INTERACTIVE DECISION SUPPORT SYSTEM BASED ON ANALYSIS AND SYNTHESIS OF DATA - DATA WAREHOUSE

Prof. Georgeta Şoavă Ph. D
University of Craiova Faculty of Economics and Business Administration,
Craiova, Romania

Abstract: : Systems to assist decision are based on data analysis and synthesis makes merging, categorizing, linking and grouping data to obtain information to highlight positive or negative factors that influence company performance. Getting results in the form of reports are the result of special techniques to explore large amounts of data. These techniques lead to highlight correlations between data, make estimates and forecasts and warnings of the managers of dysfunction. To improve business performance, management should be centered on the client, to understand customer needs, so we considered appropriate approach to data warehouses that an interactive system to assist decisions based on data analysis and synthesis. I presented the concept of data warehouse and basic features, representation of data in the data warehouse, differences between data warehouse and database, and finally to address environment data storage and use of data warehouse and to conclude with a set of conclusions.

JEL classification: M15, C81

Key words: decision; data warehouse; business; decision support; performance

1. INTRODUCTION

The solutions provided by information technologies for data synthesis, whose volume explored is huge, and for which conventional methods have become ineffective, are specific and dedicated software; queries which allow data clustering by criteria established and provides functions for fields thus created, total and subtotal functions provided by the generators permit reports indicate grouping hierarchy criteria. In recent years, gained more ground modern technologies as data warehousing and OLAP (On-Line Analytical Processing) software as the media became transactional data support systems.

In a broader concept applies to new research (intelligence) business, BI (Business Intelligence) becomes a critical component of all daily operations of business, so you develop real-time data warehouse that provides USER final rapid updates and issuing alarm signals (alerts) generated from transactional systems (TPS). Data warehouses in real time, RTDW (Real-Time Data Warehouse), and BI supports business performance plan economic organization. A simple form of data analysis is to compare the data with similar data; comparison is made keeping all the same criteria, only one with different values. Comparison is made between comparable data sets and compared technologies are equipped with observation techniques for signaling patterns, correlations, association by similarity or notices deviations, exceptions. Computer came

to meet these requirements with graphical presentation techniques that transform the information qualitative quantitative information. Observation techniques have appeared analytical data are based on mathematical theories which actual data are compared with theoretical data produced by a hypothetical model.

Developing techniques of observation led to the observation techniques based automatic data-driven. The result of such techniques is found in a general model. Observation data analytic techniques is found in a modern technology called Data Mining. The outcome of the observation is to obtain analytical patterns, correlations and models that can sometimes infer trends or can predict with a certain probability how the data will show a further period. The model allows interpretation of data, which is a cognitive process with a general appreciation of the situation, and identifies problems, opportunities or potential causes of failure.

2. DATA WAREHOUSE - DEFINITION, CHARACTERISTICS

This Data Warehouse is a set of special data products to support management decision making. Data warehouse contains historical and current data of potential interest to managers in the company. The data are structured to be permanently available for on-line analytical processing (OLAP), Data Mining, queries, reports and other applications to support decision making (Turban E., 2007).

The concept of data warehouse (Data Warehouse) refers to the final result - data stored on hard information, data characteristics distinct from transactional databases, while the concept of data storage (Data Warehousing) concerning the whole process of creating, maintenance and operation of a data warehouse.

Data warehouse (Data Warehouse) is a branch of applied computer information systems for decision support, SIAD or DSS, through which complex business and ensure the access from the outside, timely and effective information and knowledge necessary business (business information and business knowledge).

The data storage (Data Warehousing) involves the following components (Figure 1):

1) Data sources;

2) Extracting, transforming and loading data from operational databases, ETL (Extraction, Transformation and Load);

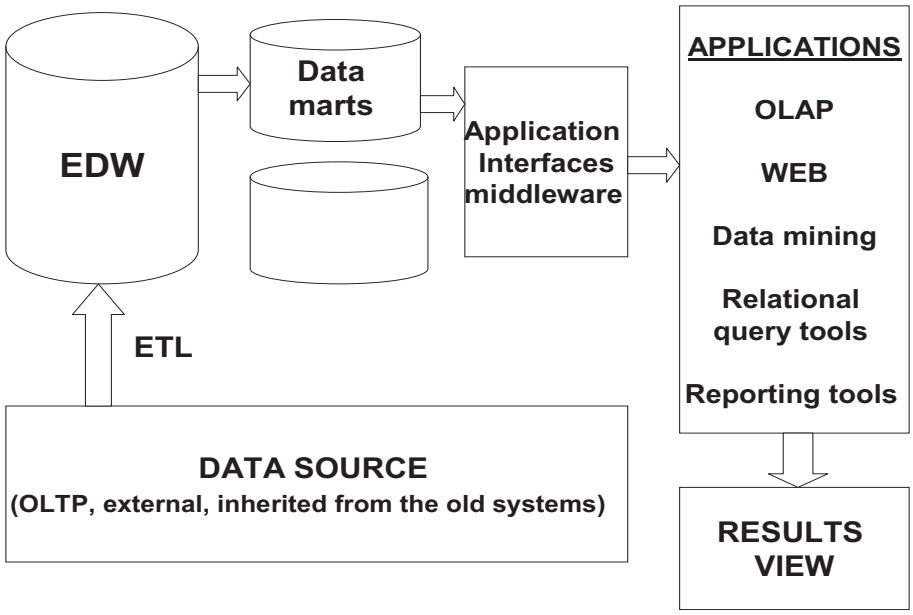
3) Type enterprise data warehouse, EDW (Enterprise Data Warehouse);

4) Metadata (software programs for data and rules for organizing data summaries. Are easily indexed and found, including Web tools);

5) Type middleware tools that provide access to the data warehouse (OLAP, data mining, software tools for reporting and data visualization).

Need for data warehouses is given the huge volume of data accumulated for economic organizations, whose integration into a structure to guide the decision making has become the focus of new information technologies.

Data warehouse integrates different types of databases at company level, providing timely and relevant data (real time or near real-time response) for decision support systems for management, SIAD. Implementation of data warehouse provides strategic value of economic organization and helps reduce costs.



***Source**: Turban E., Aronson, J.E., Liang, T.P., Sharda, R. - Decision Support and Business Intelligence Systems, Pearson Prentice Hall, New Jersey, 2007*

Figure no. 1 Data Warehouse

Fundamental characteristics of data warehouses (Turban - 2007) can be summarized as follows:

a) focus on issues - contain only relevant information for managerial decision making;

b) integration, orientation feature directly related topics - data from various sources are arranged in a consistent format, thus eliminating conflicts and discrepancies that may occur due to different measurement units. A data warehouse must be fully integrated;

c) non-volatile: once data has been entered ("written") in a data warehouse, users can not modify or update data;

d) variability in time (time series) - a data warehouse has the ability to maintain historical data of economic organization, and they are detected based on trends, deviations, long term relationships that lead to comparisons and forecasts decision. For each data store there is a temporal quality. Time is one of the important dimensions that need to have all data warehouses. Data analysis from multiple sources that contain multiple time references (egg, hits daily, weekly, monthly, etc.).

e) include Web-based applications.

f) use architecture client/server.

g) structures using relational databases or multidimensional databases.

h) use of metadata (data about data).

The characterization data warehouse should be left to their ability to store large volumes of data from archives and/or databases of computer applications specific to current business and economic organization from external sources. Exploiting these huge volumes of data is ensured by the existence of special engines that enable the

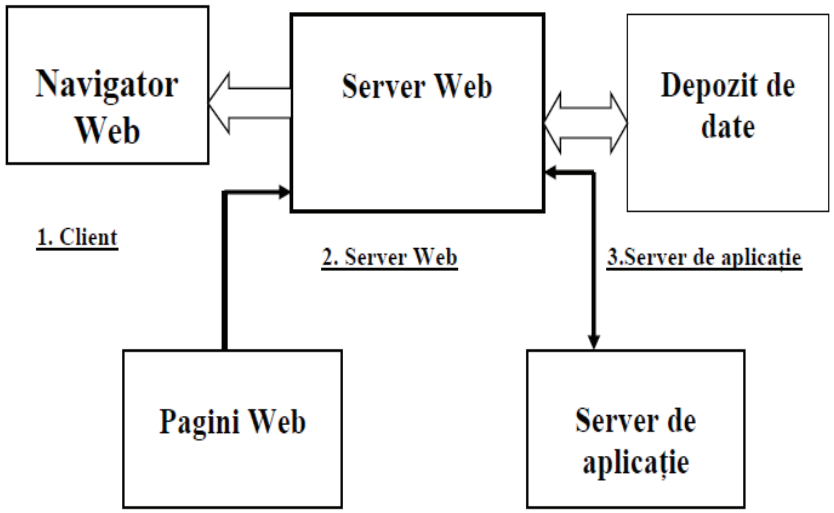
masses can be queried, and the existence of special services on-line data analysis (OLAP). Performance support software supports these services by transforming data, correlating and supplementing them and by creating data dictionary, all providing access to the primary. Data are extracted from heterogeneous databases created by existing information systems within the organization on various hardware and software platforms.

Input data warehouse is done under the control of specific applications and DBMS's, ensuring the integrity services, storage and safe work best. The data that form the primary support transactions are then processed to obtain necessary information synthesis and decision making and planning instruments are treated DBMS.

Since the operation of a huge amount of data, to obtain various reports, is provided by the integrity and consistency of the database, meeting all these data leads to the exploitation of large numbers of tables, creating multiple virtual connections and temporary tables, basically a large work that may materialize in the main drawback of the data, namely time deposits greater need their exploitation. Another drawback is the database engine and aggregation with centralized tasks that slows as current transactions. Appears thus need data storage that are dedicated to planning and strategic decisions in a different operating system so that operation of both systems took place without inconvenience. Data warehouse can store the archives of previous activity data and data on subsequent transactions without the user can intervene.

Are repositories of data concentration organize, strengthen and centralize data from heterogeneous sources that will be necessary so the analytical processing of decision processes, he practically builds gradually, by completing and future developments. To ensure high quality data which are undergoing a process of purification and transformation, indicating the manner of obtaining the data collected from the existing process that results in reducing the time required to obtain final reports. The data warehouse is made explicit in the data processing codes and nomenclatures data integration in data transactions, a process called normalization and is characterized by the fact that not change the data integrity and retrieval speeds. In a data warehouse data redundancy is allowed.

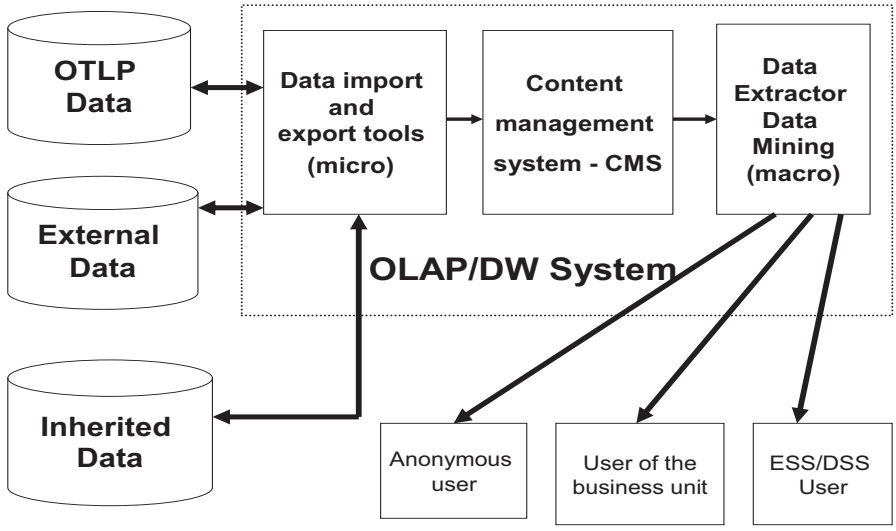
In the literature found several data warehouse architecture, grouped into two categories: type enterprise data warehouses, EDW (Enterprise Data Warehouse) and racks of data (Data Marts). An example of data warehouse architecture based on Web (Turbans, 2007) is shown in Figure 2 is 3-entity architecture (three-tiers) including client, Web server and application server. On the client side there is an internet connection and a Web browser-based graphical user interface GUI (Graphical User Interface). Medium of communication between the client (workstation - Work Station) and server type is Internet/Intranet/Extranet, on the server using a Web server for managing information flows between client and server, followed by the application server and storage Data.



Source: Turban E., Aronson, J.E., Liang, T.P., Sharda, R. - Decision Support and Business Intelligence Systems, Pearson Prentice Hall, New Jersey, 2007

Figure no. 2 Architecture of Web-based data warehouse

Generalized data warehouse architecture (Figure 3), is considering applying triad "storage-processing-presentation", the content management system, CMS (Content Management System) as a central element of data processing.



Source: Lenz, H., J., Thalheim, B., OLTP-OLAP schemes for sound applications. In TEAA 2005, volume LNCS 3888, pages 99-113

Figure no. 3 Generalized architecture of the data store

A CMS consists of two basic elements: the content management application, CMA (Content Management Application) and application content delivery, CDA (Content Delivery Application) (searchsoa.target.com). A CMS system indexes all data within the integrated economic organization, being used patterns or sets of patterns (templates) approved by the management organization, working as guides (wizards) and other tools to create and modify Web content.

The format management feature documents, CMS formats provide working with old documents (inherited), scanned paper documents that can be converted to HTML or PDF (Portable Document Format), so it can provide an update to the latest version document or restore a previous version document.

The decision to choose data warehouse architecture is influenced by several factors, of which there are:

- Information requirements of top management (top management);
- Interdependence information between departments and functional entities of economic organization;
- The degree of limitation of resources economic organization;
- Compatibility with existing systems already in operation in the organization;
- Deep motivation of employees in developing a data warehouse.

Integrating data in a data warehouse contains three major processes:

a) access to data;

b) achievement data federation;

c) timely reflection on the data store of significant changes in the data from data sources such enterprise.

Data integration and metadata in a data warehouse there are several technologies:

1) type enterprise application integration, EAI (Enterprise Application Integration) provides unified way of taking data from various sources and storage in data warehouse type business, DEW. This type enterprise application integration is at the application programming interface, API (Application Programming Interface). EAI is combined with service-oriented architecture, SOA (Service-Oriented Architecture);

2) service oriented architecture, SOA (Service-Oriented Architecture) is focused on a coherent set of business processes oriented Web services;

3) extracting, transforming and loading data into the data warehouse, ETL (Extraction, Transformation and Load) is the integral component of any project centered organization, storage and processing, as data storage (Data Warehousing). Data extraction means for reading data from a data warehouse or more databases. Transforming data is extracted from its original data conversion in standard form required for storing the data warehouse or other database that works directly store data. Loading data storing means, previously transformed into standard form, clean and refined, the data warehouse. Data sources may be relevant for data storage (shelf data) transactional database (from the computer system transaction processing, TPS), data from ERP from CRM from Excel tables (in a broader framework, the specific database OAS), the specific knowledge base KWS, the string of messages from external files, etc..;

4) type enterprise information integration, EII (Enterprise Information Integration).

Data stored in data store are consistent (the standard form gained through ETL process) the business rules that define how to use the stored data, the rules of compiling

abstracts (summaries, reports), with standardization rules coded attributes, the rules for making calculations. All these rules are stored in a metadata database and are applied consistently throughout the data warehouse.

Turban (2007) distinguishes three main types of data warehouses: data racks, DM (Data Marts), operational data buffer memories, ODS (Operational Data Stores) and type enterprise data warehouses, EDW (Enterprise Data Warehouse).

3. MARTS DATA

Starting from the organization of an enterprise, namely functional departments and entities, and data are submitted on the shelves of data (data warehouse), DM (Data Marts). If the data warehouse (Data Warehouse) combines databases across economic organizations (i.e. the integrated system of economic organization), but data racks smaller than data warehouses, intended for a specific department of the company or a subject defined the end-user desire. Using data warehouses or shelves (Data Marts) leads to increased operational performance.

These repositories are usually built with relational technology, so most shelves are shelves dependent data (that are generated directly from the data store). There are, however, and independent data store shelves organization's data that are designed specifically for a strategic business unit, SBU (Strategic Business Unit) and not using data from the data store of economic organization. Shelf independent data can be an alternative "low cost" for companies which can afford the purchase, maintenance and operation of a data warehouse.

4. OPERATIONAL DATA BUFFER MEMORY

Between operational databases (transactional) on the economic organizations operational and data warehouse management on the organizational environment can provide operational data buffer memory, ODS (Operational Data Stores), which may serve, for example, for implementation and use customer information files, CIF (Customer Information File) can be updated according to the evolution of business. As a result, operational data buffer memories are used for decision support in the short term, especially in applications with critical points.

Memoirs of operational data buffer stores the latest data on the subject considered, data from multiple sources. Data from operational data buffer memories are data from various sources that have undergone a process of extraction, transformation and loading, ETL (Extraction, Transformation and Load), similar to the data warehouse. When operational data are analyzed multidimensional data buffer memories become operational shelf operational data (Operational marts) (Turban, 2007).

5. VERSUS DATA WAREHOUSE DATABASE

An enterprise data warehouse type, EDW (Enterprise Data Warehouse) is an integrated data warehouse widely, with a large volume of data that is used in environmental management and strategic levels to support decision making.

EDW data in standard format are coming from various sources, and are used as input for most types of economic systems for decision support.

Differences between data warehouse and database can be summarized as follows:

a) contained in a data processing system transactions, OLTP (On-Line Transaction Processing) type are operational and the data contained in a data warehouse

to assist specific decisions are centralized data derived from operational data, not change time and are intended for end users;

b) for transactional systems, performance refers to the integrity, confidentiality, reliability and response time as a large number of users enter data into the system, while for SIAD (i.e. repositories) number of end users (managers) is

very small. This security and reliability and are not subject to major risks, backup and recovery procedures being less used than when transactional systems;

c) transactional systems process data sets are relatively small, recently introduced compact, so that processing is fast enough. Decision-making processes, their data requirements are large, dispersed stored leading to slower processing;

d) built databases for transactional systems are designed and developed based on known and clear requirements, changes that occur due to system adapt to the changes repeat certain stages of life., but once implemented they work long periods without changes . The SIAD requirements are only partially known at the time of design and their realization, which require data storage requirements to adapt on the fly, so it is found that the data managed transactional systems are viewed as a whole, while those in the data warehouse are organized in sections as they are organized according to subject analysis;

e) transactional systems usually reflect the flow of data from current activities, while data warehouses are focused on topics such as egg resources, products, customers, suppliers.

f) The operational systems of the organization dedicated management areas, data are often fragmented, so that managers make decisions based on partial information (incomplete). Data stores eliminate this disadvantage by accessing, integrating and organizing operational data to key role in a form that is characterized by consistency, reliability, suitability, availability and quick reference while (Timely).

6. CONCEPTUAL MODELING OF DATA WAREHOUSE

Data Warehouse is a collection of data oriented topics, integrated, correlated and non-volatile while supporting the decision. Dumped deposit data are integrated using this convention for measurements, attributes. Structure that provides data warehouse provides identification data stored punctual and, especially, quick access to them. Designing data warehouse structure is multidimensional modeling, implementing its structure is a database that provides large amounts of data storage and quick access to them, so-called database client/server.

Population data warehouse is by taking the transactional systems, but will be subject to complex processes of transformation that meets the deposit structure was designed. After this stage the deposit to join the service to obtain analyzes and reports. The steps listed above (design, popular, mining) are assisted by specialized software from browsers and generating reports until specific data mining tools.

The current operation wills frequently stored information is new requirements that will inevitably the structure, the popular extensions including historical data and the integration of new data incorporated into analysis applications. During the existence of a data warehouse is incremental and cyclical.

The design phase of a data warehouse using dimensional models that groups data in relational tables or diagrams of star-snowflake. In these schemes can be found quantitative data such as quantities or values or grouped by various criteria (customer, by product, service types, etc.). Quantitative data in dimensional databases are of

average number of transactions, centralization by certain characteristics, and totals are measures of activity. On the other hand, the aggregation will be called dimensions. Measures identified by size will be stored in a relational table which is called facts table and associated codes or criteria used are data aggregation type classification tables are associated with tables of facts and thus will be of a relational schema star. If such schemes meet several star catalogs use the same pattern form a constellation. If classifications can be divided into sub-nomenclatures then there is dependence between them. Note that the same code there may be several alternative nomenclatures. If these overtones and size integrate alternative scheme creates a form of snowflake.

Type schemes star, snowflake or constellation are multidimensional conceptual models of data warehouses, with the role of data organization on topics necessary decision-making process are open, i.e., change throughout life data warehouse.

7. DATA STORAGE ENVIRONMENT AND HOW TO USE THE DATA STORE

Environment that builds and operates a data warehouse contains the following elements: sources of transactional data, design tools, development, extraction and transformation tool data management system of database access tools and data analysis and management tools (Zaharie, D., Albescu, F., 2001).

If belonging to Microsoft's suite, all listed components are integrated in a working environment for data storage (Data Warehousing) for various versions of SQL Server. This environment offers assistance to design, implement and manage repositories throughout their life cycle.

Using this framework for Data Warehousing provides an architecture that can be integrated relatively easily with products coming from other platforms, provides Services in the import-export data validation and transformation, provides integrated metadata repository design and manage support, task- and events.

If Oracle business suite, Oracle Express product is a database management multidimensional SGBDMD, which is based on multidimensional data model, the client / server, Web application development capabilities, which include the following architectural components : utilities for managing (Express Instance Manager, Express Administrator and Relational Access Manager), development tools (Oracle Express Analyzer, Oracle Express Objects, etc.) and core (Express programming language). Oracle Express OLAP tool called Analyzer provides selection, analysis and visualization of data stored in multidimensional database and Oracle Express Objects tool provides OLAP application development using Express programming language. The tool integration with Oracle Discoverer Oracle Express Objects is facilitated development of applications for research (smart) business (Integrated Oracle Business Intelligence Tools). There will also be a connection editor, Express Connection Editor, which provides the definition of multidimensional database connections.

For a data warehouse can be processed is necessary to have a specialized set of tools: physical and logical description of data sources, warehouses or data warehouse where they will be incorporated, validation, cleansing and transforming data to be stored in data warehouse, users, tools that allow them access to data stored in that site. These are specialized tools for application development environments, software and specialized data analysis applications personal (individual).

8. CONCLUSIONS

Software tools for decision support have classic main purpose providing analysis techniques, optimization and simulation and graphical representation of results. The new assistance systems using decision fusion techniques for the data stored in non-uniform structures to use default information not specified in existing data.

Data warehouses contain unique structures, integrated and cumulative process of decision necessary, assuming the administrator as its main task the establishment of shared access to the categories of managers by providing passwords and access rights.

Using data warehouse (Data Warehouse) may facilitate:

- Improving the economic value of performance management organization focused on customers, through a better understanding of customer needs;
- Analysis of customer perceptions of value products and services offered or that could be offered in the future;
- Integrating marketing with information and communication technologies (ICT), the operational data sources, achieving enterprise data warehouse type, EDW (Enterprise Data Warehouse);
- Contribute to increase the strategic value of economic organization.

Data warehouse managers are selectively accessed according to their needs. In this way creates specialized collections in various fields which are called data warehouses (Data Marts). Data warehouses can be used as an intermediate structure for collecting data from primary sources and whose content is periodically downloaded data warehouse. Data warehouses can take birth through comprehensive data storage of transactional systems for applying data mining technology. Using Data Mining technology means that data processing is done without user intervention, the background and the results are stored for future reference on request.

REFERENCES

1. Lenz, H.J., Thalheim, B. OLTP-OLAP schemes for sound applications. In TEAA 2005, volume LNCS 3888
2. Turban, E., Aronson, J.E., Liang, T.P., Sharda, R. Decision Support and Business Intelligence Systems, Pearson Prentice Hall, New Jersey, 2007
3. Zaharie D, Albescu F, colectiv Sisteme informatice pentru asistarea deciziei , Ed. Dual Tech, 2001
4. * * * http://www. wikipedia.com
5. * * * http://www. searchsoa.target.com