DATA WAREHOUSES SECURITY IMPLEMENTATION

Burtescu Emil
University of Pitesti emil.burtescu@yahoo.com

Data warehouses were initially implemented and developed by the big firms and they were used for working out the managerial problems and for making decisions. Later on, because of the economic tendencies and of the technological progress, the data warehouses could also be set up by the medium and small companies. Setting up a data warehouse is a difficult and laborious operation and its security is even more difficult to achieve. The exaggerations concerning the security measures are not always beneficent to business. This is why the security measures projection for a data warehouse must be realized very carefully. The literature contains guides and lots of advice regarding data warehouse security. A substantial support can be offered by the specialized firms that are part of the present-day market.

Keywords: data warehouse, security, security architecture, performance, productivity.

JEL code: M15

Data warehouse

Data warehouse represents a collection of databases that come from different sources; they are integrated in the organization and they are used for working out certain problems and making decisions. Data warehouses contain a great variety of information which offers a coherent image of the company's economic conditions at a certain moment. Data warehouse refers to a combination of more different databases corresponding to the entire organization. This type of system is used for the supervision of the organization activity and supplies standard reports about it. It is also used for analyzing certain aspects regarding the organization's activity and some managerial decisions can be made on this analysis basis.

The data warehouse initiation was possible owing to the economic tendencies and to the technological progress. Nowadays, the new economic tendencies involve: trade globalization, the intensification of the competition, the spectacular shortening of the duration of products because of the technological evolution and also the imposing of high quality standards. The technological progress has allowed the increase of the calculation capacity and the prices reduction. We should also take into account the operating systems, the free-access to the data, the need for information that is acute, the storage capacity which is great and cheap and the software implements that are accessible. Considering all these, we can say that it's time to set up a data warehouse.

The main characteristics of a data warehouse:

- It is not focused on daily operations and transactions. Its focus is on the data that are used for analysis and making decisions.

- The main operation is data adding. Data are neither cleared nor overwritten. A data archive is maintained.

- The system is integrated. Data are collected from different sources (operating systems, databases, files etc.) and by the data warehouse they are transformed into standard reports of representation and they are centralized in the system. **Data integration represents the most important problem when setting up a data warehouse.**

Data warehouse costs

In most cases, the costs for setting up a data warehouse can be prohibitive and because of this aspect, many firms can't afford it. If we also consider the subsequent costs, then its image will seem apocalyptic for a medium company. On the other hand, the one who wants security for his business will allocate financial resources for the setting up and for the maintenance of a data warehouse. A data warehouse is composed of databases that contain from 1 to 10 Terabytes, in quite many cases values can reach tens Terabytes. The setting up of a data warehouse costs about 3 million dollars. The costs for its initiation and maintenance are divided into costs necessary to **hardware systems** (equipments and high storage media), **professional services** and **software for the data extraction, processing, storing and analysis.**

Data warehouse structure

Concerning its structure, a data warehouse has the following levels: **data source**, **transformation**, **data warehouse**, **reporting**, **metadata** and **operations** (Figure 1).
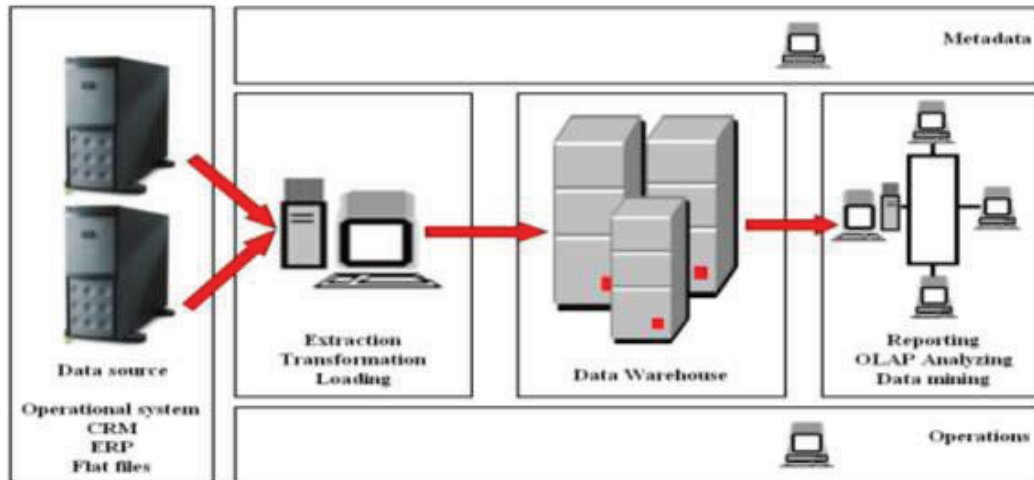
Figure 1. Data warehouse structure

Data source. Data source level refers to any information, usually in electronic format, that must be collected and stored in the data warehouse. Sources: mainframe databases (DB2, IBMS, VSAM, ISAM, Ababas etc.); client-server databases (Oracle, Informix); databases coming from personal computers (MS Access), spreadsheets and other types of files.

Transformation. This level deals with the standard transformation of the data collected at the Data Source Level. Data transformation consists in bringing them to the standard format of the data warehouse. To perform this level, one can either write special applications or use ETL implements (Extract, Transform and Load).

Data transformation may also involve data "cleaning" operations (correctness, consistence, totals, subtotals etc.).

Data warehouse. It deals in fact with data storing. For storing one can use relational databases. Data are stored for a fixed period.

Reporting. This level does the analysis and generates reports used for the supervision of the company activity. For this purpose one can use specific implementations:

- Business Intelligence Tools - applications that simplify the generating reports process from the data warehouse;

- Executive Information Systems - software applications created to generate complex reports for the company top level management;

- Enterprise Information Systems, similar to Executive Information Systems, the difference is that the first is more simplified for the company departments.

- OLAP (On Line Analytical Processing) allows multi-dimensional representations (cubes, hypercubes) and also allows fast and interactive analyses basing on different types of operations such as roll up, drill down, drill up, slice, dice etc., Data Mining, KDD (Knowledge Discovery in Databases), that use statistics analysis techniques and "forms" recognition in order to find correlations, rules, knowledge etc.

Metadata. This level includes data warehouse   management information (the latest update, the number of the users connected to the system etc.).

Operations. The main purposes of this level are data loading in the data warehouse and also data processing and extraction. Another purpose is represented by the users management, the security, the capacity and other management functions.

Data warehouse security

Securing a data warehouse is an operation that requires much care and the concentration of the efforts because there are a few particular aspects that may appear.

The first aspect refers to a direct approaching of the security because in a data warehouse the information is stored in one location. This aspect involves multiple risks for the company. Focusing on the data warehouse is enough for an intruder and he doesn't need to waiste his time to attack the operating system because he can extract any data he wants from the data warehouse.

The second aspect refers to the fact that the access to data warehouse must be facile. The implementation of the security solutions may lead to the reduction of the data warehouse performance. According to recent studies, the period of interrogation has increased from 20% to 500% because of the security measures. The point from which the use of a data warehouse proves to be unprofitable must be established.

The experts agree with the fact that **a standard method for setting up a data warehouse doesn't exist**. We can conclude that **a standard method for a data warehouse security doesn't exist either.**

However, a few principles should be taken into consideration:

- **Data warehouse team must contain a "security architect".**

- **The specific security politics for departments and/or users must be defined.**
- **One security politics for data that belong to the same category.**
Security architect must establish the following aspects concerning security:
- **Security roles.** Data that can be accessed and modified and the type of users that can do this must be defined.
- **Security technology.** The following aspects are established: how is the access path from the user to the data warehouse server protected against address spoofing, masquerading, eavesdropping and any other threat as replay attack.
- **Security administration.** The one who administrates passwords and security roles must be established. It is also established the security mode in order to assure the copies security. The employees are communicated the new security implemented measures.
In order to ensure the security in any domain, technology and the mode of its implementation have the most important role. Security architecture for a company is exemplified in the figure below (Figure 2):
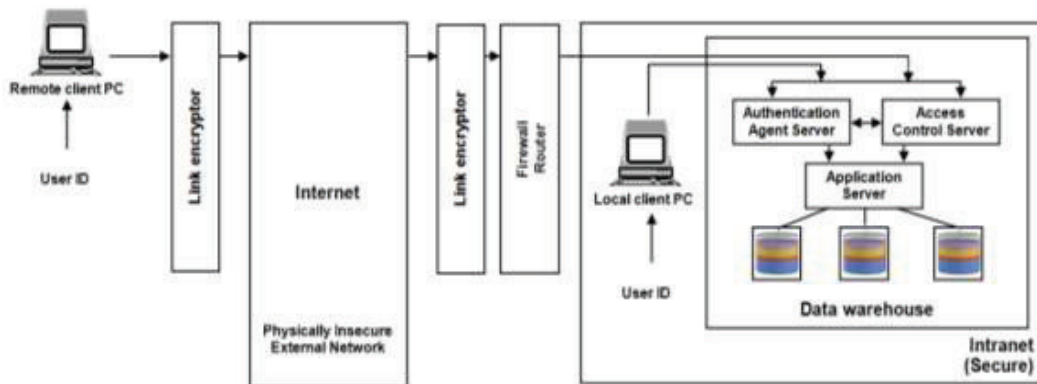

Figure 2. Security architecture for enterprise data warehouse

Another aspect that must be considered refers to the implementation security mode: whether it's done for the application which has access to the data warehouse or not. In the first case the security is integrated in the application and is specific to the data accessed by it and to the functions that operate with the respective data.
The second case is when 5 points are accumulated after answering the following questions from the table (Table 1

Table 1. Security implementation table

Questions	Answer (0-No, 1-Yes)
There are more applications that access the data warehouse?	0/1
Is the data warehouse approachable by standard methods such as ODBC?	0/1
There are more than 100 users?	0/1
There are more than 4 dimensions?	0/1
There are more than 4 hierarchies in the structure of the same dimension?	0/1
Sum	0….5

If five points are accumulated then an internal security may be applied. In this case the security addresses to all applications and has a single checking point.
The literature are indicated seven stages for the implementation of a secured data warehouse:
1. **Data Identification.** The identification of all the data that are stored by the company. A total data is realized. It is absolutely necessary to take a rigorous inventory of all data to be available to users of the data store. Data warehouse has a component of its minitoring software that can provide information about data.
2. **Data Classification.** The classification is done considering the security. For example: Strict Secret Data, Confidential Data, Special Data, Public Data. Operation of the classification of data into different categories of data an operation is not easy sometimes. In quite many cases the data are a combination of several categories.
3. **Quantifyind the Value of Data.** It involves the evaluation of the costs caused by the data loss. Depending on the type of the incident, we can estimate the replacement (substitution) cost – the cost for the replacement of the equipment that is unavailable or destroyed, or we can estimate the recovery cost – the costs necessary to the recovery of the information that has been lost. Companies use both quantitative analysis and qualitative analysis to calculate the Annulized Loss Expectancy (ALE).
4. **Vulnerabilities Identification.** It represents the identification and the documentation of the vulnerabilities. Among the types of attacks we can mention: **Inference Attacks** – involves the deduction of confidential data using

information which apparently seems public. Attacks specific to data warehouse; **Internal Attacks** – these attacks have great success because the intruders may   have certain rights and knowledge regarding security architecture; **External Attacks** – Hackers, crackers, industrial spying; **Natural Factors**/**defects** – earthquakes, flooding, interruptions and variations of the current etc. In most cases data warehouse bases on the security based on views (VIEW based). Dump data can avoid this method. Another types is: In-built DBMS security, DBMS limitations, dual security engines, availability factor, human factors.

5. **The identification of the protection means against the vulnerabilities and the costs evaluation.** It represents the identification of the protection methods against the vulnerabilities. We must identify the vulnerabilities. As principal methods we mention: Users classification and the access control implementation, Data partition, Data cryptography, control of the software development process, the executive communication, informing and training. All these have one purpose: to ensure the security and to prevent the clearing and the modification of the information from the data warehouse. Finally, the costs necessary to the security measures implementation are estimated.

6. **The selection of efficient protection measures that are to be implemented.** It involves the selection of measures that represent the security implementation, generally speaking, basing on the implementation costs. In quite many cases the criterion of cost should not be the main criterion for selection. May be taken account of the compatibility and adaptability to new controls adopted by the old. It is well known that can achieve substantial cost reductions if they choose a control to cover more vulnerabilities not control for each vulnerability.

7. **The evaluation of the implementation efficiency.** It involves a continuous verification and auditing of the security solutions that have been implemented. A communications executive to data warehouse users of new security measures will be adopted as to increase their effectiveness and productivity of the initial losses to reduce.

Before creating a data warehouse, we must know firstly very well if the data warehouse is absolutely necessary to the company. If we need a data warehouse then we must select the data we want to be stored and processed. The system will generate reports containing results that help me to make the proper managerial decisions. Storing "all that can be stored" involves supplementary costs and may lead to the reduction of the performance. On the other hand, if the stored data is insufficient, the results may be uncertain.

From the data which are supplied by the operating system, I select the data which are essential for making decisions. We have to take into account the possibility of a posterior development – we may need the data which were omitted at the beginning and later on they prove to be necessary to making decisions.

Although the initial implementation plan is well defined, during the implementation it may be modified because of the difficulties that may appear. Another reason of this decision is caused by the fact that the company which hasn't experts in security implementation wants the implementation of a paranoid security system. You should try to create a product for a certain company without referring to security. At the end, if you bring in to dicussion the security you will notice that all the attention and the efforts will be focused on security, although this can lead to the reduction of the performance application.

Conclusions

Data warehouse is a system created in order to offer an immediate access to important and reliable data which are used for decisions analysis, planning and evaluation. A data warehouse can facilitate very much the work but at the same time it can make it more difficult. An under-evaluation of the threats leads to an under-protection level for data warehouse security and business and creates possibilities of access for the intruders. An overestimation of the threats, sometimes reaching paranoid values, leads to an overprotection level that ensures an excessive security but will be as harmful as the under-protection level. In this case, the security means can be annoying and unbeneficial to the company production. It would be ideal to find out an optimal way for a data warehouse security. In quite many cases, because of the lack of funds, the companies can adopt a minimal measures level which will not be enough for attaining our supposed purpose.

REFFERENCE:
1. L. McCarthy, IT Security: Risking the Corporation, Prentice Hall PTR, 2003.
2. Security Complete, Second Edition, SYBEX Inc., 2002.
3. http://all.net/books/audit/kits/dw.html
4. http://www.information-management.com/infodirect/19990601/780-1.html
5. http://www.dbmsmag.com/9704d05.html