American Journal of Engineering Research (AJER)
e-ISSN : 2320-0847 p-ISSN : 2320-0936
Volume-03, Issue-06, pp-280-284
www.ajer.org

Research Paper Open Access

Framework for The Integrated And Validated Model of Data Warehouse

Poornima Sharma Nitin Anand
Asstt professor, CSE Department, Sri Venkateswara Engineering College,DCRUST,Murthal, Haryana, India
CSE Department, A.I.A.C.T&R, Geeta colony, Delhi, India

Abstract: **-** A Data warehouse has its own set of challenges for security. Organizational data warehouse are often very large systems. Once we have extracted and loaded the data from heterogeneous sources we need to integrated the data as it seems to similar but coming from the different sources. Validation is required at all the design phases of the development of the data warehouse. With the rapid development and implementation of the data warehouse, security problems arise in populating a warehouse with the enterprise data. For the periods many researchers work on the designing phases of the data warehouse but no research gathers the causes of the security issues. There are some malicious attempts and the vulnerabilities occurring at the levels of abstraction. In this paper we identified the malicious attempts and the vulnerabilities at validation phase through conceptual level modeling and propose the framework of the integrated and validated model of data warehouse.

Keywords**: -** *Data warehouse, designing,modeling approaches, proposed model,security issues, validation*

I. INTRODUCTION

Data warehouse is integrated databases that is designed, organized and honed for retrieval and analysis of data use to support management decision making process [1]. Data warehouse contains the wide variety of data that an organization personnel can use to gain a better understanding of their conditions. A Data warehouse comprises the data that is subject oriented, integrated, time variant and non volatile.Data warehouse development Process conducted in an iterative form after the initialization of business requirements. Then it specifies the cyclical planning , Design, construction, testing and validation and then implementation [2]. The basic view of the development of data warehouse is shown in fig:1. In this paper, we are focusing on the validation process of the development cycle of data warehouse. Validation process helps to reconcile and validate the information before it enter for the implementation. It provides the resistance for the intruded information having some vulnerabilities from the external sources.

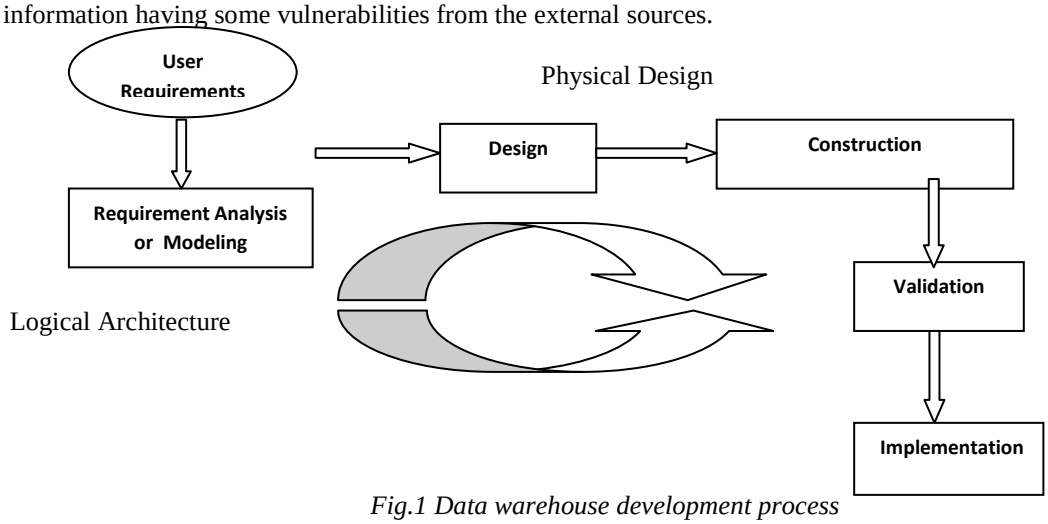

Fig.1 Data warehouse development process

Also there are some existing malicious acts at the same process of development that reduces the functionality and growth of the data warehouse environment [3]. So there is a need of security at the desired phase. To secure the personnel information, primary requirement i.e. CIA, business information, decision making information, usage, decisions and technical information. Some of the techniques at the various design faces can be use to provide the security which is mandatory for the development of the data a warehouse.
Section 2 focuses on the existing approaches on securing the data warehouse. Section 3 explains the dimensional modeling of data warehouse. Section 4 explains multi-dimensional modeling concept with the basic approaches. Further section 5 specifies the security issues. Section 6 explains the framework of the integrated and validated model of data warehouse. Section 7 has conclusion.

II. EXISTING APPROACHES ON SECURING THE DATA WAREHOUSE

This is done with the literature survey of the modeling approaches use to design the secure data warehouse. E.Soler faces on the requirement analysis for the data warehouse which includes the security requirements of the data warehouse by using the MDA approach [4]. A frame work for the requirement analysis upto the conceptual design face is proposed by Ariham Sarkar [5]. Here, the complexity arise in the development of the data warehouse as it is not supported the security of the data warehouse. The access control and audit model explains the security rules and authorization rules at the conceptual level only [6].
In automated data validation and data migration security the validation check of the existing data maintains the integrity of the data [3]. This validation can be done by the mapping and transformation methods. Validation is must at all the design phases upto the physical levels.

2.1 Dimensional Modeling

Dimensional modeling is a design concept used by many data warehouse designers to build their data warehouse [7].All data is contains in two types of table i.e fact table and dimension table.

Table	***Description***
1.Fact Table	• Stores the measures of the business and point to the key value. • Collection of related data items. • Consisting of measure s for the process of decision making.
2.Dimension Table	• Categories each item in a data set into non over-lapping region. • Contains master data with detailed information in a structure.

Dimensional Modeling is a model of tables and relations optimizing for decision support system also constituted to

1. Remove redundancy.
2. Facilitate retrieval of individual records.
3. Optimize OLTP.

2.2 Multidimensional Modeling

Multidimensional modeling is an integrated aspect of OLAP. It involves the analysis of selected facts or measures of the business area. Multidimensional modeling is a prominent factor in interactive analysis of large amount of data for decision making purpose. Basically multidimensional modeling is the foundation of the data warehouses [8]. In this section we are providing the brief reference of the most relevant models done before by the authors.

The dimensional fact model by Golfarelli et al.[9]. The multidimensional conceptual model by Enrico Franconi et al.[10]. The starER model by Tryfona et al.[11]. The model proposed by Abello et al.[12]. These approaches for multidimensional modeling considers security as an important issue but do not solve the problem of security at all stages of data warehouse development.

III. SPECIFIC SECURITY ISSUES

3.1 Malicious attempts

In the data ware house these attempts gains unauthorized controls of some ones computer [13]. These activities includes the personification of the unauthorized user which gains the access by :

a) Spoofing
b) Scanning
c) Masquerade
d) Snooping
e) Impersonalization
f) Tunneling
g) Scavenging
h) Denial of service(DOS)
i) Distributed Denial of Service(DDOS)
j) Password Cracking
k) SourceRouting.

3.2 Vulnerabilities

It is a junction of three classes a system susceptibility, attacker access, attacker capability to exploit the flaw. The vulnerabilities which spoil the behavior of data warehouse are as follows [18]:

a) Dual security engines:- Generates the complexities of security administration in data warehouse environment.
b) Interference attacks:- Posing out direct and indirect attacks.
c) Availability factors:- Deals with the confidentiality and integrity of data warehouse.
d) Human Factors :- Include the accidental and intentional acts.
e) Insider threats:- advisory who operates inside the trusted computing base, basically a trusted adversary.
f) Outsider threats:- outsider parties poses as unethical insiders.

IV. FRAMEWORK FOR THE INTEGRATED AND VALIDATED MODELFOR DATA WAREHOUSE

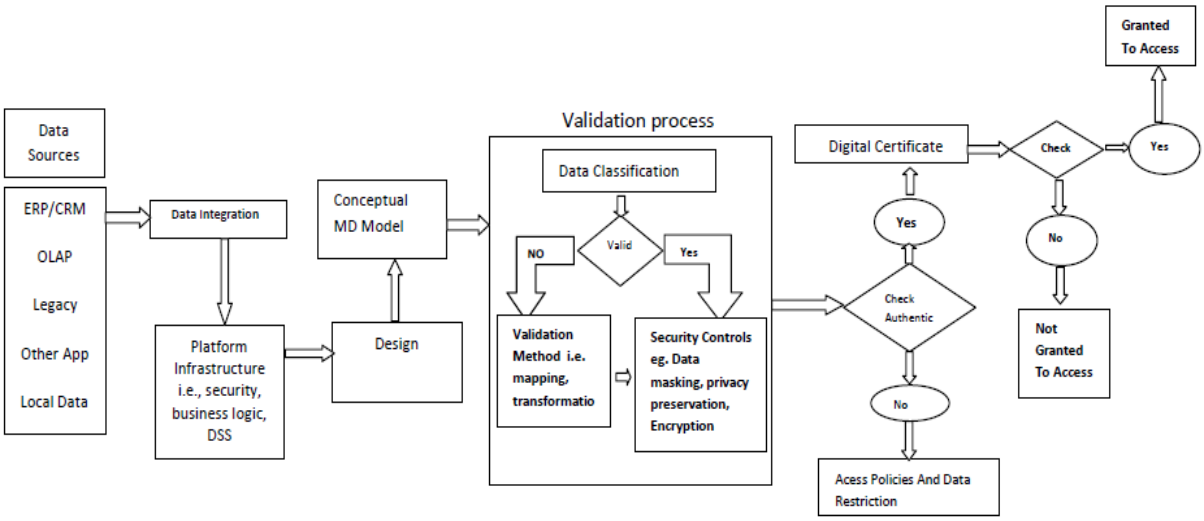

Fig: 2 Proposed model

4.1 Elaboration of the proposed model:

In this approach, the data integration is done as the data comes from the heterogeneous sources(ERP, OLAP, legacy, other applications, local data). To run the environment many issues comes into the picture like schema integration, redundancy, inconsistencies [14]. To maintain the integration of the system some of the tools are running in scenario naming as data scrubbing and data auditing tools. After the integration of the data, a platform infrastructure is built which includes the security, business logic and the decision support system. It is done at the requirement analysis phase. It manages the archieve data of different formats i.e., structured and unstructured [15]. To meet the requirement a standard platform is built in a particular behavior of the data warehouse environment. To maintain the consistency of the platform infrastructure, a regular format is fixed in a manner.

Data warehouse needs to be designed at the next level of the development process. It is the conceptual level designing. Many authors proposes there design at the conceptual level. In the conceptual muldimensional model, the specific OLAP applications has been discusses with the summarized features [16]. Modeling the requirements is the prominent factor in designing the model.

When the designing of the modeling is done, there is a need of validating the process. Validation of the process is required at each level of abstraction(conceptual, logical, physical). Validation helps to reconcile and validate the information before it enter for the implementation[3]. It provides the resistance for the intruded information having some vulnerabilities from the external sources. For validation, first we need the classification of the data to satisfy the security requirement i.e., CIA [17]. The classification of the data is catagories as *Publicly* which is the least sensitive data and can access by the end users, *Confidential* which is moderately sensitive data and those users can access this data who are in need to run their work, *Top-Secret* which is the most sensitive data and the limited users can access that data [18]. Validation check needs to be done after the classification of the data. If the validation check is negative then apply the validation methods to go for the same. Existing validation methods are mapping the data and the transformation rules [3]. If the validation check is positive then the *Security Controls* must apply to the system:

4.2 Security Controls:

1. ***Data Masking***:

2. it is the process of protecting the real data from the unconditional theft. Unlike encrypted data, masked information maintains it's usability for activities like software development and testing [19]. Techniques used for masking are Mutation, Generation, Algorithmic, Loading, Customization.

3. ***Privacy preservation***: needs of ensuring that the privacy and confidentiality needs are fulfill and proper level of data details are exposed not exposing all the details [20]. Privacy preservation is helpful to reduce the possibility of identifying sensitive information. By this method user can utilize the essential details not need to see all the background details as in the case of data abstraction.

4. ***Encryption***: it is the conversion of data into a form which is not understood by the unauthorized person.

After the validation process implementation check the authenticity, if the result is negative then apply the access policies and data restriction. By the access policies the protection of the data is done with the corresponding access rules [13]. Auditing rules cal also be use to do the same and by this the trust can be generated. And if the result is positive then we go for the digital certification of the user. *Digital certificates* are the digital file by which the identity of the user can be verified [21]. If the evaluation through the digital certificates are true then the user is granted to access the data and if it is false then the user is not able to access the same.

The basis of the data warehouse security is to understand the nature and value of the data. Proper classification is also required for implementing the access policies. The primary thing considering is the classification of data(ontology) that supports the other parameters. The classification includes the metadata that indicates semantic classification parameters. The access policies and restriction must be defined based on data to the user roles. Data integration and validation needs to run in parallel with all level of abstraction in the designing phase of data warehouse.

V. CONCLUSION

Data integration and validation is required at the different levels of abstraction throughout the whole implementation process. The proposed framework will fit for any kind of data development process and it works on the refinement process. The benefit of this model is to reduce the risk of security failures at all the stages of data warehouse development.

REFERENCES

[1] Ralph Kimball, (2004). "The Data Warehouse ETL Toolkit", Wiley India (P) ltd.

[2] Boehnlein, M., Ulbrich vom Ende. A., (2000). " Business Process Oriented Development of Data Warehouse Structures", In: Proc.of Data warehousing 2000, Physica Verlag.

[3] Manjunath T.N., Ravindra S. Hegadi, Mohan H S., (2011). " Automated Data Validation For Data Migration Security", In: International journals of Computer Applications(0975-8887), volume 30, No.6, September.

[4] Emilio Soler, Juan Trujillo, Fernandez-Medina, (2008), "Towards comprehensive requirement analysis for DW: Considering security requirement", Published in IEEE Conference.

[5] Ariban Sarkar, (2012), " Data warehouse requirement analysis framework business-object based approach", In: IJACSA, vol.3, No.1.

[6] Median, E.F., Trujillo,J., Villarroel,R., Piattini,M., (2006), "Acess control and audit model for the multidimensional modeling of Data warehouse", Decision Support Systems 42,1270-1289.

[7] Joseph M., (1998), " Dimensional Modeling and E-R Modeling In The Data Warehouse", In: White Paper No. Eight, June 22.

[8] Umashanker Sharma, Anjana Gosain, (2011), " Dimensional modeling for Data Warehouse", In: Proc. ISCET, Database.

[9] M. Golfarelli., (2009), " From User Requirements to Conceptual Design in DataWarehouse Design – a Survey", In: International Journal of Data Warehousing and Mining.

[10] Enrico Franconi, Ulrike Sattler, (1999), " A DataWarehouse Conceptual Data Model for Multidimensional Aggregation". In: Proceedings of the International Workshop on Design and Management of DataWarehouses (DMDW'99), Heidelberg, Germany, 14. - 15.6.

[11] Nectaria Tryfona, Frank Busborg, Jens G. Borch Christiansen, (1999), " starER: a conceptual model for data warehouse design", In: DOLAP '99 Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP, ACM Digital library, New York, USA.

[12] Alberto Abelló, José Samos, and Fèlix Saltor, (2006), " YAM²: A Multidimensional Conceptual Model Extending UML", (© Elsevier). In Information Systems 31 (6), September, 2006. Pages 541-567. Elsevier. ISSN 0306-4379.

[13] Available at : http://www.comptechdoc.org/independent/security/terms/attack.html.
[14] Kalinka Mihaylova Kaloyanova, (2005), " Improving Data Integration For Data Warehopuse: A Data Mining Approach". Available In: http://www.nbu.bg.
[15] Bill Inmon, (2010), " Manage Data Growth And Optimize The Data Warehouse Infrastructure And Data Warehouse Archiving", In: Informatica, White paper.
[16] Riccardo Torlon, (2003), " Conceptual Multidimensional Models", In: ACM digital Library, IGI Publishing Hershey, PA, USA.
[17] Kimmo Palletvuori, (2007), " Security of Data Warehousing Server. In: White papers, Seminar of Network Security".
[18] Slemo warigon, (1997), " Data warehouse control and security", In: LEDGER, Vol. 41, No. 2, April ; pp. 3-7.
[19] Ricardo Jorge Santos, (2011), " A data masking technique for data warehouses", In: Proceedings of the 15th Symposium on International Database Engineering & Applications, ACM Digital library, New York, USA.
[20] Raymond Chi-Wing Wong, Ada Wai-Chee Fu, (2010), " Privacy-Preserving Data Publishing: An Overview". In: Synthetic lectures on data management, Morgan ang Claypool publishers.
[21] Dr. Stefan Brands, (2004), "Non-Intrusive Identity Management, Available at http://www.credentica.com/technology/overview.pdf, published in 2004.