ISSN No. 0976-5697

Volume 7, No. 6(Special Issue), November 2016

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

Data Warehouse Security Issue

Satish Kumar
Assistant Professor
Department of computer Science
Guru Nanak College Budhlada
satishahuja06@gmail.com

Buta Singh
Assistant Professor
Department of Computer Science
Guru Nanak College Budhlada
er.bootasidhu138@gmail.com

Gagandeep Kaur
Assistant Professor
Department of Computer Science
Guru Nanak College Budhlada
gagan.manshahia@gmail.com

Satish Kumar, Boota Singh, Gagandeep Kaur

Abstract:- Data Warehouse allows storage for large amount of data from heterogeneous operational. Thus provide useful and sensitive information which help decision makers to improve the organisation business process. A Data Warehouse environment must ensure that data collected and stored in one big repository. In today A large amount of data is available on the internet and data Warehouse contains processed data from different sources. So its security is most important issue concept. Security aspects should be considered in the design phase of the Data Warehouse. In the previous time Data Warehouse were used by some high level users but now it is used by everyone for the requirements. This is the main reason to provide security in the Date Warehouse. The aim of this paper to review the security approaches specifically for Data Warehouse environment.

1. INTRODUCTION

Data Warehouse has become a necessity for every organisation due to the wide availability of huge amount of data and to provide a way integrate meaningful data from multiple sources. During the last several years companies are require to big amount of data to support big amount of data to support decision making process. There are different number of users to access the data from Data Warehouse so it is necessary to secure data. There are many definitions of the basic requirements of security. The main concept of Data Warehouse Security is CIA (Confidentiality, integrity and availability). Confidentially means only authorized users to access the Data from the Data Warehouse. Integrity means originality of Data. In other words we can say that the data has received from authentic resources .Availability means information is available all the time. A Data Warehouse may contain large amount of Data such as financial information, credit card numbers, organisation secrets and personal data thus they are unprotected to Cyber attack[1]. A Data Warehouse must ensure that sensitive data does not fall into the wrong hands when data are consolidated into one big storage location[2].

2. SECURITY APPROACHES FOR DATA WAREHOUSE

A DWH is an important part of an organization. It provide the information to the users about the business process as a whole. Security is an important requirements for Data Warehouse development, starting from requirements and go through implementation and maintenance[4] .Security controls are applied on rows ,Column or tables in OLTP(On Line Transaction Processing).While in Data Warehouse need to be accessed by different users for different data[1,5].

Data Transformation, extraction, clearing done before the data are loaded into the Data Warehouse[6]. Security concept must be applied

All layers of Data Warehouse System. Various Security Solution of DWS have been described below.

2.1 DWH Security for confidentiality issue

Confidentiality is used to protect the information form unauthorized users using direct and indirect method. [3]. To provide confidentiality to the Data Warehouse different approaches have been used dealing with access control. Access control mechanism include control on invocation and administration of the Data Warehouse and the Source of the Database. Authentication and audit mechanism is the part of access control and must be installed in a Data Warehouse. Data Warehouse has been accessed by the High Level Users of the Business. Critical access- control issues work on the Front end of the Data Warehouse. Most of the Data Warehouse assumed that there is no need to provide access-control support for a Data Warehouse Front end. Front End Data Warehouse application can provide static and dynamic reporting. Data Warehouse can be defined on a report basis so to implement access control on static reports is not a problem. It is difficult to provide access-control policies for dynamic like data mining queries. This is the main problems of data inference for example a user have not permission to access a particular information, but may retrieve it through o aggregated query.

2.2 DWH Security approaches for integrity

Integrity means to protect the data from accidental or malicious changes such as wrong data insertion, updating or deletion. The main disadvantages of access control mechanism is that they do not collect inferences on data in case of aggregated OLAP query. Inferences on data reason to the integrity issue. For more than twenty five years. Inference control approach have been in statistical and census database[7,8,9].The different approaches can be used to solve this problems.

2.2.1Restriction-Based approaches

In this approach the Safety of a query is determined based on the maximum number of values aggregated by dissimilar queries[8].

The minimum number of values aggregated by a query[10], and the highest rank of the matrix expressing answered queries[11].

Micro-Aggregation is a specific type of aggregations. In partitioning methods a partition is describe on sensitive data and a restriction is a implement on a whole block of a partition for aggregate queries[12,13]. Micro aggregation also changes cluster average. With their sensitive values[14]. Both of the methods are not useful for users and their for may contain meaning less block.

2.2.2. Combined Access and Inferences- Control Approaches

To remove a security threats access control and inferences control work together to provide a best solution. Wand and Jajodia [15] Provide a three – tier security for a data Warehouse two tier can be located in statistical database , such as sensitive data and aggregation queries. This two-tier architecture has some drawbacks: During run time inference checking query processing may result in unacceptable delays, and also under this two-tier architecture inferences-control techniques. To overcome these drawbacks, the

Research has defined a three-tier architecture to provide access control between first and second tiers. The basic lattice can be used and defined the three-tier inference-control model[16]. The first method is used for statistical database. The second method is used remove the limitations of inference-control methods. The work claim of these two methods are more appropriate for data Warehouse and OLAP System.

2.2.3 Modelling-based Approach to DWH Security:

This approach consists of three phase of DWH security. The first phase locates sensitive data from DWH with the help of security designers and experience person in the field. In the second phase to generate inference graph depends on a class diagram is constructed to detect elements which may cause inferences in future[17]. The security designers to generate a difference between precise and partial. Inferences precise inferences means that only exact information is provided but partial information to provide only partial information.

The inference graph contain a set of nodes representing the data. These nodes are connected with each other through oriented arcs that representing the direction of inference and its type (partial or precise). Data Warehouse Schemata worked automatically using UML annotations which on and off the elements that may load to both types of inferences. There are two advantages of these approaches. Independence of data domain, and useful of present data to detect inferences.

2.2.4 Data Masking and perturbation-Based Security approaches

With the help of data-masking we protect the data from unauthorized user. Using data masking original data values can be changed with new values. Data-masking are performed with the help of oracle in their DBMS[18]. Encryption is used to encrypt the data In data-masking The oracle has developed different version of Transparent Data Encryption(TDE)[19,20].Data masking technique for data-warehouse consists only numeric values proposed by Santos et al.[21]. This approach was based on mathematical operators Such as division, reminder and two simple arithmetic operation that can be used without changing user applications and DBMS Source code. They claimed that the proposed f formula providing an appropriate security level.

2.3 Data Warehouse Security Approach for the availability issue

Data availability is most important concept of Data Warehouse System. This approach is used to recover the data from real time corruption. Data replication is used to restore damage data using with the help of different solution. In this way we can perform database maintenance in a easy way. When centralized Server contain database then RAID Architecture can be used for mirroring data on system[22,23].The Aim of the organisation to implement Data Warehouse in low-cost machines for cost optimization purpose. RAID Technology is not suitable for this type of situation because only one disk drive is present in the RAID Technology. The different commercial solution for the Data Warehouse such as Oracle[24] and Aster Data[25] in today's market. With the help of hamming codes approach to recover corrupted data using error-correction codes. Marsh and Schneider[26] describe a new technique for distributed storage have the same features which described in early plus encryption methods.

3. CONCLUSION

In this papers we review the existing Data Warehouse Security solution, discuss their issue and their impact on Data Warehouse scalability and performance requirements. The proposed solutions are inefficient for use in Data Warehouse Security environments. A Data Warehouse requires specific functionality. Data Warehouse Security is and active research to any industrial project. Further research in Data Warehouse Security is needed to address the issues discussed above because many more aspects remain to be considered.

REFERENCES

[1] H. Inmon, *Building the Data Warehouse*, 3rd ed., John Wiley, USA, 2002.

[2] N. Yuhanna, *Your Enterprise Database SecurityStrategy*, Forrester Research, 2010.

[3] C. Farkas, and S. Jajodia, The Inference Problem:a Survey, *ACM SIGKDD Explorations Newsletter*,Vol. 4, Issue 2, pp. 6-11, December 2002.

[4] 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.

[5] N. Kaite, M. Stolba and A.Y. Tjoa, A Prototype Model for Data Warehouse Security Based on Metadata, *International Conference of Database and Expert Systems, Vienna, pp. 300-308*, IEEE Press, August, 1998.

[6] E.R. Weippl, Security in Data Warehouses, Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction, L. Bellatreche (Ed.), Chapter 15, pp. 272-27, *Information Science Reference*, 2010.

[7] N. M. Adam and J. C. Wortmann, Security- Control Methods for Statistical Databases: a Comparative Study, *ACM Computing Surveys*, Vol. 21, Issue 4, pp. 515–556, December, 1989.

[8] D.E. Denning and J. Schlorer, Inference Controls for Statistical Databases, *IEEE Computer*, Vol. 16, Issue 7, pp. 69–82, IEEE Computer Society 1983.

[9] L. Willenborg, and T. DeWalal, *Statistical Disclosure Control in Practice*, Springer Verlag, New York, 1996.

[10] D. Dobkin, A.K. Jones and R.J. Lipton, Secure Databases: Protection Against User Influence, *ACM Transactions on Database Systems*, Vol. 4, Issue 1, pp. 97–106, 1979.

[11] F. H. Chin and G. Ozsoyoglu, Auditing and Inference Control in Statistical Databases, *IEEE Transactions on Software Engineering*, Vol. 8, Issue 6, pp. 574–582, 1982.

[12] F. H. Chin and G. Ozsoyoglu, Statistical Database Design, *ACM Transactions on Database Systems*, Vol. 6, Issue 1, pp. 113–139, 1981.

[13] C.T. Yu. and F.Y. Chin, A Study on the Protection of Statistical Data- bases, *Proceedings of ACM SIGMOD International Conference on Management of Data*, pp. 169–181, 1977.

[14] J.M. Mateo-Sanz, J.M. and J. Domingo-Ferrer, A Method for Data-oriented Multivariate Micro Aggregation, *Proceeding of Conference on Statistical Data Protection*, pp. 89–99, 1998.

[15] L. Wang and S. Jajodia, Security in Data Warehouses and OLAP Systems, *in Handbook of Database Security*, Springer Verlag, pp. 191-212, 2008.

[16] L.Wang, S. Jajodia and D. Wijesekera, Lattice-based Inference Control in Data Cubes, in book *Preserving Privacy in On-Line Analytical Processing (OLAP)*, Springer, pp. 119-145, 2007.

[17] 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 EngineeringLecture Notes in Computer Science*, Vol. 6918, pp.71-84, Springer-Verlag, 2011.

[18] Oracle Corporation, *Oracle Advanced Security Transparent Data Encryption Best Practices*, Oracle White Paper, July 2010.

[19] Oracle Corporation, *Security and the Data Warehouse*, Oracle White Paper, April 2005.

[20] Oracle Corporation, *Data Masking Best Practices*, Oracle White Paper, July 2010.

[21] R. J. Santos, J. Bernardino and M. Vieira , A Data Masking Technique for Data Warehouses, *Proceedings of the 15th Symposium on International Database Engineering & Applications*, pp. 61-69, ACM Digital Library, 2011.

[22] IBM Corporation, *Understanding RAID Level 5*, IBM Systems Software Information Center, 2007.

[23] IBM Corporation, *Understanding RAID Level 6*,IBM Systems Software Information Center, 2007.

[24] Oracle, *Oracle Real Application Clusters (RAC)*,www.oracle.com/us/products/database/options/real -applicationclusters/index.htm, September 2010.

[25] AsterData Systems, Aster Data nCluster: Always on, for 24x7 Big Data Analytics, http://www.asterdata.com/product/alwayson.php, 2010.

[26] M.A. Marsh and F.B. Schneider, CODEX: a Robust and Secure Secret Distribution System, *IEEE Transactions on Dependable and Secure Computing*, Vol. 1, Issue 1 , pp. 34-47, 2004.