A SURVEY ON CURRENT SECURITY PERSPECTIVES IN DATA WAREHOUSES

G.Thangaraju#1, Dr.X.Agnes Kala Rani*2

#1Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore-641 021
*2Professor, Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore-641 021

gtrmsu@gmail.com
agneskala72@gmail.com

Abstract—Business Intelligence and Data Warehouse is most important role player in the enterprises. Data warehouse (DW) is a collection of integrated databases. The DWHDB are designed to support managerial decision-making and problem-solving functions. The data’s available in the DWHDB are very perceptive. Such perceptive information needs security so as to prevent this perceptive information from the unauthorized users and attackers. The Current research phenomena in data warehouse have shown number of methods to secure the data in DWHDB at different perspective. These methods are defined at physical level, logical level, conceptual level and business level. In this paper provide a comprehensive review of various methods or approaches to prevent or secure the data in the data warehouse environment [1].

Keywords: Data warehouse, Security, User, Confidentiality, Integrity, Availability.

I. INTRODUCTION

A data warehouse is a structured extensible environment designed for the analysis of nonvolatile data, logically and physically transformed from multiple source applications align with business structure, updated and maintained for a long time period, expressed in simple business terms, and summarized for quick analysis.

In a business environment the updated report is needed to take a business decision, for the reporting purpose the report server are interconnected with the data warehouse server. The stages of Data warehousing are i).Data Collection from the different OLTP systems ii).Data cleaning for taking a necessary information iii).Data Integration to integrate the data retrieved from the different sources iv).Data modeling for business needs, the modeling process contains a different methods we can choose any one method for our business need (Star Schema, Snow flex Schema. v).Building a Data mart, Data mart a logical and physical subset of the data warehouse’s presentations area. Originally, data marts where defined as highly aggregated subsets of data often chosen to answer a specific business question. vi).Metadata, Metadata is data about data, Examples of metadata include data element descriptions, data type descriptions, attribute/property descriptions, range/domain descriptions and process/method descriptions. Metadata is stored in a data dictionary and repository. It insulates the data warehouse from changes in the schema of operational systems.vii).Reporting (OLAP)-

The processes of translating data to presentation formats via a pre-defined or ad-hoc queries [2].

The users of Data warehouses are Analysts, Managers and Executives, operational personnel, customers, suppliers and end users of all domains. The above staging levels are having a security in depending upon their needs. In the next section we discussed about the security issues in data warehousing.

II. SECURITY AND DATA WAREHOUSE

The Security requirements of the Data warehouse are not unlike those of other distributed computing environment or systems [2]. The data warehouse environment has the different phases during the development phase we have consider the following basic security principles
The above mentioned seven stages are perform certain action here we can show the action based on the stages are in the following table.

<table>
<thead>
<tr>
<th>S.No</th>
<th>Stage</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Identifying the data</td>
<td>Collect the all information about the data like database name, tables, columns, rows, and profile of the DWH, the user of the data.</td>
</tr>
<tr>
<td>02</td>
<td>Classifying the data</td>
<td>Classifying the data is essential to satisfy the security requirements for data Confidentiality, Integrity, and Availability. In DWH the classification are done on the following categories a).Public or least sensitive data b).Confidential or Moderately sensitive data c).Top secret or most sensitive data</td>
</tr>
<tr>
<td>03</td>
<td>Quantifying the Data</td>
<td>Reconstruct lost data, restore the integrity of corrupted, intercepted or fabricated data, not make timely decisions due to denial of service, and pay financial liability for public disclosure of confidential data.</td>
</tr>
<tr>
<td>05</td>
<td>Identifying protective measures and their costs</td>
<td>Protective measures for DW data might include: i. The human wall ii. User access classification iii. Access controls iv. Integrity controls v. data encryption vi. Partitioning vii. Development controls</td>
</tr>
<tr>
<td>06</td>
<td>Selecting cost-effective Security measures</td>
<td>The security cost is calculated as cumulate the expenses occur in the previous stages are counted here, apart from the cost the security measure in the DWH consider the compatibility, potential impact on DWH performance and adaptability factors.</td>
</tr>
<tr>
<td>07</td>
<td>Evaluating the</td>
<td>The evaluations should be effectiveness of security measure always conducted to determine whether the measures are i. small, simple and straightforward; ii.carefully analyzed, tested and verified iii. used properly and selectively iv. Elastic.</td>
</tr>
</tbody>
</table>

Table 1: The Stages and action done by the stages of DWH development phase security considerations.

The basic theory is that the any information system implanting a data warehouse uses all the necessary means for protecting the availability and integrity of the underlying databases, as well as the confidentiality and privacy of the information [3].

The information flow between in the DWH environment is having more security with respect to an accesses control mechanism. Access control is the collection of mechanisms that permits managers of a system to exercise a directing or restraining influence over the behavior, use, and content of a system. It permits management to specify what users can do, which resources they can access, and what operations they can perform on a system [4].

In the current scenario of data warehousing technology has implemented a lot of security features in different aspects like the Access Control - MAC, RBAC, DAC in both OLTP and OLAP [5].

Security model in Conceptual, logical and physical design have the different approaches like Model Driven Approaches (MDA) provided by the Object Management Group (OMG). This approach allows as defining models at different abstraction level, In Business level used as Computer Independent Model (CIM), like Platform Independent Model (PIM) and Platform Specific Model (PSM) for conceptual and logical model vice versa[6].

The next session of this paper represent the some security solutions suggested by some author.

III. DATA WAREHOUSE SECURITY SOLUTIONS

The security solution of DWH in two aspects first one is the Preventive Data Security Solutions, the second one is Reactive Data Security Solutions.

1. Preventive Data Security Solutions

This technique are used for protecting data in advance of attacks, such as implementing concurrency constraints and referential integrity, data masking, data access policies and encryption techniques for changing original data values, and checksum for integrity checks on changed data.

The current DBMS allows defining data validation rules, RBA policies, referential integrity constrains, and comply with ACID requirements, all of which assure data confidentiality, consistency and correctness, up to certain point. For data corruptions checksums techniques are used and for distinguishing data either original or tampered data is using the signature in all records of the Data warehouse, as published in [7].The CRC, MD5 and SHA algorithms are
used for detecting correctness error. The Data masking technique is used for enforcing data privacy, in oracle 11g uses the Transparent Data Encryption (TDE) for data privacy. The Advanced Encryption Standard methods used in MySQL v5 for data encryption function. However, given the increase of sophisticated attacks and rising internal theft, preventive security techniques and traditional CIA features are no, longer enough to protect data. This has lead to the development of reactive data security techniques [7].

2. Reactive Data Security Solutions

This solution consists of auto-recovery, fault-tolerance, intrusion detection, auto-repair, among others, which try to protect data from attackers able to bypass preventive security techniques. The main task in data warehouse availability involves the real time recovery of damaged or incorrectly modified data and continuous 24/7 user access. The RAID Mechanism is used rectify these kind of problems [8].

IV. RELATED WORKS

In this section some essential contributions to improve the security perspective in DWHs.

Ricardo Jorge Santos et al [9] proposed a data masking solution to enhance data privacy in DWHs. He also uses one of the masked fact tables masking key for enabling false data injection, increasing the security strength against the attackers.

Raj et al [10] finally suggests the security measure to prevent the sensitive data from malicious attacks. He provides an log based security system architecture to prevent the data from the attackers.

David et al [11] conclude his comparative study of design approaches from security perspective, none of the approaches will be start to secure a data warehouse design from requirement level to final implementation.

Jorge Bernardino et al [12] he was concluded that the current encryption solutions are not suitable for DWHs. Data warehouses function in a well-determined specific environment with tight security, performances and scalability requirements and, therefore, need specific solutions able to cope with these directives.

Carlos Blanco et al [13] provide a MDA approach supports a secure development of DWs. This architecture initially only supported a rational path towards DBMS, but since the majority of DWs are managed over a multidimensional approach, it was recently improved with a multidimensional path towards OLAP tools. The security rules to the MDA architecture, by modifying our conceptual met model to represent these security rules and defining sets of QVT transformations which automatically generate multidimensional logical models from conceptual model.

G.Thangaraju et al [14] discussed the security and performance in ETL with the transformations operations finally conclude the unconnected lookup transformation is the best for security and fast transformation.

Sajjad Ahmad et al [15] provide a hybrid framework for security issues in data warehouse; the new hybrid technique focuses on identifying the level of users. Identifying of different levels of users introduces the proper classification of users as they are placed to access the data. The access control of the system is made much safe and secured by classifying the users. This division of users also makes sure that the authorized users can only access the data. Unauthorized or Irrelevant users can not access the data that is not concerning to their functions.

M.Sreedhar et al [16] used the Sechmatic Technique using Data type Preserving Encryption to Boost Data Warehouse Security, the first stage of his work was to secure the DWs using the basic data type prevention technique afterwards the enhanced encryption technique is used with DES 56 key, also discussed with the other preventive mechanism for DWs security, the following measures are used to boost DWs security i. prevention ii. Detection iii. Containment, Recovery, Investigation, blinder key methods also used.

Colin Puri et al [17] developed a tool for the tracking the data lineage and generate data flows effectively and efficiently on a heterogeneous environment. We are able to capture high-level patterns by clustering trace log entries and discovering temporal relationships between clusters. As, a result the tool is able to accurately discover flow and sequence structures, is robust to noise, and scales to the size of data.

Somchart Fugkeaw et al [18] presented an approach for a framework of collaborative access control for OLAP queries spanning over multi-data warehouse which is based on PKI and PMI, RBAC and multi-agent system. At a core, they are showed how multiple DW security policies are integrated and handled under the XACML policy specification. Finally they provide the implementation of the prototype A-COLD.

Microsoft SQL Server White Paper Released on April 2102 [19] gives the comparative study of security features in SQL Server DWs Concepts with MySQL 5 DWs concepts, in this comparison there are 9 security measure are taken out of 9 only three measures are supported by both SQL Server 2012 and MySQL 5, other than three are supported by SQL Server 2012 Only.

Juelog LI et al provides a data warehouse security in field service level at the application of flight service. By using the method of dimensional modeling to design FSP data warehouse, put original data in the various departments through clean and conversion, then stored in data warehouse. For the DSS of flight operations the data can be retrieved from the FSS with security.
Hallman et al concluded security issues about Data integrity as , Integrity is the most potential volatile part of the success of any database. A-well designed and maintained database can ensure key domain, and referential integrity. Accessibility and network security tools are very important aspects of data management activities. At a core he said the data integrity is maintained by the users with the common knowledge of their domain or system with the working skills of computers.

V. CONCLUSION

In this paper the survey is carried out the security perspectives of DWs with the existing CIA features of different DWH software’s and also some the authors share their concepts against the DWs in Development phase, Design phase, Modeling aspects, Access level and also in different aspects, on the survey we find the draw backs with the Integrity Management, So we propose the new methodology to enhance the Integrity Management in Data warehouse at Attribute and user Access level.

REFERENCES


Author Information

Mr.G.Thangaraju is working as Assistant Professor in the Department of Computer Applications, Thanthai Hans Rooever College, Perambalur, Tamilnadu, India. He has 17 years of experience in teaching. He has published many research articles in the National / International Conferences and journals. He is currently pursuing doctor of programme in Computer Science at Karpagam Academy of Higher Education, Coimbatore, Tamilnadu, India. His current area of research interests Data warehousing, Distributed Data Base and software metrics.

Dr.X.Agnes Kala Rani is working as Professor in the Department of Computer Applications at Karpagam Academy of Higher Education, Coimbatore. She received her Ph.D. in 2011 from Mother Teresa University. She holds the Master of Engineering degree from VMKV University and Master of Computer Applications degree from Madras University. She has several publications including scientific journals and top-tier networking conferences to her credit.

Mr.G.Thangaraju is working as Assistant Professor in the Department of Computer Applications, Thanthai Hans Rooever College, Perambalur, Tamilnadu, India. He has 17 years of experience in teaching. He has published many research articles in the National / International Conferences and journals. He is currently pursuing doctor of programme in Computer Science at Karpagam Academy of Higher Education, Coimbatore, Tamilnadu, India. His current area of research interests Data warehousing, Distributed Data Base and software metrics.

Dr.X.Agnes Kala Rani is working as Professor in the Department of Computer Applications at Karpagam Academy of Higher Education, Coimbatore. She received her Ph.D. in 2011 from Mother Teresa University. She holds the Master of Engineering degree from VMKV University and Master of Computer Applications degree from Madras University. She has several publications including scientific journals and top-tier networking conferences to her credit.