✚IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY
SECURITY IN DATA WAREHOUSE: A QUINTESSENTIAL APPROACH

Rajat Verma[*1] **& Harshita Mishra**[2]

[*1&2]M.Tech (Computer Science & Engineering) Department of Computer Science & Engineering,
Amity School of Engineering & Technology, Amity University, Lucknow

ABSTRACT

Data that means unprocessed raw facts and figures are increasing at an alarming rate. This is due to the increased population. This is also due in the enhancement in the number of devices connecting to the internet. By 2020, 50 billion devices must be connected to the internet. The internet of things commonly abbreviated as IOT is a major concern in this. The Confidentiality, availability, integrity model is mentioned in this paper. Characteristics of Data warehouse-OLAP is also mentioned in this paper. Security concern is also mentioned in this paper. Different types of attacks are also mentioned in this paper. All this will certainly contribute to the research sector.

Keywords: Data, Warehouse, OLAP, Solutions,C-I-A.

I.   INTRODUCTION

By every now and then, the unprocessed raw facts and figures are increasing at a tremendous pace. That means, if I take an example of today that is July 24, 2018 at 8:36 pm that suppose 3 lacs entries are there, and at 8:37 pm it becomes 4 lacs. That means at every millisecond or micro second some unprocessed facts are generating. It is full of noise and inconsistency. The main aim is that it should be kept secure, so that it may remain same throughout and may not be altered by an unauthorized party. This requires various parameters such as confidentiality, availability, integrity etc. [1]. It can be depicted from fig.1.

Fig.1. C-I-A Model for Data Security

- **Confidentiality**: It basically means that the processed data or we can say information, should be between sender and the receiver. That means it should not be leaked to other people. Other people may misuse it, or can make fun of the topic that was discussed.
- **Integrity**: Data Integrity means that its original form should not be changed. That means it should be precise as well as consistent.
- **Availability**: This is another aspect of data security. It means that data should be available whenever it is required, whatever the condition is!

But the point is that, many of the people doesn't take any steps to secure their data. They also don't make up the back up of any data, means their approach is very casual even in the case of private information. This private

information could be bank account numbers, Automated Teller Machine (ATM) [2] pin number etc. This can lead to very dangerous symptoms and they could become bankrupt!

Fig.2. Bankruptcy

There could be other effects also that may affect diverse things.

II.   RESEARCH BACKGROUND

Data warehouse
It is basically the repository of unprocessed raw facts and figures from diverse sources [3]. It is used to support the actions that is to be taken by the management team of the organization. Its illustration is depicted in Fig.3.

Fig.3 Data Warehouse

Index-
- ETL: Extract Transform Load [4]
- CRM: Customer Relationship Management [5]
- ERP: Enterprise Resource Planning [6]
- Database: Collection of inter related records[7]
- Data Marts: Subsets of Data Warehouse[8]

Online Analytical processing (OLAP) [9] is a term that is used in context to Data warehouse. OLAP is illustrated in Fig.4.
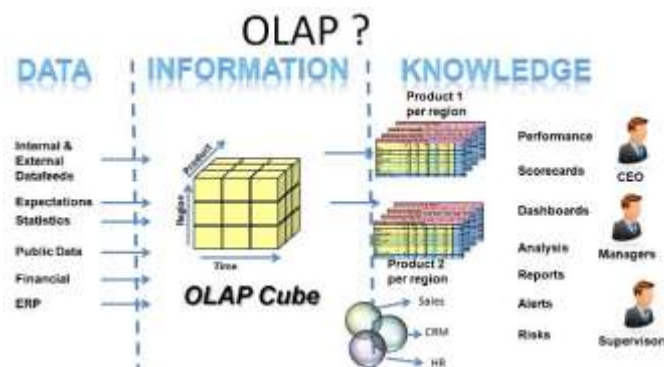

Fig.4 Online Analytical Processing

III.   CHARACTERISTICS OF OLAP

It is depicted in the tabular format.

Parameters	Online Analytical Processing
Process	Online Analysis & data retrieving
Characteristics	Large Volume of data
Functionality	Online database query management system
Method	Data Warehouse
Query	Select Operations
Table	Tables are not normalized
Source	Different OLTP databases
Data Integrity	Not an issue
Response Time	Seconds to Minutes
Data Quality	Might not be organized
Usefulness	Planning, Problem Solving, Data Support
Operation	Only Read and rarely write
Audience	Customer-Oriented
Query-Type	Complex
Back up	Not important when compared to OLTP
Design	Subject Oriented
User Type	Data knowledge users
Purpose	Analysis by category
Performance metric	Query Throughput
Number of Users	Hundreds of Users
Productivity	Helps in Increasing Productivity

IV.   SECURITY CONCERN IN DATA WAREHOUSE

Security [10] is a crucial parameter in terms of data. Data Warehouse is basically the collection of huge data. The main objective is to secure the data warehouse. If we are late in adding security parameters then it can damage the performance of data warehouse to a large extent. So it is always recommended that security policies should be added as soon as possible. It can be depicted by Fig.5.

Fig.5. Security Cycle in Data Warehouse (Oracle)

The 4 major terms in this scenario are:
- Access Control [11]
- Data Protection [12]
- Monitoring [13]
- User Management [14]

The four terminologies that are particularly affected by the security procedures are:
- User Access [15]
- Data Load [16]
- Data Movement [17]
- Query Generation [18]

V.   SECURE DATA WAREHOUSING SOLUTION
It can be depicted with the following illustration:

Fig.6. Secure Data Warehousing Solution

VI.   DIFFERENT TYPES OF ATTACKS
There are mainly two types of attacks:
- Active Attacks [19]
- Passive Attacks [20]

Active Attacks: This is basically the exploitation of the network in which the hacker attempts to make changes in the data that is sent from the sender to the receiver.
There are various kinds of Active attacks that are mentioned below-
- Denial of Service
- Distributed Denial of Service
- Session Replay
- Masquerade
- Message modification
- Trojans

It can be depicted by illustration no 7.

Fig.7. illustration of active attacks

Passive attacks: It is a type of attack in which only monitoring is being done. This can be depicted as reading of messages, tracking internet details etc. It can be depicted in Fig.8.

Fig.8. Passive attack

VII. OTHER TYPES OF THREATS

- Virus
- Worms
- Spam
- Ransomware
- Botnets
- Malwares

VIII. CONCLUSION

In today's modern world, data is increasing at an enormous velocity. For this, security is really a matter of concern as if an unauthorized access takes place, security parameters can help in tracking that. As the information can be negative for some aspect and positive for some aspect. Every data should be monitored from time to time so that data should be kept authorized in every case.

REFERENCES

[1] Avizienis, A., Laprie, J. C., Randell, B., & Landwehr, C. (2004). Basic concepts and taxonomy of dependable and secure computing. IEEE transactions on dependable and secure computing, 1(1), 11-33.

[2] Magee, P. D., Barker, D., Nelson, D. S., Mercer, S. A., Persons, M., Maczuzak, M., & Warner Jr, R. J. (2004). U.S. Patent Application No. 29/187,158.

[3] Journel, A. G. (2002). Combining knowledge from diverse sources: An alternative to traditional data independence hypotheses. Mathematical geology, 34(5), 573-596.

[4] Vassiliadis, P. (2009). A survey of extract–transform–load technology. International Journal of Data Warehousing and Mining (IJDWM), 5(3), 1-27.

[5] Buttle, F. (2004). Customer relationship management. Routledge.

[6] Arik Ragowsky, T. M. S. (2002). Enterprise resource planning. Journal of Management Information Systems, 19(1), 11-15.

[7] Bernstein, P. A., Hadzilacos, V., & Goodman, N. (1987). Concurrency control and recovery in database systems.

[8] Bonifati, A., Cattaneo, F., Ceri, S., Fuggetta, A., & Paraboschi, S. (2001). Designing data marts for data warehouses. ACM transactions on software engineering and methodology, 10(4), 452-483.

[9] Gladden, P. E. (1998). U.S. Patent No. 5,765,028. Washington, DC: U.S. Patent and Trademark Office.

[10] Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In Collaboration Technologies and Systems (CTS), 2013 International Conference on (pp. 42-47). IEEE.

[11] Sandhu, R. S., Coyne, E. J., Feinstein, H. L., & Youman, C. E. (1996). Role-based access control models. Computer, 29(2), 38-47.

[12] Bygrave, L. A. (2002). Data protection law. Wolters Kluwer Law & Business.

[13] Valdes, A., & Skinner, K. (2000, October). Adaptive, model-based monitoring for cyber-attack detection. In International Workshop on Recent Advances in Intrusion Detection (pp. 80-93). Springer, Berlin, Heidelberg.

[14] Bianchessi, N., Cordeau, J. F., Desrosiers, J., Laporte, G., & Raymond, V. (2007). A heuristic for the multi-satellite, multi-orbit and multi-user management of earth observation satellites. European Journal of Operational Research, 177(2), 750-762.

[15] Yan, T. W., Jacobsen, M., Garcia-Molina, H., & Dayal, U. (1996). From user access patterns to dynamic hypertext linking. Stanford InfoLab.

[16] Stewart, R. G., & Ipri, A. C. (1999). U.S. Patent No. 5,952,789. Washington, DC: U.S. Patent and Trademark Office.

[17] [17] movement and access service for wide area computing systems. In Proceedings of the sixth workshop on I/O in parallel and distributed systems (pp. 78-88). ACM.

[18] Kamvar, S. D., Haveliwala, T. H., & Jeh, G. M. (2010). U.S. Patent No. 7,836,044. Washington, DC: U.S. Patent and Trademark Office. .

[19] Sunar, B., Martin, W. J., & Stinson, D. R. (2007). A provably secure true random number generator with built-in tolerance to active attacks. IEEE Transactions on computers, 56(1).

[20] Kong, J., Hong, X., & Gerla, M. (2003, October). A new set of passive routing attacks in mobile ad hoc networks. In Military Communications Conference, 2003. MILCOM'03. 2003 IEEE (Vol. 2, pp. 796-801). IEEE

CITE AN ARTICLE

Verma, R., & Mishra, H. (2018). SECURITY IN DATA WAREHOUSE: A QUINTESSENTIAL APPROACH. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, 7*(8), 18-23.