Zygmunt Ryznar

DATA WAREHOUSING AND DATA MINING IN BANKING

Introductory Remarks

The classical stage of operational service has come to an end. Multichanneling within electronic channels is now mainstream. In addition, there are changes in the structure of services and fundraising, such as speculative financial instruments and securities exchanges. Banks focus their efforts on reducing the potential financial risks inherent to the globalization of financial markets. This requires banking personnel with specialized knowledge as well as adequate information support, mainly through data warehousing and data mining tools or intelligent systems.

Data warehouses can play an important role in supporting the banking executives at the strategic level (board of directors) and tactical level (directors of departments and branches). Intelligent data mining tools can be useful also in operational activities (e.g. credit risk assessment and detection of abnormal behavior of customers using stolen or counterfeit cards).

Defining the informational needs to be supported by data warehouses is a complex task for banking, usually performed during the design phase called "business discovery". Covering the most essential and “critical” or urgent requirements is one of the basic conditions for the success of implementation a data warehouse.

A data warehouse is a centralized non-transactional database for storing information, usually for the long term, in a specific analytical manner. “Analytical” does not mean that it stores elementary information, but as an analysis tool, it can - using OLAP (On Line Analytical Processing) - access databases at different analytical levels, using techniques such as a deep "drill down" e.g. starting with the bank's cumulative yearly totals, then at quarterly or monthly periods for each bank branch, inside a branch of a particular class of banking products, etc. The OLAP has its own specific data access mechanisms (ROLAP – Relational OLAP, MOLAP – Multidimensional OLAP).

Separate user groups (e.g. departments, task groups) may use two kinds of datamarts (Dms): dependent containing information extracted from the global data warehouse or independent loaded with extracts from the operational databases. DMs provide a better user experience (such as navigation of familiar information resources and faster response times) and are used to reduce costs (smaller computers) or to minimize data traffic over the network.

The global DW contains company-wide information grouped by such entities as the customer and product, and further grouped by similarity and time (details for the most recent 2-3 years, summaries for many years, etc.).

Knowledge databases and data warehousing create favorable conditions for intelligent data mining aimed at acquiring knowledge hidden in large amounts of data. Data mining technology combined with ROLAP ("star" and "snowflake" schemes) and MOLAP (multidimensional cubes built-in database) demonstrates how to move from the classical ERA (Entities, Relations and Attributes) data modeling to intelligent modeling.

According to the interpretation of the Gartner Group, "Data mining is the process of discovering new correlations, patterns and trends from large volumes of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques." From a formal technology viewpoint, "data-mining" tends to produce information such as classes, clusters (categories) as subsets for association (association events), sequences of events, similar sequences of events, ...and so on. Advanced data-mining occurs in intelligent systems, which are used in poorly defined, undefined or chaotic environments. These systems use applied learning techniques such as heuristic reasoning sequences, expert systems and neural networks.

Architecture of Data Warehouse System in banking

Banks require a variety of solutions ranging from a fully centralized global data warehouse (DW), to not integrated datamarts (DM) for the emerging needs of the departments. Both of these extreme solutions have significant disadvantages along with their benefits, so it is worth considering the indirect variant presented here based on a global data source ODS (Operational Data Store), a global DW and DMs.

Due to the limited space assigned to this paper we will briefly describe only some of the components.

ODS collects data from various systems, cleaning and restructuring the data before loading it into the warehouse. Data structures and rules of purification and transformation are contained in the metadata repository.

Infoservices DM is designed to provide an "early warning" by monitoring critical areas.

BSC DW brings together a variety of information and is needed to run the business management strategy as a whole, providing the necessary tools to analyze the impact of various factors on the global results, evaluating the development potential, etc.

CR-DW is intended primarily for automatic segmentation of customers, determining product baskets, measuring customer loyalty and risk, selecting customers for direct marketing, etc.

CF-DW - based on the ODS warehouse, generates expected operating cash flow based on individual transactions and contracts.

Concluding remarks

Data warehouses are costly and complex undertakings with the primary purpose of supporting the management. Development should be determined by the rules of efficient business, not the ambitions of technology personnel. Essential to success is the formulation of business goals and information needs, the quality of the input data and the proper use of the potential of modern technology.

Exemplary schema of Data Warehousing System