Solutions to Current Challenges in Data Warehouse Security

Mugdha M Marathe[1] Pravin S. Metkewar[2]
[1]Student [2]Associate Professor
[1,2]SICSR, Affiliated to Symbiosis International University. Pune, Maharashtra, India

Abstract— Data warehouse gives stockpiling to enormous sums of historical data from heterogeneous operational sources in the form of multidimensional views thus helping decision-makers to improve the organization's business processes. Data Warehouses are the organization's most important resources in what concerns important business data, making them an engaging target for malicious inside and outside attackers. The security of such a huge information base is crucial for the sustainability and reliability of the data warehouse. The security requirements of a Data Warehouse are different from other distributed systems. The aim of this paper is to provide a view on some of the vulnerabilities existing in data warehouse along with various solutions to make data warehouse secure.

Key words: Data Warehouse, Security, Security Issues, Privacy, Data Security, Data Integrity

I. INTRODUCTION

Data warehouses are basically databases putting away united verifiable and current business data required for decision sustenance purposes. The Data Warehouse reflects the measures and consequences of the business, and how and when it happens. Data warehouses contain a great variety of data which offers a sound picture of the organization's economic conditions at a certain moment. A company's database may contain massive amounts of organizational data such as information related to finance, numbers of credit cards, organizational trade secrets and personal data. This makes them a key asset for any enterprise.

Unfortunately, this also makes them an engaging focus for noxious inside and outside aggressors. Data breach is becoming a major concern as more businesses receive information stockpiling innovation. Cybercriminals are snatching this opportunity to attack susceptible businesses unfamiliar with data security. Most business cyber-attacks are meek in nature, but tremendously effective against networks that don't practice basic security protocols.

II. DATA WAREHOUSE SECURITY

Many publishes security statistics show that the number of attacks on data warehouses is increasing continuously. Securing a data warehouse is an operation that requires much care and concentration of the efforts because there are a few specific features that may appear.

Data security focuses mainly on three issues: confidentiality, integrity and availability. These concepts are also known by the acronym CIA.

Confidentiality emphasizes on protection of information from illegal revelation, either by indirect logical implication or by direct retrieval. Integrity involves data safeguarding from unintentional or malicious changes such as false information insertion, sullying, or obliteration. Accessibility ensures that data are accessible to all authorized users at any time.

Although available solutions have been proven to be methodically effective, they are infeasible or at least inefficient for a DWH environment.

III. SECURITY APPROACHES FOR DWH

Different security arrangements have been proposed in the DWH writing and are portrayed underneath, ordered by how they address the fundamental security concerns, for example, CIA.

A. DWH security approaches for Confidentiality Issues

Confidentiality underlines security of data from unapproved revelation, either by roundabout sensible derivation or by direct recovery.

Keeping in mind the end goal to address DWH confidentiality concerns, numerous methodologies have been proposed managing access control. Access control mechanisms involve controlling both invocation and administration of the DWH and the source databases. Verification and review systems additionally fall under access control and should be introduced in a DWH domain. Access control instruments include controlling both summon and organization of the DWH and the source databases. Validation and review instruments additionally fall under access control and should be introduced in a DWH domain. Basic access-control issues additionally emerge at the front end of a DWH.

Most DWH or OLAP sellers expect that there is no compelling reason to give fine-grained access-control support for a DWH front end since it ruins revelation of investigative data. Be that as it may, this suspicion is not suitable in light of the fact that numerous clients can get to scientific apparatuses to question the DWH. Front-end DWH applications can give both static and element reporting. Forcing access control on static reports is not an issue since it can be characterized on a report premise. For element reporting like information mining questions, it is hard to give fitting access-control approaches. This prompts the issue of information deduction; for instance, a client may not be approved to acquire specific data, but rather might recover it through a collected inquiry.

1) Mandatory Access Control approach

MAC is the stringiest of all the access control based approaches. At the point when a client endeavours to get to an asset under Mandatory Access Control the operating system checks the client's classification and classes and looks at them to the properties of the object's security name. If the client's accreditations match the MAC security mark properties of the article access is permitted.

2) Discretionary Access Control Approach

In Discretionary Access Control (DAC) every client or client group is permitted to control access of their own information. Rather than a security name on account of MAC, a client has a list which known as the access control list(ACL) through which it can choose which client to offer consent to get to its individual information together with the

level of access gave to that client (whether read just or read, write and so forth).

3) Role Based Access Control Approach

RBAC is based on the client's role in the organization. It is the most widely used mechanism. It takes practical approach in building the access control. A client or a group of client's is provided access to an object based on their roles in the organization.

4) Rule Based Access Control Approach

In this approach, a set of rules is defined. For example, rules are defined in an Access Control List. Unlike in MAC, the rules here are not stringent and can be modified as and when necessary.

B. DWH Security approaches for Integrity

Integrity involves data insurance from unintentional or malignant changes, for example, false information insertion, sullying or annihilation. The drawback of access-control components is that they don't catch derivations on information on account of amasses of an OLAP question.

The proposed approaches can be categorized into restriction based and perturbation based techniques.

Restriction based inference control techniques simply deny unsafe queries to prevent malicious inference. Perturbation based techniques add noise to data, swap data or modify the original data and can also apply data modification to each query dynamically.

Integrity includes information security from unplanned or noxious changes, for example, false information insertion, pollution, or devastation. The hindrance of access-control instruments is that they don't catch surmising on information on account of an amassed OLAP question. Surmising's on information lead to the honesty issue. For over thirty years, deduction control approaches have been examined in factual and enumeration databases [7, 8, and 9]. The proposed methodologies can be sorted into limitation based and bother based systems. Limitation based surmising control strategies basically deny risky inquiries to anticipate vindictive induction. Annoyance strategies add clamour to information, swap information, or change the first information and can likewise apply information adjustment to every question progressively. The methodologies exhibited to fathom the respectability issue can be ordered further as depicted beneath.

1) Restriction based Approach

In restriction based methods, the security of an inquiry is resolved in view of the most extreme number of qualities totalled by unique inquiries the base number of qualities collected by a question and the most noteworthy rank of the grid communicating addressed inquiries.

Miniaturized scale conglomeration and dividing considers particular kind of totals. In parcelling strategies, an allotment is characterized on delicate information, and a limitation is connected on a complete square of a segment for total questions Small scale total additionally replaces group midpoints with their touchy qualities [14]. Both strategies are not taking into account dimensional pecking orders and in this manner might contain good for nothing hinders that are not valuable for clients.

2) Combined Access and Inference-Control based Approach

So as to uproot security dangers, access control and deduction control together can give a decent arrangement. Guaranteeing security ought not to influence the handiness of DWH and OLAP frameworks. Normally, two levels can be found in measurable databases, for example, delicate information and accumulation questions. This two-level design has some intrinsic disadvantages: derivation checking amid run-time inquiry preparing might bring about unsatisfactory postponements, furthermore under this two-level engineering, surmising control systems can't profit by the unique qualities of OLAP. To defeat these downsides, the exploration has characterized a three-level design to give access control between the first and second levels and derivation control between the second and third levels.

The fundamental grid based deduction strategy can be utilized and actualized on the three-level surmising control model. The main system utilized existing deduction control techniques for factual databases, while the second approach was intended to evacuate the confinements of existing induction control strategies. The work asserts that both techniques could be connected on the premise of a three-level induction control engineering that is more suitable for DWH and OLAP frameworks particularly.

3) Modelling Based Approach

This approach consists of three phases. In the first phase, identification of sensitive data is done from Data warehouse schema. In the second phase inference graph is constructed, in order to detect elements, based on class diagram. In third phase, Data warehouse schema are augmented automatically by UML annotations.

4) Data Masking and Perturbation-Based Security Approaches

Data revelation can be easily evaded by data-masking approaches. Using data masking, original data values can be substituted or altered. In data masking, encryption is an advanced form of imposing confidentiality. Some well-known encryption algorithms are AES and 3DES.

K-anonymity-based approach also reveals sensitive information without intimidating privacy. K-anonymity and inference-control methods can be combined to obtain a better solution.

C. DWH security approaches for Availability issues

Data availability is extreme important in any Data Warehouse system. This involves data retrieval from real-time corruption or improper data modification and continuous user access. Data duplication is done to be able to restore impaired data using many anticipated solutions. Also database downtime because of maintenance intervention can be avoided.

1) RAID Based Approach

This approach is used for mirroring data in the network where the database is stored on the centralized server. In any case, organizations have been executing their DWHs in minimal effort machines for cost-enhancement purposes.

2) Hamming Code Approach

This is another approach to recover impaired data. It uses error-correction codes for doing the same. The proposed data storing framework makes it conceivable to recover

tainted information hinders by utilizing blunder adjusting codes, remapping bad pieces, and duplicating blocks.

IV. CONCLUSION

This study has given a review of existing DWH security arrangements, talking about their issues and their effect on DWH adaptability and execution prerequisites. It has ended up obvious that the proposed arrangements are infeasible for use in DWH situations. A DWH requires particular usefulness with tight versatility and execution necessities. A complete arrangement is in this way required makes it conceivable to address these mandates. DWH security is a dynamic examination to any mechanical task. Further research in DWH security is expected to address the issues talked about in light of the fact that numerous more perspectives stay to be considered, and there numerous open inquiries to be replied.

REFERENCES

[1] H. Inmon, Building the Data Warehouse, 3rd ed., John Wiley, USA, 2002.
[2] N. Yuhanna, Your Enterprise Database Security Strategy, Forrester Research, 2010.
[3] S. Triki, H. Ben-Abdallah, N. Harbi, and O. Boussaid, Securing the Data Warehouse: a SemiAutomatic 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.
[4] C. Farkas, and S. Jajodia, The Inference Problem: a Survey, ACM SIGKDD Explorations Newsletter, Vol. 4, Issue 2, pp. 6-11, December 2002.
[5] N. M. Adam and J. C. Wortmann, SecurityControl Methods for Statistical Databases: a Comparative Study, ACM Computing Surveys, Vol. 21, Issue 4, pp. 515–556, December, 1989.
[6] 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.
[7] 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.
[8] Prabhakaran, L.N. Bairavasundaram, N. Agrawal, H.S. Gunawi,, A.C. Arpaci-Dusseau and R.H. Arpaci-Dusseau, IRON file systems, International Symposis on Operating System Principles (SOSP), pp. 206-220, Brighton, UK, October, 2005.
[9] IBM Corporation, "Understanding RAID level 6", IBM Systems Software Information Center, 2007.
[10] Oracle Corporation, "Data masking best practices", Oracle White Paper, July 2010.
[11] IBM Corporation, "Understanding RAID level 5", IBM Systems Software Information Center, 2007.
[12] S. Lujan-Mora, S, J. Trujillo and I.Y. Song, "A UML profile for multidimensional modeling in Data Warehouses", Data & Knowledge Engineering, Vol. 59, Issue 3, pp. 725–769, Science Direct, 2006.
[13] E. Fernandez-Medina, J. Trujillo, P. Villarroel and M. Piattini, "Extending UML for designing a secure data warehouse, conceptual modeling", Lecture Notes in Computer Science, Vol. 3288, pp. 217230, 2006.
[14] E. Fernandez-Medina, E., J. Trujillo and R. Villarroel, "Developing secure data warehouses with a UML extension", Information Systems, Vol. 32, pp. 826-856, Science Direct, 2007.

