

Data Warehouse - The Source of Business Information

Malgorzata Nycz and Barbara Smok
The Economic University of Wroclaw, Wroclaw, Poland

malgorzata.nycz@ae.wroc.pl barbara.smok@ae.wroc.pl

Abstract

Recently the role and importance of database application in supporting the decision making process has been still increasing. Data warehouse can be an answer for this challenge. Data collected in a warehouse deliver different information e.g. about market, demand, selling goods, trends, anomalies on market. The paper presents data warehouse as a potential source of business information.

Keywords: data warehouse, economic information, OLAP

Introduction

The economic activity in dynamically changing environment puts high demands in front of enterprise that may have decisive impact on its development as well as on its possibility to participate in a competitive market game. Permanently changing environment and increase of competition have been reflected in strategic, tactic and also operational activities. Due to this the enterprise tries to implement the newest technologies to be more competitive and gain a market success. There are many different databases in enterprise that are often not compatible and from which only some information are of important value. Sometimes carrying out complex analysis on data coming from different information systems is very difficult or impossible. The data warehouse concept helps to overcome these problems. The data warehouse is a set of key information that can be used in enterprise management. This information can be of great value e.g. within carrying a promotion campaign or when deciding about the level of goods in a storehouse.

The enterprise value can be measured by the capacity of database collecting information about e.g. clients and the ability to use it. Along with the development of information technology, data gathered on different carriers have been one of the most important values in enterprise. Along with the enterprise development and its often reorganizations, data consolidation becomes a more and more complicated process. Usually business data are created and stored in many places that also may cause difficulties in the analysis creation. Oracle Warehouse Builder (OWB) is a tool that makes this process easier. It enables joining different data sources and transforming neces-

sary to their unification. OWB manages the data store as well and integrates it with the analytic tools assuring that data are uniform and actual.

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact Publisher@InformingScience.org to request redistribution permission.

Business Information

There are some basic definitions met in the literature that mean more or less the same: economic information, managerial information, business information.

As “an economic information” we can understand (Czerniak, 1996):

- Knowledge of the observer about the economic system and its environment,
- Content: quantitative and qualitative characteristics of the economic system components,
- The whole of the related components of the economic system,
- Knowledge about the economic object necessary for the observer to solve a particular task,
- Knowledge decreasing the indefinite situation in economic system when solving a task.

The glossary of mathematics and economic cybernetics (Fiedorenko, 1985, p.194) defines the term *economic information* as “information about social processes of production, share and exchange as well as consumption of material goods”. This definition is not very precise. Other definition says that managerial information is information about economic sciences achievements and their implementations in economic practice. According to J. Olenski “economic information sensu stricte is any information that is the complex economic object or the economic category; features are the economic category gauges measured on these objects”(Olenski, 2001). For example, for the object *enterprise*, the economic information is a production value, profit, credits, etc.

The objects of business information are processes, events or any other economic phenomena. The subject of economic information (the user) is an economically active unit conducting economic activity covering decision undertaking and being an object of influences of other social and economic subjects. Economic information is formulated in economic language and it causes results of business nature. It is any information that is an integral component of the economic system (Januszko, 2003).

In economic systems where the exchange and partition of goods and services are the basic functions, technical and business information dominates in contacts between the economic objects. It is also basic information asset in commodity turnover and in financial sector. The higher level of management system, the more important is economic information in relation to the other sorts of information (e.g. technical information). Economic information dominates at the macroeconomic level. Economic and production process control has been realized by technical information. Business information is applied first of all to making social-economic analysis and is often of great importance when decision making. A concrete decision has been always expressed as business and technical-economical information.

Information and knowledge should be actual and of high quality because they are very important competitive factors. The enterprise has difficult task to overcome. It must collect information and knowledge and it must be able to use it properly to have the advantage over other enterprises and to survive on dynamically changing market.

The fast development of information systems covers as well the technologies that support management in enterprises. We can observe that nowadays the process from data to wisdom is getting more and more important. Within it the following are realized: acquirement, collecting, updating, sharing and operating of information assets. Knowledge that is a subject of this process usually possesses many sources. (See Figure 1.)

The managers everyday undertake many decisions influencing the success (or its lack) of the enterprise. The decision making process requires not only knowledge and information but also imagination and knowing the context due to the fact that it is a complex process joining the today enterprise activity with it past activity. Information has been collected in databases. The enterprise environment is being permanently changing, with high level of risk and uncertainty. In such a situation the decision making process requires from the manager possessing large knowledge,

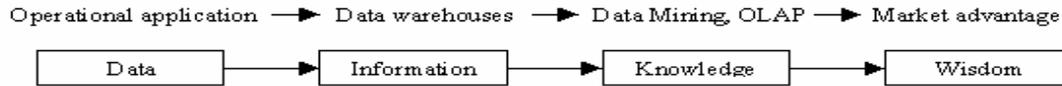


Figure 1. Knowledge and decisive processes

wisdom and thinking. The information systems keep data and enable registration of all updates. The basic problems transactions and the speed of information searching. Applications of this type are called operational here are, among the others, maintenance of integrity of changed data (the integrity chains, ones and the databases – operational databases. Besides this type of applications, there are also analytical applications having as a main goal the drawing information from data about patterns, business trends that can support the decision making. Adequately processed (and situated in a context) data are becoming information. Modern technology and IT enable transformation of information into knowledge. Wisdom can be treated as a way of seeing the problem or phenomena from many different perspectives and manifest itself in even acceptance of knowledge and ignorance. Intelligent supporting decision making systems emphasis on modern methods and techniques of information processing (data warehouses, Data Mining, OLAP) that support the decision-maker in different areas e.g. in management.

Data Warehouse

As a data warehouse we understand a collection of data that is:

- “subject-oriented,
- integrated,
- nonvolatile,
- time-variant”

and it supports the management’s decisions (Inmon, 2002, p.54). Data warehouses (DW) are not database systems in traditional understanding this term. They differ from databases among other things with e.g. data models, entering information methods, processing models, query optimization algorithms, visualization techniques (Grothe&Gentsch, 2000). Transformation from data to knowledge has been presented in Figure 2.

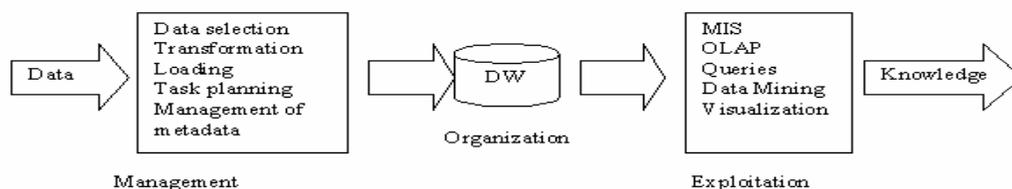


Figure 2. The example of the managerial information system based on data warehouse

Systems and applications based on the Oracle databases are created for processing a very large amount of information and for supporting the decision making processes. A warehouse can be treated as a specific central database where many different types of databases as e.g. copies of operational databases with data from the past, aggregated data (with the possibility to divide them into different dimensions), metadata. DW integrates the enterprise data that are processed using analytical applications within the decision support systems. Data that have been gathered within the data warehouse ought to be specifically prepared for further use in data warehouse. Such a preparation process must be correctly designed. Oracle Warehouse Builder (OWB) is a tool that can be used to create and design a data warehouse. The ETL (Extraction, Transformation, Loading) tools are complex integrated environments enhanced by the functionality and different data type drivers. There are many methods of exchanging data between data sources and target data warehouse. Many producers of integration tools encourage potential users to implement their products, especially in a situation where the data flow prepared in different formats and data are coming from different systems. Often it is impossible or difficult to see in a coherent way the context that data possess. (See Figure 3.)

The ETL process is responsible for the whole way of data from sources to the data warehouse environment. It covers such stages as capturing, transformation and move of data. Designing and implementation of this process are one of the most difficult stages within the warehouse life cycle.

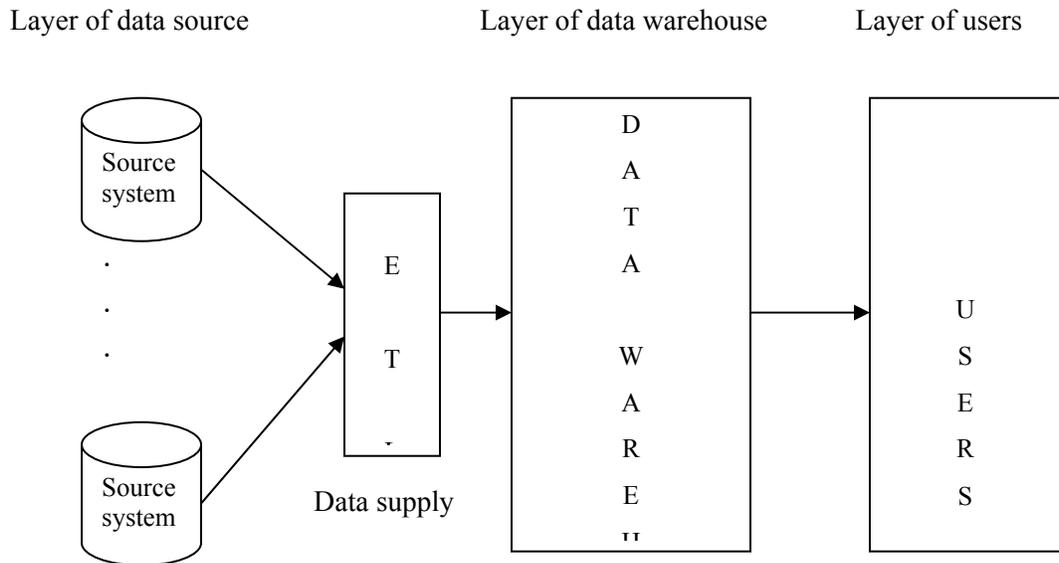


Figure 3. Location of the ETL tools in the data warehouse

The OWB structure is following. It consists of two main environments: the project environment and the realization one (Oracle Warehouse Builder, 2001). The processes connected with the logical architecture designing are realized within the project environment. The second environment enables the analysis realization required by the user. The project environment consists of metadata, project tools and repository that is the most important element of the data warehouse. Metadata are different information supporting the designing process, management, application and everyday usage of the DW as well as enabling control of supplying data. Metadata can be of different shape: databases list from which the DW is supplied, these database structures, data delivery rules between their sources and the DW, the data warehouse schema, query and report sets,

indexes, rules of the data warehouse objects division into partitions, rules of taking, transforming and correcting of the source data, etc.

OWB offers a special group of tools called validators that check data quality and completeness in repository. Within the DW one can distinguish the following (See Figure 4):

- Extraction – that extracts data from one or more sources and moves them to the warehouse,
- Transformation – meaning data transforming and converting into one common format,
- Loading –moving data to warehouse,

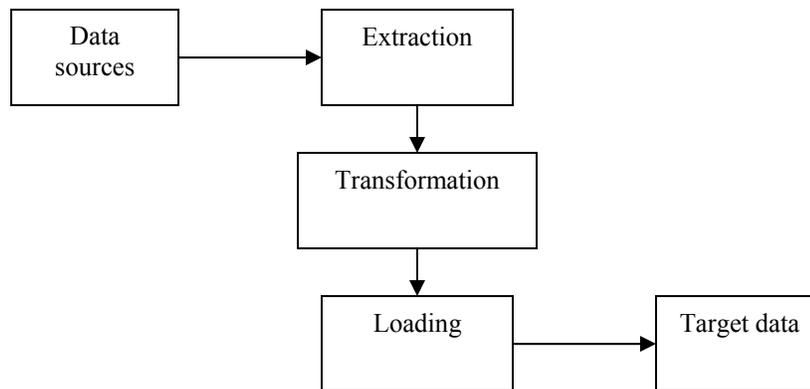


Figure 4. Schema of data extraction

The realization environment consists of databases on which the DW actually works and concrete source data delivered in the DW.

Data within the DW are organized in shape of the multidimensional data model that can be seen as a multidimensional OLAP (On-Line Analytical Processing) cube (Figure 5).

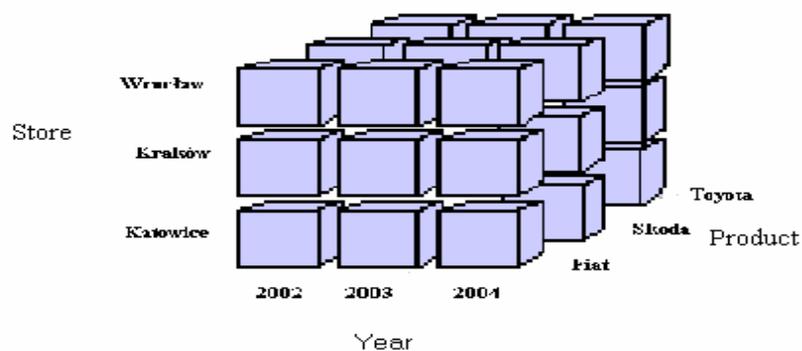


Figure 5. Multidimensional data model

A set of numeric measures called facts is a subject of analysis. Facts represent information that is to be analyzed (as e.g. product sales, student average notes, gain, etc.) and are characterized by measures (as e.g. number of bought products). The value of each measure depends on dimensions establishing the context of the analyze. The basic logical structure of the data warehouse is a star schema. As an example a star schema has been shown in Figure 6.

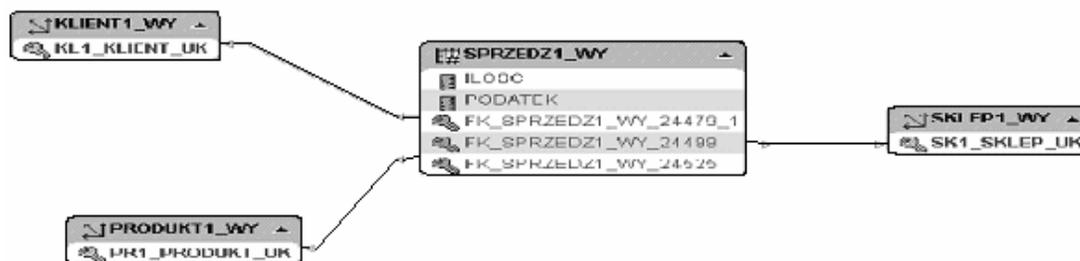


Figure 6. The example of the star schema in the Oracle Warehouse Builder

The star consists of a central fact table and some dimension tables. The star structure is one of the most popular models useful when designing the conceptual schema of the data warehouse. There are two other models, more complicated when compare with the star one: so called snowflake and constellation. The star structure operates on facts and dimensions. A fact describes a single information stored in the data warehouse. For example, the sales results in previous years is the fact. These results can be seen from different points of view, in different dimensions as e.g. goods, quarter, client (Nycz, Smok & Pondel, 2003). Facts are the quantitative measures and their value depends on the set of dimensions.

Usually the data warehouses use multidimensional data structures as a star or snowflake schemas. These structures describe quantitative information – so called measures (e.g. value of sales) using dimensions (e.g. town, product category, quarter, method of sales). Such a type of data processing has been called current analytical processing.

Data warehouses are one of the basic technologies of integrated information system creation. The main goal of this integration is to make possible the multidimensional analysis and comparison of data for the decision making reasons. The OLAP model is to support the analyze processes by delivering tools that enable analyzing database in many dimensions defined by the user (as e.g. time, place, product). Facts and dimensions can be stored on the relational database server ROLAP (Relational OLAP) or on the multidimensional database server MOLAP (Multidimensional OLAP). ROLAP is usually used for collecting row and aggregated data; MOLAP stores in data marts data that are partly aggregated.

But having the database is not enough to carry on analysis: it is necessary to have any decision support system: a reporting system (like e.g. Oracle Reports Builder), an analytical processing system OLAP (e.g. Oracle Discoverer or Express), a data exploration system (e.g. Oracle Data Mining). The basic advantage of any data warehouse is the possibility of collecting information coming from different heterogeneous sources. The integrated data are stored in the structures that enable quick data analyze. Data in relational databases are organized in rows and columns. The solutions delivered by the multidimensional analyze (OLAP) can be seen as enhancement of the relational databases possibilities. The OLAP database possesses the multidimensional structure and the data cube is its basic element. The data cube enables realization of such operations as slice and dice as well as the data aggregation. The cube can have three dimensions, but it is possible to have more dimensions. As some examples of dimensions used in sales analyze can be given as follows: clients, vendors, budget (planning, comparing the real results with those planned), discounts (the analyze of the sales or gains depending on a given discount, or for a given the price category), the sales regions. To make a typical analyze of sales using a cube, it is enough when the cube consists of seven at the most dimensions. The multidimensional recording in a cube makes it possible to very quick survey the data in all possible arrangements. The changed of presented data rely only on rotation the cube without the necessity of calculation of all values. Thanks to the multidimensional analyze we can:

- analyze huge amount of data having the answers generated very quickly,
- use the slice and dice as well as drill up, drill down modes due to different dimensions defined in the data structure, and
- quickly identify trends and recognize problems that could be impossible to diagnose without such analyzes.

The multidimensional data structure can be stored either in multidimensional databases and they are in this case a source of the OLAP tools or in enhanced relational databases – being a source of the ROLAP tools. Exact definition of information needs is required when using the functionality of these systems in decision making.

Multidimensional Approach to Information

Multidimensional approach means that information is stored in multidimensional tables (cubes). Such a cube is a set of data ordered by dimensions (Inmon, Welch, & Glassey, 1997). The dimension possesses an ordered hierarchy of the data aggregation levels. Due to this fact it is possible to see the dimension on different levels of details. Multidimensional data analyses are supported by special operators such as (Analizy, 2005, p.8):

- pivoting – it relies on showing the measure and defining the dimensions in which the measure will be presented,
- rolling up – it enables the navigation up the given dimension hierarchy to calculate the more and more aggregated value,
- drilling down – that enables to go deep into a hierarchy of the dimension to carry out the more detailed analysis, e.g. searching more detailed information about sales in each store, everyday for a particular category of products,
- rotating – enabling the presentation of data in different axes of dimension,
- slicing and dicing – that narrows the analyzed data to a chosen dimensions and in frame of each the narrowing the analysis to a concrete values,
- ranking – enabling the ordering of information in a given dimension.

In Figure 5 we present an example of the multidimensional analysis (slice and dice). We can see in it the aggregated data depicting the scale of the car sales in three following categories (dimensions): store, period and product. On these data we can build a cube that is a multidimensional database and we are able to obtain some information on e.g. how high was the sale of the product named Fiat in the store in Wroclaw in 2003. It is possible due to the fact that such information has already existed in the multidimensional database and it is to read. In traditional database this operation requires a complex query to be realized that needs much more time then reading the data from the multidimensional database. The analysis “in deep” allows mining the aggregated data through lower levels to identify the details. This technique enables searching the answers on questions like e.g. “why the sales of shoes produced by the firm “Rylko” in northern part of the country was lower in two first quarters in the year 2001”. Decomposing the aggregated data from the whole country into regions we can observe whether this decline depicted all regions in the country or only selected ones. If we have the aggregated data of sales split into provinces, it is easy to notice the decline e.g. in the Poznan province. When using this analysis it is possible mining through information to the level of towns, communes, etc.

Summary

Managerial decision-making requires the permanent access to all necessary information depicting all managerial areas, among other things sales, personnel, production process management, finances or accountability. This information is collected mostly in many non-integrated information systems and databases. There are more and more tools offered by different producers enabling automatic analysis. Data warehouse also belongs to this group. It can be treated as a set of the materialized perspectives being the answers to the questions directed to the sources databases. But it is not enough to collect data and to analyze them to find some regularity among data. This what decides about the success of the enterprise is the access in a short time to information and knowledge that enables to make right decisions.

References

- Analizy danych w modelach ROLAP i MOLAP.* (2005). IV edycja konferencji Hurtownie danych i Business Intelligence, Warszawa 2005 (in Polish).
- Berry, M.J.A., & Linoff G. (1997). *Data mining techniques for marketing, sales and customer support.* Wiley Computer Publishing.
- Czerniak, J. (1996). *Informacja i zarządzanie.* Warszawa (in Polish).
- Fiedorenko, N.P. (Ed.). (1985). *Słownik matematyki i cybernetyki ekonomicznej,* Warszawa. PWE (in Polish).
- Grothe, M., & Gentsch, P. (2000). *Business intelligence.* Addison-Wesley.
- Inmon, W.H., Welch, J.D., & Glassey, K.L. (1997). *Managing the data warehouse.* Wiley Computer Publishing.
- Inmon, W.H. (2002). *Building the data warehouse.* Wiley & Sons.
- Januszko, W. (2003). *Systemy informacji gospodarczej.* PWE, Warszawa (in Polish).
- Kimball, R., & Ross, M. (2002). *The data warehouse toolkit.* Wiley Computer Publishing.
- Nycz, M., Smok, B., & Pondel M. (2003). *Wykorzystanie Oracle 9i Data Mining do pozyskiwania wiedzy z baz danych.* In Zastosowania technik informacyjnych informacyjnych gospodarce i zarządzanie wiedzą. Red. Drelichowski L. Wyd. Akademii Techniczno-Rolniczej w Bydgoszczy (in Polish).
- Oleński, J. (2001). *Ekonomika informacji. Podstawy.* PWN Warszawa (in Polish).
- Oracle warehouse builder. OWB user's guide, Release3i.* (2001). Oracle.

Biographies

Malgorzata Nycz, Ph.D. Eng. is a lecturer in the Faculty of Management and Computer Science, Wrocław University of Economics, Poland. Dr Nycz has authored over 50 publications mostly oriented on intelligent systems topics, data warehouses and distance learning issues. Her current research is in intelligent systems with focus on data analysis, information and knowledge discovery from databases, and modern education including e-learning processes.

Barbara Smok, Ph.D. is a lecturer in the Faculty of Management and Computer Science, Wrocław University of Economics, Poland. Dr Smok has authored over 45 publications mostly oriented on databases, intelligent systems topics and distance learning issues. Her current research is in intelligent systems with focus on data warehouses, knowledge bases, and modern education including e-learning processes.