"HENRI COANDA"
AIR FORCE ACADEMY
ROMANIA

GERMANY

"GENERAL M.R. STEFANIK"
ARMED FORCES ACADEMY
SLOVAK REPUBLIC

INTERNATIONAL CONFERENCE of SCIENTIFIC PAPER
AFASES 2011
Brasov, 26-28 May 2011

DESIGNING A DATA WAREHOUSE. CASE STUDY: SC "AMBIENT" SA

Carmen RADUT, Mihaela ALBICI, Delia TESELIOS, Anca VALCU

Faculty of Management Marketing in Economic Affairs, "Constantin Brancoveanu" University, Ramnicu Valcea, Romania,

Abstract: *The current paper approaches the concept of data warehouse and includes the design of the data warehouse in SC "AMBIENT" SA . The core of a data warehouse is a data base of large dimensions containing information used by end-users: clients, suppliers, advertising companies etc. In order to build the warehouse, the following steps have been made: extracting the data in operational data bases/external sources and copying them into the warehouse; cleaning the data and in order to be secure, the data should be correct when making decisions; loading the correct data into the warehouse; creating the data aggregates such as pre-calculated totals, subtotals, average values etc. Designing a data warehouse is a suggestion to obtain the warehouse needed by managers, analysts and specialists involved in strategic decision making regarding the development and future of an organization.*

Mathematics Subject Classification 2010: *68P05, 68P20l, 62P20, 93B15, 97R50, 97R99.*

Keywords: *Databases, Data analysis, Data Warehouses.*

1. INTRODUCTION

In the economic field, the decision-making process is based upon superior capitalization of data by synthesizing, analysis and interpreting which leads to transforming the information gathered into knowledge. This data synthesizing process implies data centralization using different criteria with the help of: specific and dedicated programs; interrogations that offer the possibility of grouping data following established criteria for the created domains; total and subtotal functions offered by the report generators which allow indication of grouping criteria hierarchies.

The immediate problem is that the volume of information is great, which leads to the failure of conventional methods and resort to the use of modern technologies like Data Warehousing and OLAP (On-Line Analytical Processing) for specific data transaction systems.

The data analysis consists of finding connections between synthesized data like: associations, structural correlations. The Data Warehousing represents a special data set which contains historic and present data of potential interest, produced in order to aid the managers through the decision-making process.

There are three different types of Data Warehouses in terms of area coverage of information: Enterprise Warehouses, Data Mart and Virtual Data Warehouses.

The Enterprise Warehouse collects all the information regarding matters that directly influence the well being of the company. The

Data Mart contains a subset regarding the volume of information of the company about a group of users. The Virtual Warehouse represents a number of views regarding the operational data bases.

Designing a Data Warehouse implies understanding and analysis of the economic processes and the construction of an economic analysis scheme. Designing a Data Warehouse implies engineering the following components: identification of the data sources; extracting, transforming and uploading the information in the operational data base (ETL – Extraction, Transformation, Load); Designing the Enterprise Data Warehouse; Designing metadata (general rules of organization of the information); use of middleware type instruments to insure access to the Data Warehouse (OLAP, Data Mining, software instruments used for data visualization and information feedback).

2. DATA WAREHOUSE CREATION

In order to design and create the Data Warehouses, two steps must be completed: the economic analysis (analysis of the economic processes and creating an economical analysis blueprint), designing method (implies the following procedures: top-down view, data source view, data warehouse and business query view) and the actual design of the Data Warehouse.

For example, in order to design a Data Warehouse using the top-down approach you start with the design and complete planning and it is used when the technology is fully matured and well-known and the economic aspects that must be undertaken are fully understood. The bottom-up approach implies a series of experiments and prototypes and it is utilized at the first stages of technological development and modeling. A combined approach of the two methods blends the planned and strategic aspect of the top-down method as long as it is kept in mind the advantage of a fast implementation.

Thus the designing and creating a Data Warehouse consists of: planning; requirements study, problem analysis, Data Warehouse designing, data compiling and testing and finally the use of the Data Warehouse.

The used methodologies are Cascade method, Spiral method. The Cascade method consists of executing a thorough structural analysis every step of the way before moving on to the next phase. The Spiral method generates functional systems which are progressively more complex, at short intervals, between two successive versions.

The actual designing process of the Data Warehouse consists of the following steps:

- Choose the economic process that needs to be modeled (stocks, sales, etc.). If the economic process is organizational and implies complex and multiple object collections there must be created a Data Warehouse. If the process is department based and focused on a single domain there must be created a Data Mart.
- Choose the level of granularity, the fundamental data, which is used to represent atomic data table for each work process.
- Choose the dimension (time, article, client, vendor, deposit, transaction type and state) applied to each registration of the data table.
- Choose the value (values, i.e.: sales_currency or sold quantity) which populates every registered line of the Data Table.

The Data Warehouses are based upon the multidimensional data model where data is seen as a data cube. This allows modeling and visualization of data in multiple dimensions and it is defined by dimensions (i.e.: perspectives and entities needed to store recorded information needed by the company) and facts (collections of quantified activities and dimensions which identify the mode in which they took place).

The dimension has a table associated which is called Dimension Table. The definition of a Data Cube is as follows: It is a data set organized and summarized in a multidimensional structure via a set of dimensions and measurements which represent a easy-to-use mechanism in order to access information in a short amount of time. The cube has a blueprint represented by the set of information tables. The central information table is the fact table and it is the source of the cube's measurements.

For example, if the granularity level indicates the detail level of the data found in the table, the granularity level is directly proportional with the level of detail of the hierarchy in the dimension table. The granularity is determined by the level of detail of the "Time" dimension. The hierarchy of the "Time" dimension in a decision-making dedicated system in the financial-bookkeeping domain is limited to the level of monthly reports. A more detailed data retrieving process can cause additional difficulties in data collection. If requested by the user, a daily data report can be considered.

That is why an important concept is data aggregation (pre-calculated values established upon the data on the analytic level of the deposit), which determine an important raise in terms of performance when it comes to response time of the informational request. With the help of aggregation information with a synthetic character can be obtained via analyzing the aggregated values without further intermediary calculus. Due to the fact that there are certain technical products that have similar fabrication procedures, the Product dimension details the production steps involved. Data is acquired from multiple data sources, then they are uploaded in the Data Warehouse, after which, based upon OLAP and Data Mining, the data is transformed in information and is directed to the beneficiary as a report.

3. CASE STUDY – BUILDING DATA WAREHOUSE FOR SC "AMBIENT" SA

SC "AMBIENT" SA has as a specific trade material and products for house construction and improving. Since the Data Cube allows modeling and viewing of the data in multiple dimensions being defined by dimensions and facts, in our case, SC "AMBIENT" SA, we can create a Data Warehouse for sales that contain registration based upon the following dimensions: time, article, branch and zone. With the help of these dimensions we will store the monthly sales on articles, branches and zones. Each dimension has a specific table associated with it named Dimension Table, which describes the dimensions. In the case of the dimension table for articles the following features will appear: article name, brand and type. The multidimensional data model is organized around a certain central theme, in our case, sales. Therefore we have a table full of facts, and the fact has a numeric measure and indicated the measurements thru which we want to analyze the relationships between certain dimensions.

Thus, the facts from the Data Warehouse include: sales in lei (sales-currency), quantity sold (number of units sold), total sales planed, and the facts table contains the names of the facts or measurements as well as the keys for each and every table form the dimensions table that are connected to it.

Figure 1 - Highlights how a Data Deposit works and the multidimensional schemes.

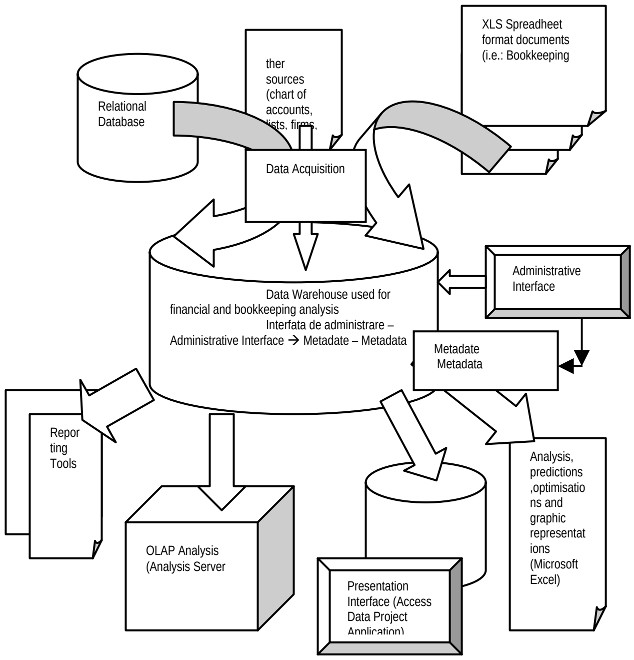

Figure 1. The architecture of the system proposed for implementation

Even though the given deposits in the cube are n-dimensional, usually, we use 3D cubes. The 4D cube built as a series of 3D cubes represented in Figure 2, we can continue to display any n-D data as a series of (n-1) D cubes.

For example, a sales scheme which is considered to have 4 dimensions (Time, item, branch, zone) is represented in the picture above. It contains a central table for sales which also contains keys for the other 4 dimensions.Before the two measures: sales-currency and quantity-sold. The “Zone” table contains the Key-Zone, City, County, Postal Code, Country. This definition could create a certain redundancy (i.e.: The cities Medias and Sibiu are both in the Sibiu County).

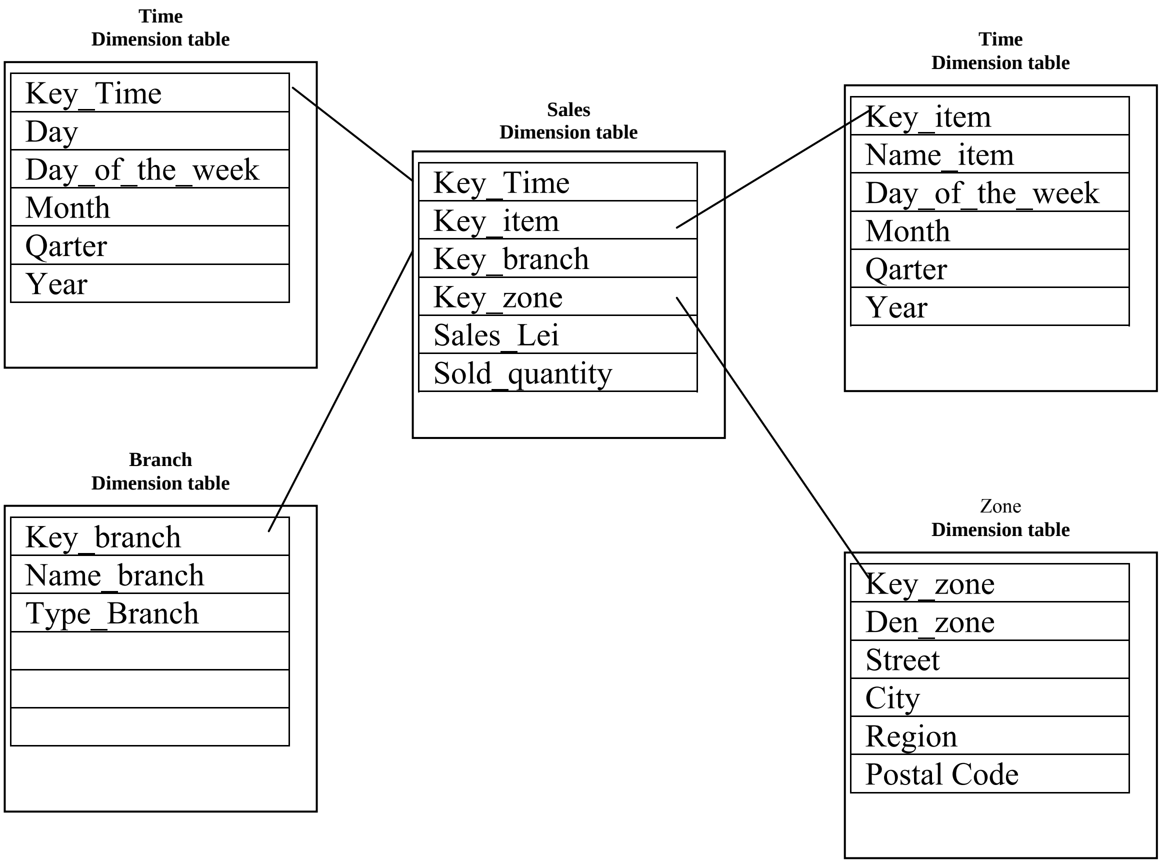

Figure 2. Star based blueprint of a sales data deposit

The multidimensional model requires that the data is organized in multiple dimensions, where each dimension contains multiple levels of abstraction defined my hierarchy. OLAP operations upon the data cube offers different views allowing interactive interrogation and analysis (Drill-up (roll-up); Drill-down; Dice; Slice; Pivot; Rotate). Of course the scientific demarche will continue with the data extracting process and transformation, for which extraction instruments are used, transform, integrated, cleanup and uploading data form Source Systems to one more data bases of the Data Warehouse.

4. CONCLUSIONS

For adequate decision making we need historical facts that, normally, data bases do not contain. Operational data, although very large, are far from adequate for complete decision-making. However Decision-making needs data from different sources, resulting high quality data, clean and integrated by the imperious necessity of Data Warehouse use, because operational data bases only contain detailed unprocessed data(primary data), such as transactions that need to be consolidate before analysis.

REFERENCES

1. Dănăiață, D., Mircea, G., Hurbean C., Popovics A., Hauer I.. *Tehnologia informației și a comunicațiilor pentru economişti*, Editura Mirton, Timişoara (2005).
2. McMahon, D.. *Rapid Application Development*, McGraw-Hill, International Edition (2000).
3. Oprea D.. *Analiza şi proiectarea sistemelor informaționale economice*, Editura Polirom, Iaşi (1999).
4. Turban E., Aronson, J.E., Liang, T.P., Sharda, R. *Decision Support and Business Intelligence Systems*, Pearson Prentice Hall, New Jersey (2007).