive information security	Unauthorized access, loss of data	Authentication, data masking, encryption
Integrity	Operational control	Data is no longer accurate	Audit logs
Availability	Business analysis and planning	Business disturbance	Back up storage

V. SECURITY APPROACHES FOR CIA ISSUES

There is huge need for the security of development of the data warehouse from requirement then implementation and maintenance [17].

- **Confidentiality Dealing Approaches:** To maintain the data more confidential by controlling the access, many approaches have been projected. Audit mechanism and authentication method helps to control the data access and these must be installed in data warehouse environment.
- **Integrity Dealing Approaches**: Integrity provides protection to the precious data from malicious activities. In case of an aggregated OLAP query, no inference captured on data and this is the main disadvantage of access control mechanism. In restriction based, techniques are used to prevent from the malicious inference.
- **Restriction based Approaches**: maximum numbers of values aggregated by a query and the top rank of the matrix is used to evaluate the safety of a query [19].
- **Combination of Access and Inference Control Approaches**: If these two approaches are used together then better solution may be generated that removes the security threats. Two-tier architecture having some limitations like unacceptable delay occurs when checking the inference during the run time queries and inference control techniques are not beneficial for the OLAP special features.
- **Approaches Based on Perturbation and Data Masking**: Data masking approach is very beneficial which avoids the data 938 Fig. 4. Security measures for Efficient Decisions disclosure. Oracle used this technique in their database management system. A new data masking technique proposed that consisting only numerical values. Other benefits are: requires low computational efforts, less response time overhead and provides an appropriate security level [20].
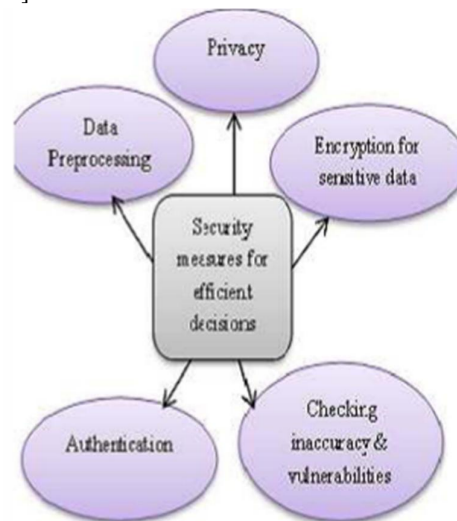

Fig. 4. Data Warehouse security areas

5.1 Considerations for DWH Security

In any system, the most beneficial thing is Security because it provides guidance against the malicious activity like wrong modification and misplacing of data. We have considered some security measures that make our data analysis more efficient.

- For better security protection, data mart must be on single data warehouse.
- In the data warehouse, the most vital thing is correct data. Thus, filter must be used to remove the incorrect data. Data correctness should be ensured before entry into the data warehouse.
- After collecting the data from multiple data sources, all the data must be clean for providing better accuracy.
- Privacy is concerned with individual user. Data privacy internally maintained with the help of different types of integrity constraints.
- To check errors and security vulnerabilities, a log must be taken on each and every activity.

- Proper authentication must be provided when a customer and employee wants to access the required information from the precious data repositories.

VI. PROPOSED METHODOLOGY

In the development of data warehouse one highly recommended method based upon the review of the authors works and further research is to systematically implement the warehouse in an incremental and evolutionary manner. Primarily a very high level of corporate data model is defined for a limited period of time Secondly, data is collected from operational databases and external sources to create many Data Marts in a parallel manner. After which distributed data marts can be constructed to integrate many data marts via hub servers. Finally, a Multi-Tier Virtual Data Warehouse is constructed by using the data of these data marts.

6.1 The Virtual Data Warehouse Security

As most of our very much valuable data are stored in multiple data marts and subsequently these data marts are stored in data warehouses here if only the authorized person wants /information to be searched data marts can be used otherwise he/she can search it from the data warehouse. Most importantly Advantages of virtual data warehouse approach. A big virtual data warehouse will be created by placing all these data marts into one place i.e. Virtual Data warehouse. I have listed some advantages of Virtual Data Warehouse i.e. it is quite easy to build. Other benefits are as follows:

1. It will take less time to access the required information.
2. Less complexity in the big Data warehouse.
3. Easy maintenance of the data warehouse.
4. Data analysis and our strategic decisions will be more efficient.
5. Data can be assembled quickly for one time reports.
6. Less risk of data being lost.
7. From the user perspective, it is scalable and cost-effective(less expensive) data integration.
8. Always altering source systems.
9. As firewall and filters are used, therefore our data will be more protected and accurate and effective decisions can be made.
10. Better possibility of fast and effective business strategic decisions.
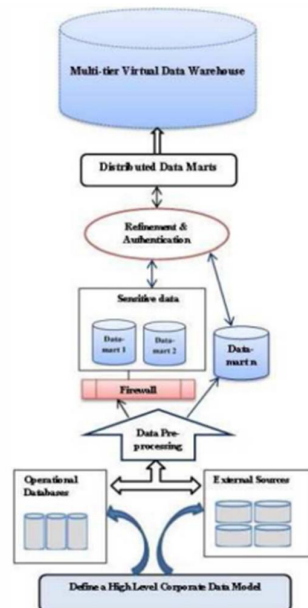
6.2 Hybrid Approach

Fig 5. Approach suggested for DWH security and development

This virtual DWH approach (also called as federated approach) is more feasible than traditional EDWH and there is no need to store data in the centralized repository. All the data tasks like data modeling and ETL are minimized in the case of virtual DWH but still there are few drawbacks in the virtual DWH like repetitive transformation and integration operations, impact to source systems and accessing data shipping over networks. Both the approaches (traditional EDWH and virtual DWH) having some advantages as well as some drawbacks.

Thus the suggestion made by the authors are that of hybrid approach so as to consider the advantages of both approaches. In this approach data is sectored into two data storage units, frequent data to be stored in Virtual Data Warehouse and that which is rarely used can be accessed using traditional data warehouse. Based on approach it is highly likely that data access will be faster and accurate and our data analysis and decisions will be faster and accurate.

VII. 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. Future research in data warehouse security will address several issues. First, with the increasing size of DWHs containing very personal information, privacy-preserving techniques will become more important. This area of research has also received more attention because nation-wide data gathering programs for national security are established. Second, while this theoretical research is certainly important, there are many more aspects to security that need to be considered. A nationwide DWH needs to be secured as an entire system including the mechanisms of data delivery, data querying, and usage. Data warehouse is one of the most critical elements of the Information Technology infrastructure of Organization. As companies are facing problem due to late accessing the required information because very large data is stored in the data warehouse. So this whole data should be separated into two parts i.e. traditional data warehouse and virtual data warehouse. Therefore a hybrid approach will be beneficial for data analysts and managers to get the data faster and to make their decisions more efficient.

REFERENCES

[1]G. K Gupta, Introduction to Data Mining with case studies, Second Edition, Published by PHI Learnings New Delhi, ISBN:978-81-203- 4326-9.

[2] Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques Second Edition, Published by Morgan Kaufinann, San Francisco, ISBN: 978-I-SS860-901-3.

[3] PaulrajPonniah,DATA WAREHOUSING, WILEY STUDENT EDITION,ISBN: 978-81-26S-0919-8.

[4] Surajit Chaudhuri, UmeshwarDayal, An Overview of Data Warehousing and OLAP Technology (Appears in ACM Sigmod Record, March 1997).

[5].ManyaSethi,DATAWAREHOUSING AND OLAP TECHNOLOGY, International Journal of Engineering Research and Applications (IJERA), 2012.

[6] SeemaMaitri,Data Warehousing and Data Mining Concepts and Techniques. Published by NandaniPrakashan, New Delhi,2012.

[7] John D. Porter and John J. Rome Lessons from a Successful Data Warehouse Implementation.

[8] Oracle Corporation, Security and Data Warehouse, Oracle white paper, 200S

[9] Ephraim Mbaka Timothy, Kavisha Duggal, Review on Security Levels of Data Warehouse, International Journal of Advanced Research in Computer Science and Software Engineering, November 2013.

[10] Kimmo Palletvuori, Security of Data Warehousing Server.

[11] Raj Rani, Data Warehouse Security Using Log Based Analysis: A Review, International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, April 2014.

[12] S. Amritpal, Nitin Umesh Implementing Log Based Security in Data Warehouse, International Journal of Advanced Computer, 2013.

[13] Lee, S. Y Low, W. L. , and Wong, P. Y, Learning Fingerprints for a Database Intrusion Detection System.

[14] Mario Piattini+, Jose Antonio Roderof, AUDITING DATA WAREHOUSE SECURlTY ,0-7803-S247-S/99/$1O.00 01 999 IEEE.

[15] SAIQA ALEEM LUIZ FERNANDO CAPRETZ, FAHEEM AHMED, Security Issues in Data Warehouse.

[16] P. Devbandu, and S. Stubblebine, Software Engineering for Security: a Road Map, Proceedings of Conference on the Future of Software Engineering, pp. 227-239. ACM Press, NY, 2000.

[17] E.R. Weippl, Security in Data Warehouses, Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction, L. Bellatreche (Ed.), Chapter IS, pp. 272-27, Information Science Reference, 20 I O.

[18] F. H. Chin and G. Ozsoyoglu, Auditing and lnference Control in Statistical Databases, IEEE Transactions on Software Engineering, Vol. 8, Issue 6, pp.S74S82, 1982.

[19] S.Triki, H. Ben-Abdallah, N. Harbi, and O. Boussaid, Securing the Data Warehouse: a Semi-Automatic Approach for Inference Prevention at the Design Level, Model and Data Engineering Lecture Notes in Computer Science, Vol. 6918, pp. 71-84, Springer-Verlag, 2011.