# The Importance of Data Warehouses in the Development of Computerized Decision Support Solutions. A Comparison between Data Warehouses and Data Marts

Alexandru Adrian ȚOLE
Romanian – American University, Bucharest, Romania
alexandru.tole@gmail.com

*In the last decade, the amount of data that an organization processes and stores has grown exponentially. In most cases, the data stored is used to support the business process through accurate and up-to-date information about the business environment and activity of the company. In order for a company's managers to be capable of generating the reports they need to make decisions, one needs a computer system able to store complex and very large quantities of data. At the same time, for the development of such an information system, one must take into account the cost of it.*
***Keywords:*** *Data Warehouse, Data Mart, Top-down, Bottom-up, database, architecture, management system*

## 1 Introduction

Data Warehouses (DW) by means of its data organization and usage elements, are meant to provide helpful tools to assist managers in the decision-making process. Many companies deem the implementation of DW tools as most useful because the instruments that these systems offer help with the maintenance and development of the economic unit. Companies invest large sums in the development and implementation of such solutions as they provide essential elements that companies can use to retain customers and, at the same time, increase their number.

Data Warehouses were originally intended for use in areas such as banking, telecommunications and retailing. However, later, they became hugely popular in other areas as well. According to the literature, it is the banking field and the telecommunications domain that uses these solutions most frequently, allocating the largest amounts for the development of data warehouses.

Inmon estimates the costs for the development of a solution of the Data Warehouse type as being between 500,000 $ and $ 1,000,000 per terabyte [1] and this includes only the infrastructure. From this perspective, the companies that do not have such financial resources may resort to the implementation of Data Marts (DM), whose costs are much lower.

As far as the definition is concerned, the data warehouse term can be presented as "a database that is maintained separately from the operational databases of the organization" [2]. By this, we mean that the information is extracted from operational databases, goes through an ETL type process, following to be loaded into the Data Warehouse. These data warehouses are used to support the decision-making process by storing information in the decision-making process.

As the main features of data bases, we distinguish the following [3]:

- Integration;
- Data persistence;
- Historical character;
- Guidance on topics.

The "integration" characteristic of data warehouses refers to the fact that the source of the information stored is not singular. The data existing in a DW originate from sources such as relational databases, files (e.g. excel, word, xml), etc. This information, before being loaded into the DW, goes through

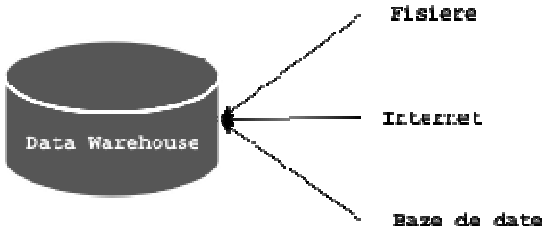
processing to be used later in the analysis performed using other Informatics solutions.

Another feature of Data Warehouses is that they focus on topics that are specific for the activity of the company, removing information that is not relevant for the decision-making process. Traditional systems focused more on the particular data requested by certain departments (compartments) of the economic unit. Over time, these systems have been developed to support the implementation of processes from the beginning to the end.

The historical character of Data Warehouses is that they store information for an undetermined period of time. This helps in the achievement of the decision-making process because the decision-makers can make statistical calculations for certain key indicators.

The information stored in a data warehouse is permanent and cannot be changed. Therefore, any changes made in the DW data sources will have new data as a result, without modifying or deleting the existing data. It follows that a data warehouse is stored separately from the data processed by other applications. Normally, the operations in a data warehouse are the loading of data and the access to this information.

Another feature of a DW is that related to the origin of the data. It can have simultaneously, internal and external information sources (Fig. 2 - Sources of information) to help the decision-making process. At the same time, DW includes tools with which users can quickly access data.




(Files, Databases)

**Fig. 1** Sources of information

The notion of "data warehousing" is very often used in the making of a data warehouse. The development of a data warehouse involves processes such as data integration, cleansing and consolidating them. In order to use a data warehouse one needs most often various applications that help with the interpretation of this information. These applications allow the decision-making structures to use information in a convenient and easy way for the decisions-making process. As far as the "data warehousing" term is concerned, this is interpreted as the process of the making of the data warehouse. For the use of the DW it is the *Warehouse Database Management System* [3] term that is used..

The process of data warehousing is useful because it helps the integration of heterogeneous data sources. It is well-known that large companies collect various types of information and load them into databases for later use in the development of statistics and/or calculation of trends. The integration of information, as well as an as easy as possible access to them constitute a priority for companies.

With regard to traditional databases, the integration of heterogeneous databases is carried out by means of two elements: wrappers and integrators. As for example, when performing a site query, it is a dictionary of metadata that is used to transform this search into a query appropriate for the sites involved. The result received from various sources is integrated in the global response of the query performed.

The concept of data query will result in a complex process of integration and filtering that is reflected on the processing resources. Thus, it is ineffective and can become expensive, in terms of hardware architecture, to achieve frequent queries, especially if they require aggregations. Data warehousing provides an alternative to the traditional concept regarding the integration of databases via the "update-driven" [4]

characteristic, through which the data from sources are queried and stored in the data warehouse to be analyzed later. For example, in contrast to the databases that store information on accessing the email by Yahoo users, a data warehouse does not present information updated in real time.

Although a data warehouse has the disadvantage of supplying recent data, it provides a high performance by integrating databases whereas data is copied, processed, summarized and restructured into a semantic data collection [5]. At the same time, the process of data warehouse query does not interfere with processes from local sources. A data warehouse is also advantageous due to the fact that it stores and integrates historical data and provides facilities regarding advanced query features.

Data warehouse provides access to the integrated data of the company, which had been previously blocked or restricted. Users can establish a secure connection to the data warehouse via a PC. Security is provided on the one hand by means of the user interface (UI) and on the other hand through the database server.

The data warehouse has the quality of providing a single version of the truth. The information available in the DW have a high quality due to the transformation process they have gone through. This is so if a single data source is used, thus putting an end to the debate on the veracity of the data. The data warehouse comes to be a unique source of information for the top-management. It should be noted that the existence of a single variant of truth is dependent upon firm agreement on the terms used. For example, the term "possible client" can have multiple meanings, such as former clients, customers whose contracts are about to expire, the entities addressed offering deals etc. Surely these issues influence the decision-making process.

The fact that a data warehouse integrates historical data determines the relevance of decisions taken by managers because, in general, decisions are based on previous experiences as well. For example, comparing monthly reports leads to the making of decisions regarding investments. Data warehouses can be used to accurately record the past, making the OLTP systems available to focus on the correct registration of current transactions. Historical data are loaded and integrated with other existing data in the DW to provide rapid access.

The types of dynamic reports as well as the OLAP query tools allow users to visualize the data from DW from different perspectives and at different levels of detail. These possibilities offered by Data Warehouses reduce time and effort concerning the collection, processing and filtering of information coming from various sources.

The operational processes are vastly different as compared to the decision-making processes. The attempt to integrate the decision-making information with the operational information makes system maintenance became a major problem. Thus, starting from the operational processes, the Data Warehouse provides a separate architecture in relation to the implementation of decisions.

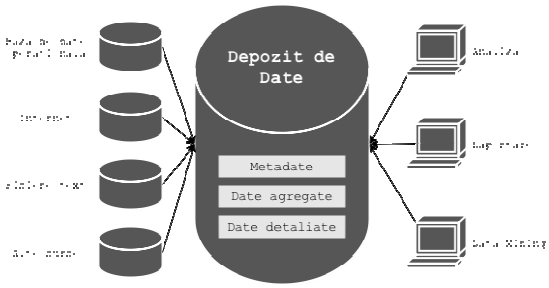
## 2 Databases architecture

A data warehouse consists of a very large database, which contains data that can be used by end-users. In a Data Warehouse there are several types of information that correspond to the users ' needs, such as the following:

- Detailed data;
- Aggregate data;
- Metadata.

Metadata is information embedded in the data warehouse containing data relating to the content stored. Metadata includes information showing the structure of the existing data in the data warehouse, their origin, transformation rules, aggregation and calculation. They have a very important role

in powering the data warehouse, being used in all the processes of loading of data and are updated throughout the life of the data warehouse.
The existence of aggregate data in the data warehouse increases data redundancy, but it is required, however, to reduce response times regarding Data Warehouse queries. A logical flowchart regarding the basic architecture of a data warehouse can be seen below (Fig. 3 – Basic architecture of a DW).




(Operational databases, Text files, Other sources – Data Warehouse – Metadata, Aggregated data, Detailed data – Analysis, Reporting)

**Fig. 2** Basic architecture of a DW

Data warehouses are generally intended for use by analysts or persons engaged in decision-making processes concerning the development of the economic entity. To achieve this, they need powerful tools that facilitate access to and use of the information stored in data warehouses. These tools are mostly provided by the data warehouse. At the same time, a DW solution can also integrate tools which meet the need of users for rapid access to information or quickly generate reports. There are specialized tools that can transform the information in the data warehouse so as to be presented in the form of graphics and/or diagrames. Here one can fiind specialized OLAP instruments and data mining.
The OLAP-type instruments focus more on a multidimensional representation of information and allow for rapid analysis of data by means of processes such as drill down, slice, etc. Thus, the user can obtain results rapidly and can work at different levels of synthesis.
Data mining -type tools are helping transform information into knowledge, thus the term Data Mining is often confused with the Knowledge Discovery in Databases term [6].

## 3 Architecture of data warehouses on three levels

From the architecture described above we can build a software solution that is capable of meeting the requirements of all users of the data warehouse. Figure 4 shows a data warehouse architecture on three levels: Bottom Tier, Middle Tier și Top Tier[7].

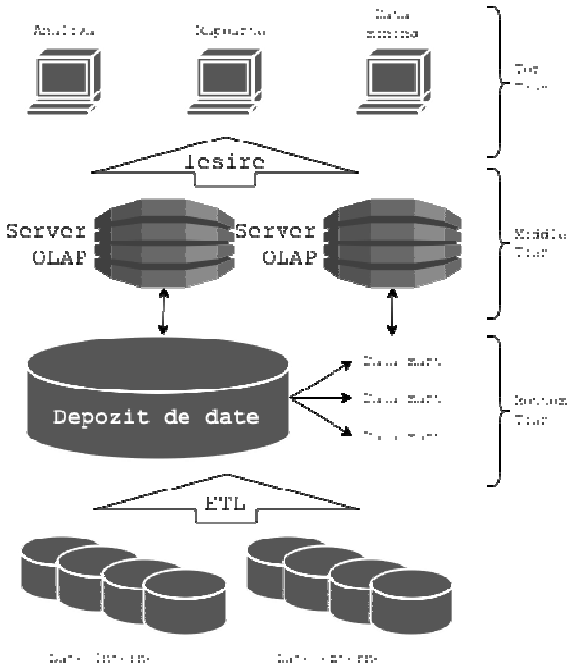



**Fig. 3** Architecture of data warehouses on three levels

(Analysis, Reports, Exit, OLAP Server, Data Warehouse, Internal data, External Data)

a) Bottom tier:

Includes the server of the data warehouse and of the data marts. Typically, the databases used for building the data warehouse are relational databases. The data warehouse is loaded with internal information (from the company, from various departments) and external (third party applications used by employees, information collected online, files etc.). The information is loaded into the data

warehouse after they were extracted and processed. At the same time, from the database one obtains the information to be loaded on the data marts to be made available to users.

b) Middle tier:

It is made up of OLAP servers. It uses either a relational model (ROLAP) or a multidimensional one (MOLAP).

c) Top tier:

It is the level at which the user interacts with the computer system. At this level you can generate reports, you can create a data analysis by means of the tools or you can achieve the data mining.

Ralph Kimball and William Inmon had different opinions with regard to the design and architecture of data warehouses. Inmon supported a "Data Mart" structure, dependent on data, a method which is called "top-down". This method describes an approach by means of which the data warehouse is done first, followed by the creation of Data Marts, as satellites that contain data. Kimball has approached the problem differently, in the sense that he started from the development of Data Marts to the realization of the data warehouse. This approach is called "bottom-up".

## 4 Top-Down Method

Inmon has noticed that it takes the transfer of data from various OLTP systems and centralizing them for further analysis. He considers that the data must be organized in structures that are "embedded, non-volatile, subject-oriented and variable in time "[8]. At the same time, he thinks that the data must be accessed in a detailed enough level to allow the use of data mining tools. The Data Marts, in this case, are seen as being data sub-sets of the Data Warehouse. Data Marts are developed for in each department, so as to later meet the analysis requirements of the department for which the data warehouse has been developed .

The top-down method (Fig. 5-Top-Down), in an OLAP environment, starts with data extraction from the operational data sources. Afterwards, they are loaded into the waiting area, where they are validated and consolidated in order to ensure a high level of data quality. They are then transferred to the Operational Data Store (ODS). This stage (ODS) is often omitted in cases where there occurs a doubling of operational databases. At the same time, detailed data are constantly extracted from the operational data warehouse and hosted temporarily in the transfer area, following extraction and upload (ETL) in the data warehouse.

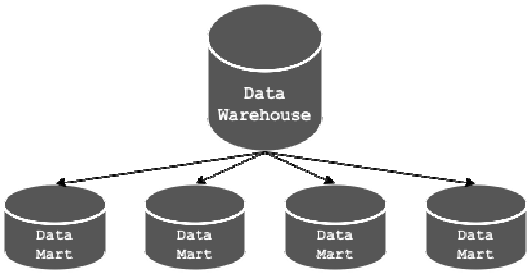



**Fig. 4** Top-Down

The need for an Operational Data Warehouse is given by the needs of the business process. Whenever the situation calls for the existence of detailed data in the data warehouse, then the implementation of an ODS is justified.

Once the data warehouse processes concerning the aggregation and centralization come to an end, the cycle of Data Marts updateis resumed by means of data extraction and upload into the transfer zone in order to subject them to transformations. This helps to structure data. Upon successful completion of these processes, Data Marts can be loaded with data with a view to be available to users who work in an OLAP environment.

## 5 Bottom-Up Method

Ralph Kimball designed data warehouses using Data Marts connected to it by means of a "bus" [9] type architecture (Fig. 6-Bus Architecture). This architecture encompassed all the common elements of a Data Mart. Kimball considered that by

using these elements, users can interrogate all the Data Marts at the same time. This renders the data store more virtual than physical. Thus, all Data Marts can be found on a single server or may be located on different servers, spread across the organization, forming a virtual data warehouse.
Applying the Kimball model, we can consider as a Data Mart the cubes as well, built by using OLAP. This model provides flexibility by means of the rapid realization of Data Marts.

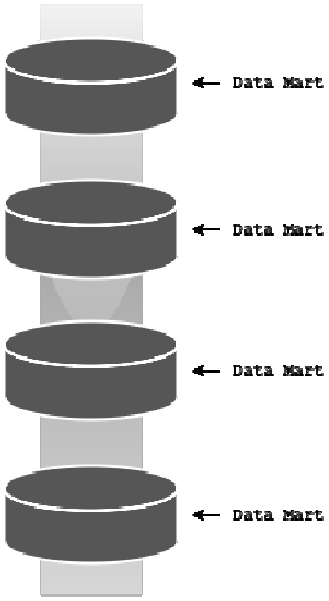


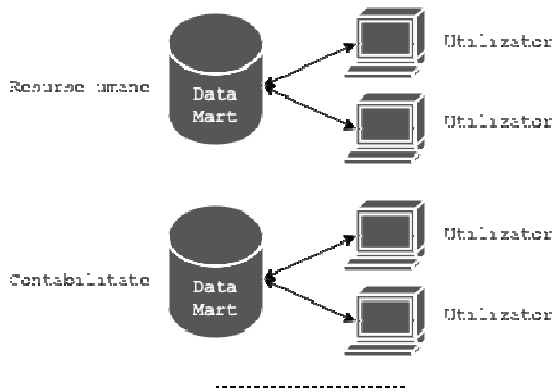

**Fig. 6** Bus Architecture

By means of this model, Data Marts can be carried out more quickly and the structuring of common data, according to this architecture, eliminates the effort made whenever the need to achieve more Data Marts occurs.
Bottom-up approach reverses the hierarchical relationships between Data Warehouse (DW) and Data Mart (DM). DM are loaded directly, through the area of transfer, with the data obtained from different sources. In this case, the existence of ODS is optional, being subject to the company's needs. However, this approach increases the complexity of the process of transformation of the data. The standard procedure by which the DM are updated from the Operational Data Warehouse, not directly from relational databases, provides data consolidation, being the recommended approach for use of the resulting data.
Data stream in the Bottom-Up method begins by the extraction of data from operational databases, their upload them in the transfer area where they are processed and then consolidated and loaded into the Operational Data Warehouse. The information on the ODS is replaced or supplemented by the newly-loaded data. Upon completion of these operations, the data from the ODS are again loaded in the transfer area and processed to fit the structure of the data available in Data Marts. The data uploaded into the DM are transferred into the waiting area to be subjected to processes of aggregation, summary, etc. to be able to be loaded into the DW and made available to the end user to be analyzed.
Therefore, the bottom-up approach starts from the company's need to process certain information and highlights Data Marts as being the primary source of information for making an analysis[10].

## 6 Data Mart

Inmon defines the Data Mart as "a subset of the data warehouse that has been converted to meet the needs of a department" [11]. Therefore, a relevant data mart contains information from a particular area/section. This information will either be used only by the members of the department, or it will help to make the reports at the level of management to assist in the decision-making process.
It is very important that when you develop a computer system having the role of a data warehouse, it must be sufficiently flexible so as to be able to adapt to changes that may take place. The flexibility of the solution is also given by its capacity to connect to all levels of the company and retrieve the necessary data. If new servers are added to compensate for the need of capacity or computing power, processes concerning the

system configuration, optimization, and administration may become difficult if the operations are repeated whenever a hardware architecture component is replaced or added. To overcome this problem you can opt for the development of data subsets called Data Marts (Fig. 7- Data Mart). A data mart is a data warehouse that contains information specific to an organization department or activity, as opposed to a data warehouse that can meet the needs of the entire company. In this context, the resources required to develop a data mart are much fewer than those needed for a data warehouse.




(Human Resources, User, Accounting, User)
**Fig. 7** Data Mart

Consequently, you can use existing information in common, connecting the data marts for each department, in order to achieve an infrastructure based on which one can develop solutions to assist the decision-making process.
Due to the very low cost of data mart development, most companies opt out for developing them. At the same time, the costs relating to the administration are also low and the flexibility offered by them is large enough to help the transition to Data Warehouse if needed.

## 7 Data Warehouse vs. Data Mart

Choosing a solution of the type of Data Warehouse or Data Mart depends, to a large extent, on the needs of the company. If the company aims at processing and aggregating the data in order to achieve a solid system to assist in the decision-making process, then it is recommended that you implement a Data Warehouse solution. At the same time, a Data Warehouse includes historical data, which helps the development of statistics. If the economic entity only needs data storage for each department or if all the departments use the same computerized solution to enter specific data, then a Data Mart will be capable enough to meet such needs.

**Table 1** Data Warehouse and Data Mart Characteristics

| Data Warehouse | Data Mart |
|---|---|
| ▪ **Comprises both internal (the company departments) and external data (internet, other files);**<br>▪ **Stores detailed data;**<br>▪ **Contains historical data istorice (that can span the entire Data Warehouse life-duration);**<br>▪ **It does not necessarily use a dimensional model;**<br>▪ **Contains metadata;**<br>▪ **High implementation costs.** | ▪ Comprises data specific to a department or area of interest;<br>▪ Uses a dimensional model (the*„Star"* model, most of the time*)*;<br>▪ Contains informations that can be transmitted to the data warehouse (if it exists);<br>▪ Low implementation costs and duration. |

As can be seen in the table above, a Data Warehouse contains more types of information than a Data Mart. Moreover, in order to achieve a decision support solution, a DW is more useful because it contains data that have undergone

processing before being loaded. Thus, the quality of information increases, the generated reports showing the actual data. A Data Mart is superior to a Data Warehouse in terms of implementation and maintenance costs. At the same time, a Data Mart can compete fairly easily with a DW with regard to the information support provided for the achievement of decision-making solutions because small and medium-sized firms often use a single software to enter and record the business process information. Usually, small and medium-sized businesses use a software which comprises several modules specific to departments. In this sense, the implementation of an extra module (if the computer solution allows it) with a view to generate reports and graphs is quite easy and the costs are considerably lower than the development of a solution which operates on the basis of a Data Warehouse.

## 8 Conclusions and Proposals

With regard to the development of a company, the choice of appropriate computer solutions represents a very important element. Computer solutions refer to both the hardware infrastructure (servers, network, firewall, etc.) and the systems software (operating systems, databases, specialized software) that will help the business process. In this context, there are several elements that must be taken into account in the development of a computer solution:

- Company needs;
- Financial resources of the company regarding the development of a computer-based solution;
- Complexity and prevalence of information.

When a company wishes to implement a solution that works with information available both within the company and externally, and leading to the implementation of software systems to support the decision-making process, then a Data Warehouse solution is appropriate. The advantage it offers is given by the quality of the stored data (going through a process of ETL), includes historical data (data recorded during the whole DW lifespan), provides long-term stability.

If the economic entity does not work with complex information and/or the information source is compact (only one department or multiple departments working with one soft) then you can deploy a Data Mart. By means of the data existing in such a Data Mart there can be built, as well as in the case of a DW, a software solution to support managers in the decision-making process. At the same time, the costs of implementing a Data Mart solution are lower as compared to those of a Data Warehouse solution.

The advantages that a Data Mart has got as compared to a DW are: low implementation costs and reduced duration of deployment. At the same time, on the basis of the Data Marts existing in a company a Data Warehouse can be develop at all times.

## References

[1] William H. INMON, „*Some Straight Talk About The Costs Of Data Warehousing*”, http://goo.gl/cYGGH6;

[2] *Abordări de tip Data Warehousing - Implementare în Microsoft SQL Server 2005* - http://goo.gl/NPl9x1;

[3] David J. DEWITT, Samuel MADDEN, Michael STONEBRAKER, „*How to Build a High-Performance Data Warehouse*”, http://goo.gl/PT60BZ;

[4] Morgan KAUFMANN, „*Data warehousing*”, https://goo.gl/bY1E8k;

[5] Victoria NEBOT, Rafael BERLANGA, Juan Manuel PÉREZ, María José ARAMBURU, „*Multidimensional Integrated Ontologies: A Framework For Designing Semantic Data Warehouses*”, http://goo.gl/ChCYdb;

[6] Usama FAYYAD, Gregory PIATETSKY-SHAPIRO, Padhraic SMYTH, *„From Data Mining to Knowledge Discovery in Databases”*, http://goo.gl/EEJH4f;

[7] Sandeepak BHANDARI, Tarun SHARMA, Jagpreet SINGH, Sarabjit KAUR, *„A Review: Data Warehousing, Its Issues, Architecture and Tools*", International Journal for Innovative Research in Science & Technology, Volume 1, Issue 3, August 2014, ISSN: 2349-6010;

[8] Joseph M. FIRESTONE, *„DKMS Brief No. Six: Data Warehouses, Data Marts, and Data Warehousing: New Definitions and New Conceptions"*, http://goo.gl/VcSOkp;

[9] Ralph KIMBALL, Margy ROSS, *„The Data Warehouse Toolkit: Second Edition"*, Wiley Computer Publishing, 2002, ISBN 0-471-20024-7;

[10] *TDWI Data Warehousing Architectures: Choosing The Right Data Warehousing Approach*, 2005, http://goo.gl/rpdmV6;

[11] W. H. INMON, Claudia IMHOFF, Ryan SOUSA, " *Corporate Information Factory, 2nd Edition*", John Wiley & Sons, 2001, ISBN: 978-0-471-39961-2;

**Alexandru Adrian ȚOLE** (born in 1986, Romania) graduated from the School of Domestic and International Business, Banking and Finance, the Romanian – American University, in 2009. He also graduated the Scientific Master Program in Finance, Banking and Insurance. He works at the Ministry for Information Society. He is pursuing a PhD Program in the area of Executive Information Systems.