

Using ORACLE tools to generate Multidimensional Model in Warehouse.

Abstract. The paper presents modern techniques of data modelling and processing, collected by the company. It presents the process of multidimensional data modelling (include the transformation of logical structure to relational structure and next transformation to Multidimensional Model of warehouse in star or snowflake schema). It also shows the ETL process and methods of creating OLAP cubes by use of ORACLE tools to support decision making by business analysts. An approach based on data mining techniques allows analysts to capture certain features in customers, to offer dedicated products for the customer groups. Based on customer behaviour can be concluded about his tendencies to their certain behaviours and preferences.

Streszczenie. W pracy zaprezentowano współczesne techniki modelowania i przetwarzania danych pozyskiwanych przez instytucje. Artykuł przedstawia proces modelowania wielowymiarowych danych oraz proces ETL do momentu ich prezentacji z użyciem narzędzi firmy ORACLE. Przedstawiono metody tworzenia kostek OLAP wspomagające proces podejmowania decyzji przez analityków biznesowych. Podejście oparte na technikach data mining pozwala analitykom uchwycić pewne cechy wspólne dla klientów, w celu złożenia oferty produktów dedykowanych dla konkretnych grup. Na podstawie zachowań klientów można wnioskować o ich skłonnościach do pewnych zachowań i preferencji. (Modelowanie wielowymiarowych danych do ich prezentacji z użyciem narzędzi firmy ORACLE)

Keywords: Database in Management, Business Intelligence, Data Warehouse, OLAP.

Słowa kluczowe: Bazy Danych w Zarządzaniu, Business Intelligence Hurtowni Danych, OLAP.

Introduction

Today's companies need tools to support sales, production, management. The high costs of carrying out a marketing campaign and its preparation time will need a tool to facilitate this task. The overall aim of such activities is to maximize profits at the lowest cost. There are many ways to reach customers. The development of information technology has enabled the analysis of data for company profile. Clients of institutions that use BI tools also received new offers dedicated to their expectations. New technologies based on an analysis of data will save the time needed to complete tasks.

Companies gain a competitive advantage in the market by the more fit the profile of their activities to the customers' expectations. Capture trends in business such as the existing terms of the dynamics of growing sales of certain types of products it is now possible with the use of data analysis techniques. Each product has its own characteristics and cannot be addressed to all customers. This situation requires tools to support corporate control of production processes, sales and quality management.

An approach based on data mining techniques and expertise allows analysts to capture certain features in common customers, to offer dedicated products for the customer groups. Based on customer behavior can be concluded about his tendencies to certain products. The fulfillment of conditions provides that there is some probability that a customer buys a product.

Multidimensional data analysis techniques based on OLAP cubes and the construction of warehouse and data modeling are presented in this article using Oracle tools.

Modeling the Business and multidimensional structures

OLAP consists objects that are a part of dimensional model. The dimensional data model (include: dimensions, attributes, levels, hierarchies, measures and cubes) is highly structured and implies rules that govern the relationships among the data and control how the data can be queried. The fact table is referred to a cube, and the columns (in table) are referred to measures. The cube has edges, which are referred to dimensions. The fact table include measures that are linked to a dimensions. Each dimension is a grouping of related columns from one or more tables. Analysts know which business measures they are interested in examining.

Dimensions and attributes make the data meaningful, and how the dimensions of their business are organized into levels and hierarchies.

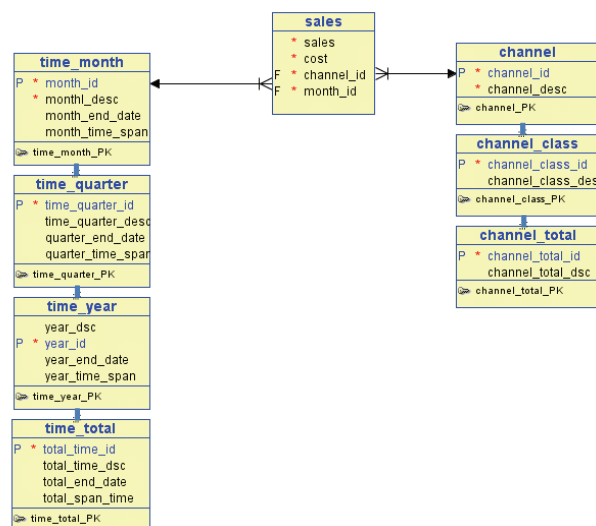


Fig. 1. Diagram of a Relational schema.

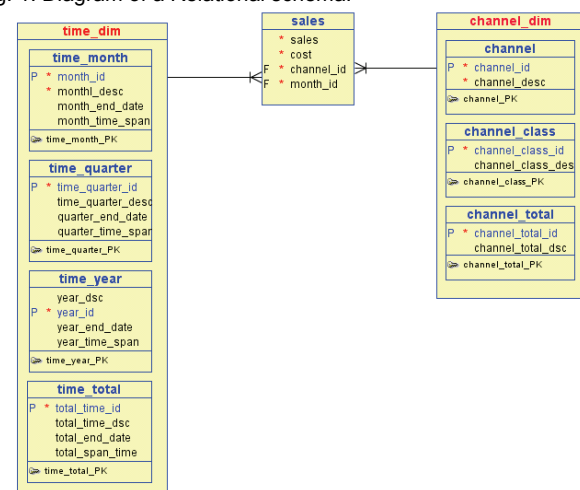


Fig. 2. Multi-dimensional schema.

Cubes are the physical implementation of the dimensional model and are highly optimized for dimensional queries. The OLAP engine that underlies cubes is based on an indexed multidimensional array model, which provides direct cell access. Diagram of a relational and multidimensional schema are presented on Fig1 and Fig.2.

The OLAP engine manipulates dimensional objects and are optimized for analysis, analytic and row functions can be calculated much faster in OLAP than in SQL.

Overview of the Multidimensional Data Model

The dimensional objects include cubes, measures, dimensions, attributes, levels, and hierarchies (Fig.3).

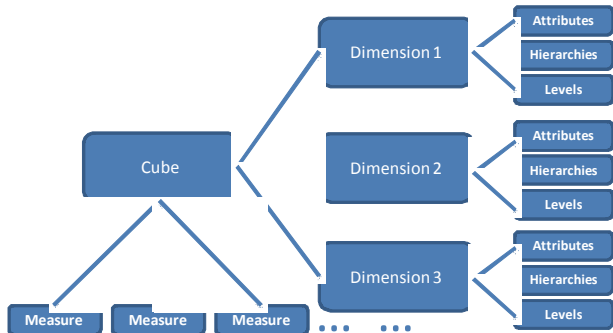


Fig. 3. Diagram of the OLAP Dimensional Model

Dimensions:

Dimensions contain a set of unique values that identify and categorize data, they form the edges of a cube. Base-level dimension values correspond to the unique keys of a fact table.

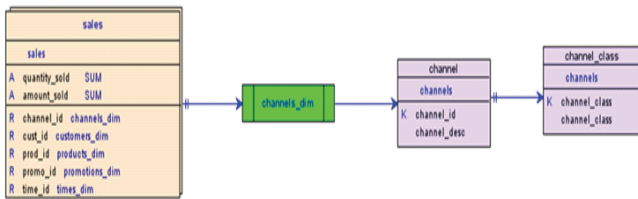


Fig. 4. OLAP multidimensional model

Time Dimensions:

Time dimension require additional information about the members in terms of each member's time span and end-date. With these additional attributes multidimensional model is able to provide additional features compared to other dimensions such as:

- Prior Value,
- Difference from Prior Period,
- Percent Difference from Prior Period,
- Future Value,
- Moving Average,
- Moving Maximum,
- Moving Minimum,
- Moving Total,
- Year to Date.

It is possible to automatically define the time dimension using OWB environment and fill it with data from a given range of years. To do this, perform the following steps using the Create Time Dimension Wizard:

- provide a name and description (e.g. time_dim)
- set the storage type (MOLAP – multidimensional storage, ROLAP – relational storage). If you have high volumes of data combined with high refresh rates or if you have detailed high volume data. Choose MOLAP to store aggregated data that is used for analysis.

- defined the range of date stored in the time dimension as Fig. 5.

- choose the levels in the hierarchy (Fig.6.)

Fig. 5. Parameters window for the dimension of time.

Fig. 6. Time hierarchy and levels window for the dimension of time.

The time dimension, giving levels, hierarchies, attributes are presented in Fig. 7.

Fig. 8. presents the mapping of the logical structure to the physical structure of time dimension, which is stored in a database.

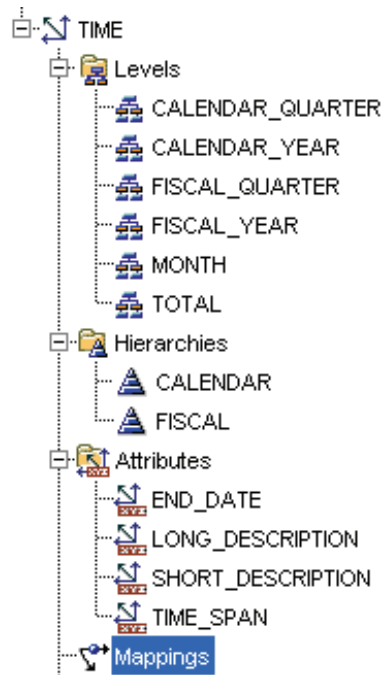


Fig. 7. Dimensional Time - Structure

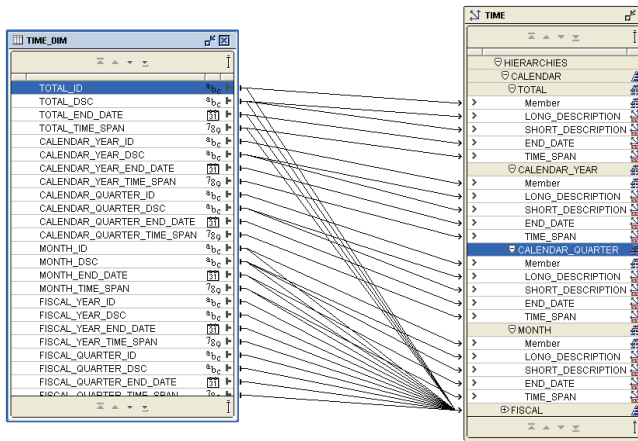


Fig. 8. Dimensional Time - Mapping

Hierarchies and Levels:

A hierarchy (Fig.9.) is a way to organize data at different levels of aggregation. In viewing data, analysts use dimension hierarchies to recognize trends at one level, drill down to lower levels to identify reasons for these trends, and roll up to higher levels to see what affect these trends have on a larger sector of the business.

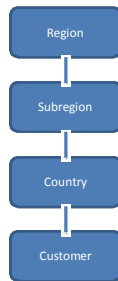


Fig. 9. Levels in a Dimension Hierarchy

Attributes.

An attribute provides additional information about the data. Some attributes are used for display. You might also have attributes like package, sizes etc.

Time attributes can provide information about the Time dimension that may be useful in some types of analysis, such as number of days in each time period.

Each attribute corresponds to a column in dimension table.

Cubes

Cubes provide a means of organizing measures that have the same shape, they have the exact same dimensions. A cube corresponds to a single fact table.

Measures

Measures populate the cells of a cube with the facts collected about business operations. Measures are organized by dimensions, which typically include a Time dimension.

An analytic database contains snapshots of historical data, derived from data in a transactional database, legacy system, syndicated sources, or other data sources. Three years of historical data is generally considered to be appropriate for analytic applications.

Table 1. Calculated measures

Group Name	Calculation Name
Basic Arithmetic	Addition
	Subtraction
	Multiplication
	Division
	Ratio
Advanced Arithmetic	Cumulative total
	Index

	Percent Markup Percent Variance Rank Share Variance
Prior/Future Comparison	Prior Value Difference from Prior Period Percent Difference from Prior Period Future Value
Time Frame	Moving Average Moving Maximum Moving Minimum Moving Total Year to Date

Defining OLAP objects

The following Oracle tools to define the analytical workspace (AW) are available: Oracle Analytic Workspace Manager (AWM), Oracle Warehouse Bulider (OWB), Oracle SQL Developer Data Modeler and other package.

All objects can be created in a dedicated and most powerful tool, which is Oracle Warehouse Bulider (including the model of ROLAP and MOLAP warehouse in the structure of a star or snowflake). OWB can perform the Extraction Transformation and Loading (ETL) engine using Oracle Workflow in order to fill the data warehouse. It allows both logical and physical design of the warehouse.

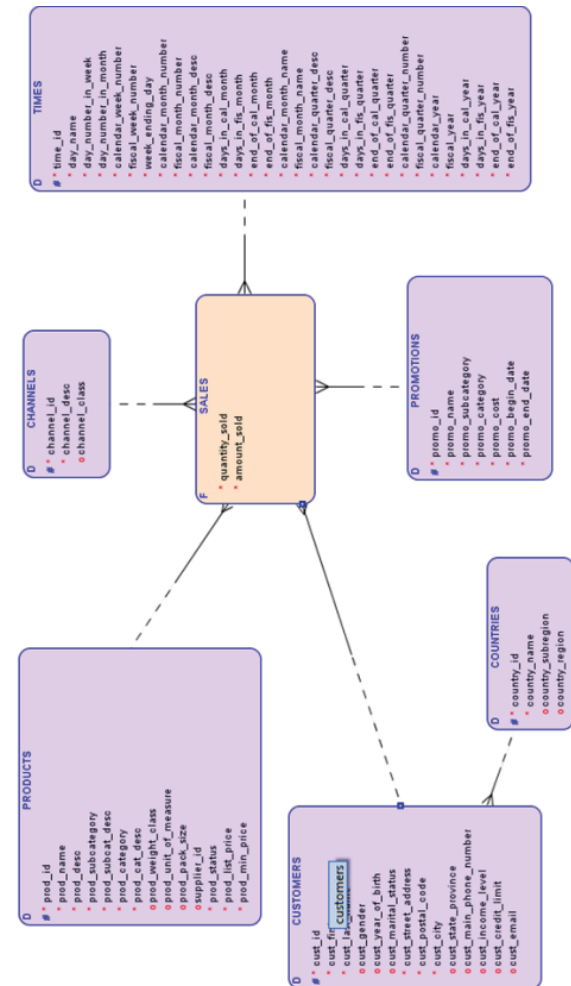


Fig. 10. Warehouse logical schema in the form of a snowflake.

The modeling process by using Oracle SQL Developer Data Modeler

Oracle SQL Developer Data Modeler allows modeling the logical structure of warehouse in star or snowflake schema . Entity tab allows to select (on object definition) the

type classification as Dimension or Fact. The next step is to set the definition of the Entity Attributes and Data Type. Each object in the form of such Dimension or Cube should also be set to a Unique Identifiers. Cube object usually has not unique identifier key. It is also necessary to carry out the mapping process, which is performed by selecting one type of relation 1: N relation, relation 1:1, M: N relation.

The logical model is transformed to the relational schema (Engineer to Relational Model) . In the process, you can choose which objects will be involved in transformation (Fig. 11).

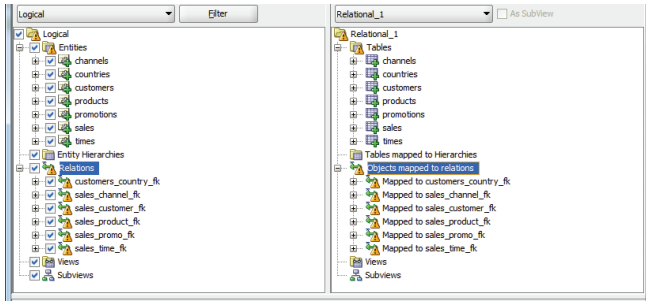


Fig. 11. Engineer to Relational Model process.

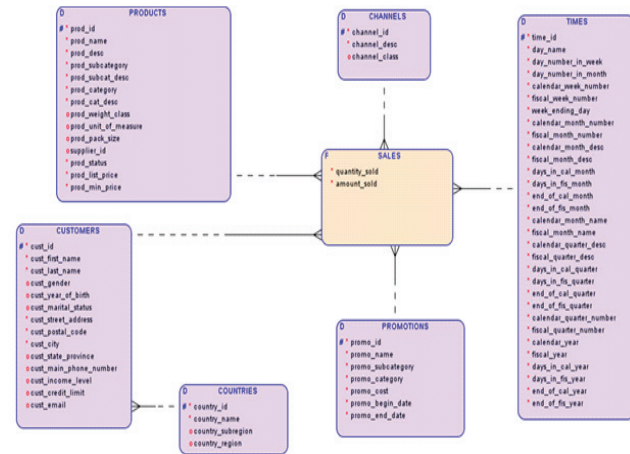


Fig. 12. Relational schema after transformation process of the logical model.

The next step is to create Multidimensional Model (MD) from the Relational Model (RD) in which is defined Physical Models based on the: attributes, levels and hierarchies. Such defined structure, for example, can be exported to Analytical Workspace (Oracle AW) and also be viewed in the AWM editor or can be modified as objects of OLAP.

Designing ETL

When source and target objects is defined in Oracle SQL Developer Data Modeler then the process of data movement and transformation logic should be started (ETL). ETL stands for Extract, Transform, and Load. ETL is designed in Warehouse Builder (OWB) and is used to create mappings to extract data from sources, transform it, and load it into targets. ETL process flows logic that interrelate mappings sources with targets and other activities can be designed using the OWB Process Flow Editor.

To build ETL process logic the Workflow Server should be started. The Mapping Editor (Fig.15.) interface in Warehouse Builder is used to design a mapping. You select operators from the Mapping Editor Palette panel and place them on the canvas.

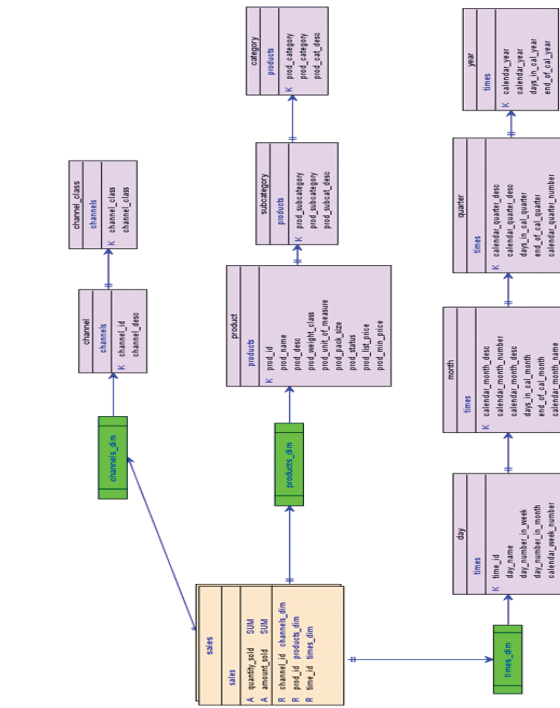


Fig. 13. Multidimensional Model.

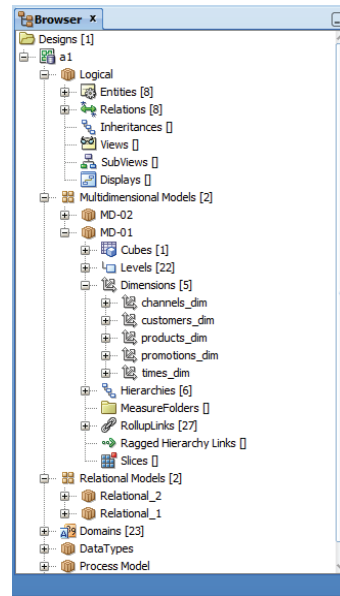


Fig. 14. Multidimensional Model.

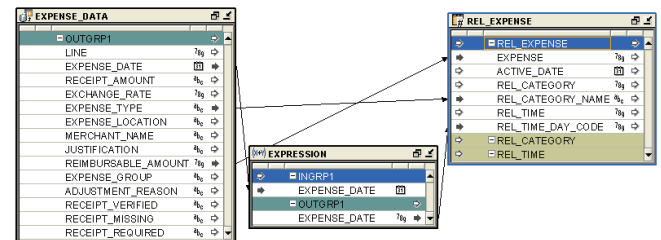


Fig. 15. The process of mapping sources with targets objects.

When mappings is done (that define the operations for moving data from sources to targets) the process flows can be created and defined. Process flows interrelate mappings and activities external to Warehouse Builder. In OWB, a

process flow is used to manage the execution of mappings and related activities.

The activities editor of process flow includes a variety of activities that can be added and then connected with transitions to designing of a flow. Transitions indicate the sequence and conditions in which to launch activities, an activity represents units of work in a process flow and these units can involve components internal or external to Warehouse Builder.

The AWM application can present data in tabular or graphically form. Such applications are used by business analysts to make decisions about functioning of the company and its development strategy. The analysts can manipulate cube objects with use of drag and drop methods. They may also limit the scope of the presented data using filters that limit data on individual dimensions, hierarchies and levels. The OWB has a similar methods of data presentation. Another way is to build own application with an interface to the data warehouse structure, e.g. using JDBC OCI8.

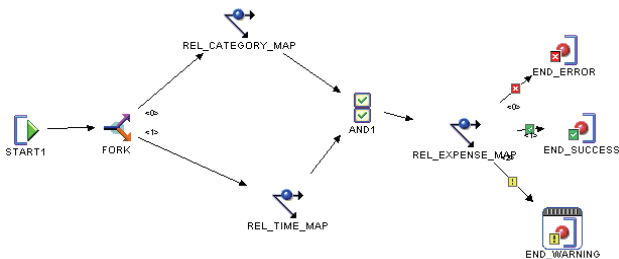
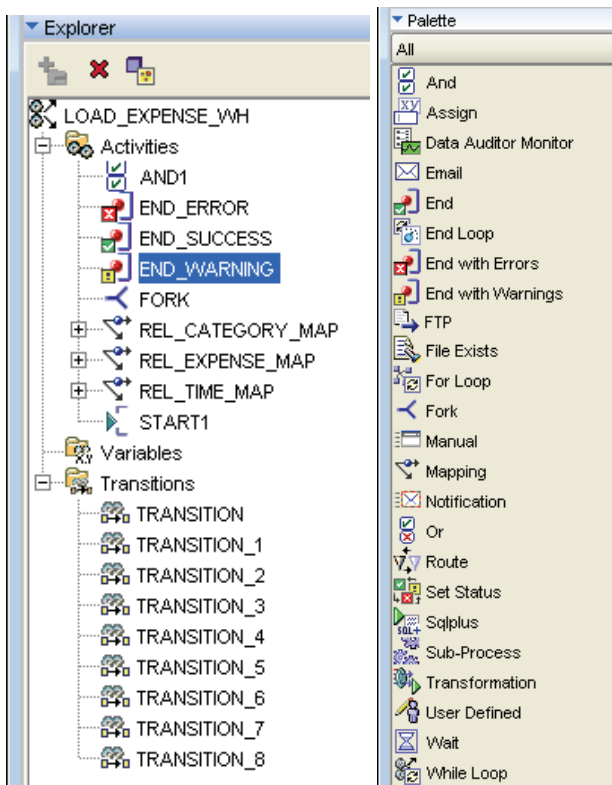


Fig. 16. Design a Process Flow in OWB using ORACLE Workflow Server

In the window Measure Data Viewer, data can be presented in the form of many types of charts (eg bar, line, pie). Fig. 18. shows the possibility to present data in a pie chart.

Fig. 19. presents a method to build reports based on the Excel worksheet in which the imported data retrieved from the materialized views created in the process of maintenance OLAP cubes. All necessary data is stored in a xls file. Reports can be created in a dynamic manner depending on the needs and presented graphically. The xls file consists all imported data from Cube.

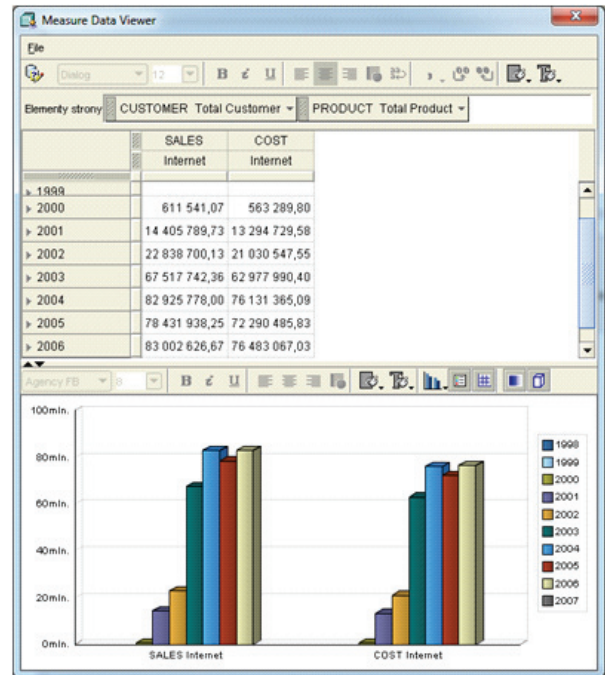


Fig. 17. Sample data from OLAP cubes (AWM or OWB application)

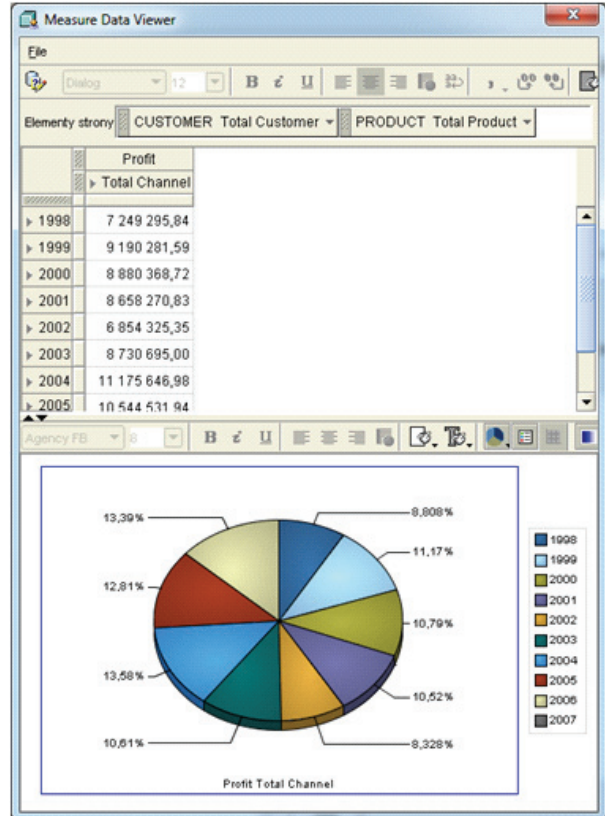


Fig. 18. Sample data from OLAP cubes (pie chart)

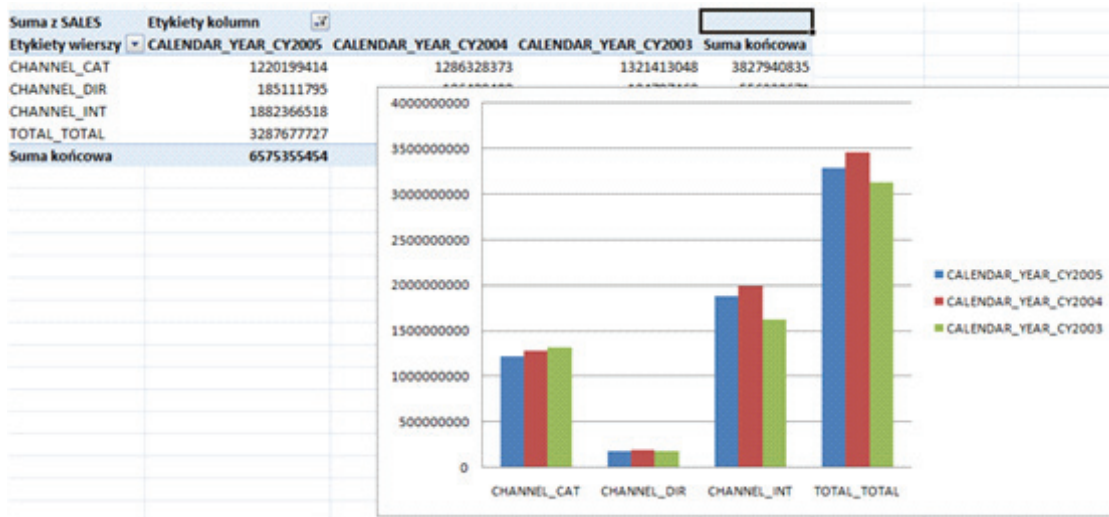


Fig. 19. Sample data from OLAP cubes in Microsoft Excel.

Summary

The paper presents modern techniques of data modeling and processing, collected by the company. It presents the process of multidimensional data modeling (include the transformation of logical structure to relational structure and next transformation to Multidimensional Model of warehouse in star or snowflake schema). It also shows the ETL process and methods of creating OLAP cubes by use of ORACLE tools to support decision making by business analysts. An approach based on data mining techniques allows analysts to capture certain features in customers, to offer dedicated products for the customer groups. Based on customer behavior can be concluded about his tendencies to their certain behaviors and preferences.

REFERENCES

- [1] Oracle OLAP - User's Guide 11g. Release 1 (11.1). B28124-03
- [2] Oracle Database - Data Warehousing Guide 11g. Release 1 (11.1). B28313-02
- [3] Oracle Warehouse Builder - User's Guide 11g. Release 1 (11.1). B31278-06
- [4] Oracle SQL Developer - User's Guide Release 3.0. E17472-05
- [5] Oracle SQL Developer - Data Modeler User's Guide. Release 3.0. E15802-03

- [6] Chuck Ballard, Dirk Herreman, Don Schau, Rhonda Bell, Eunsang Kim, Ann Valencic. Data Modeling Techniques for Data Warehousing. IBM - International Technical Support Organization - 1998. <http://www.redbooks.ibm.com>
- [7] Smarter Planet IBM: <http://www.ibm.com/smarterplanet/us/en/overview/ideas/>
- [8] Mendrala D., Szeliga M. Serwer SQL 2008. Usługi biznesowe. Analiza i eksploracja danych, Helion 2009, Gliwice.
- [9] Chuck Ballard, Daniel M. Farrell, Amit Gupta, Carlos Mazuela, Stanislav Vohnik. Dimensional Modeling: In a Business Intelligence Environment. IBM - International Technical Support Organization - 2006. <http://www.redbooks.ibm.com>

Authors: prof. dr hab. inż. Sławomir Wiak, Technical University of Lodz, Institute of Mechatronics and Information Systems, ul. Stefanowskiego 18/22, 90-924 Łódź, E-mail: slawomir.wiak@p.lodz.pl; dr inż. Paweł Drzymala, Technical University of Lodz, Institute of Mechatronics and Information Systems, Stefanowskiego 18/22, 90-924 Łódź, E-mail: pawel.drzymala@p.lodz.pl; dr inż. Henryk Welfle, Technical University of Lodz, Institute of Mechatronics and Information Systems, Stefanowskiego 18/22, 90-924 Łódź, E-mail: henryk.welfle@p.lodz.pl.