

The Role of Data warehousing in Educational Data Analysis

Shirin Mirabedini

Department of Computer Engineering, Payame Noor University, Tehran, Iran

Corresponding author: Shirin Mirabedini

ABSTRACT: In today's world where a large amount of data centers and educational institutions on-line transaction processing systems are produced and rapidly changing and competitive environment, many of these centers are looking for an efficient tool which are able to move quickly to the high volume of data collection and analysis, convert it into useful information to support sound decisions. The most important criteria for identifying successful and effective leaders in the field of education is something that comes from the decisions of the directors. Collect and integrate data, analyze the information and pass it to the training centers and academic decisions affecting the other hand, it is possible to investigate the factors influencing students' success and success in attracting students and the optimal resource allocation decisions in order to increase productivity. Business Intelligence and Data Warehousing strategy is an effective way in order to overcome the problems of existing systems are on college campuses. In this paper, research on business intelligence, we will try to explore how to build a data warehouse and online analytical processing tools (OLAP) to analyze data in line with the academic community.

Keywords: on-line analytical processing, data cube, business intelligence, analytic database, data warehouse.

INTRODUCTION

In a successful education system takes all decisions made based on the data and data analysis methods have played a major role in deciding the quality of these systems. Currently, the data warehouse with on-line analytical processing tools for database analysis is one of the best ways to collect and analyze information on education systems have been proposed and the main aim is to support decision making in such systems. Actually Analytic Database is a collection of subject-oriented, integrated, based on different time intervals (time-varying) and immutable (non-volatile), which is used for management and decision support. Over a relatively long time Management Information Systems (MIS) major role in the management of training centers were responsible for tasks in the same way as the tool of decision support systems (DSS), Expert System (Expert System), Executive Information Systems (EIS) Ltd. came into existence to help senior executives, but now with the changing competitive conditions, these systems are able to meet and support the new educational demands. Some of today's functional requirements can be named as follows:

- Perform fast and efficient decisions under competitive pressures.
- Monitoring the competitive environment
- Data analysis and feasibility assessment of various aspects of educational centers from different perspectives
- Collect and analyze data from all the online resources and use a variety of techniques, data mining and data visualization

According to the needs of today's conventional systems are not dissimilar and different complexes and their inability to find hidden relationships among data and other shortcomings of today's systems. In fact, the centers and educational institutions to be able to quickly and accurately react to the changes in business areas and the right

management decisions need to be taken in ways that have collected data from internal and external sources and is able to analyze them. In this article we learn how to set up a database analysis to evaluate college.

Literature Review

Business Intelligence Strategy:

A business intelligence strategy, a package of computer application that supports different aspects of information needs. The package is marked as in Figure 1, including transaction processing systems, data warehouse, Data Mart, a report generation tool, online analytical processing systems (OLAP) and data mining is elsewhere.

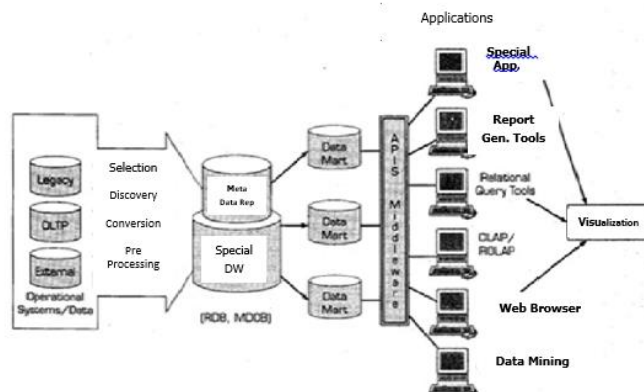


Figure 1. A strategy for business intelligence

Educational systems today, on the one hand to improve the quality and efficiency are faced with new challenges and the needs of the changing economy and other domestic and regional markets are working. This requires establishing and institutionalizing a comprehensive decision support systems and decentralized training.

Decision Support Systems:

Early 1970s, Scott – Morton for first important concept to build decision support systems, his decision support systems can be defined as a computer-based interactive system to decision makers in analyzing data and helps use models to solve problems of unstructured. According to another definition of the Scout - Morton (1978) decision support system is a system that aimed at improving the quality of decision making, human resources, intellectual capabilities that links computers. Such systems play an important role in gathering useful information from raw data, documents, personal knowledge and problem solving to play business models. For such systems, strategy, business intelligence and data warehouse mechanisms is essential.

Data Warehouse:

Since the mid-1980s there was a need to store data and found that software systems should be separated into two parts, operating systems and information systems that will support the daily activities of business operating systems and rapid response to real-time communication and presentation of pre-defined conditions are appropriate current and Information Systems to manage and control the business and data analysis are used to make decisions on the spot and futures firm performance. The next task is to work with the database, and add applications and data warehouses, data were presented that were in charge of accelerating the data, and reports of decision-making are effective. The main purpose of data warehouses provide a suitable ground to produce the information that managers, executives and analysts to factor that helps correct decision.

As it is clear from the information that is stored in a data warehouse databases are kept of what is far more important and more valuable. All information contained in data warehouses organized the first ever to be kept united and organized. This repository of information from all sources and are collected in a long period and therefore the amount of data is enormous. In a data warehouse, the information must vary in a specific structure to store transaction processing environments , as well as the need to access information using suitable methods .For this purpose, data warehouse using multidimensional cube structure designs different data structure is created .For example, data stored in the warehouse for registration of the education system as shown in Figure 2, registered students , semester , course registration and tuition rates of the different aspects considered most cases , this have a hierarchical include semester Date and term of the school year.

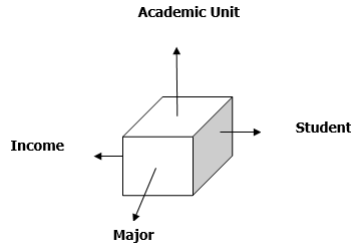


Figure 2. Cube of income enrollees' university education system

Yet through this Rubik's Cube various reports will be generated from many different perspectives. After determining the data storage requirements in the logic design data, information schema through three possibilities are the following:

Star schema:

Common schema, star schema where the warehouse with a central table called the fact table contains a large amount of data without repetition and smaller tables called dimension tables are defined.

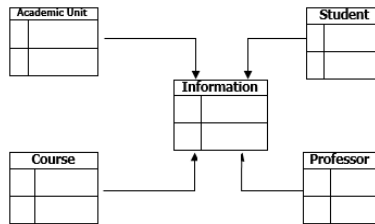


Figure 3. View of the education system using star schema

As part of the education system in Figure 3 is characterized by the shape of your face is a star, the truth table in its center and radius of each of the tables that are related by the following.

Snow flake schema:

The snow flake scheme, a type of star schema where some dimension tables are normalized. And therefore further divide the extra tables are separated from the next table.

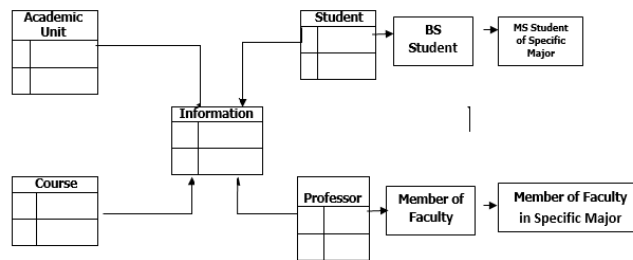


Figure 4. View of the education system using snow flake scheme

As is shown in Figure 4 star schema is a schema similar to the snow flake of the table is split into sub-tables. The difference is that your tables are normalized snow flake scheme and redundancy in them has decreased, this work is in the field of data storage space is useful, however, because this method requires more calculation join queries are confronted with reduced efficiency.

Galaxy schema:

In complex applications to share a dimension, it is necessary to require multiple fact tables. These guys set your star schema or a system is therefore called galaxy schema. This allows you to us the dimension tables are shared

between different facts tables. After creating a data warehouse logical design data aforementioned three methods to turn the physical design of a database management system is implemented. Data warehousing technology, a set of concepts and new tools that provide information to members of the Knowledge Workers in executive decisions, will help managers and analysts. After preparation of the data warehouse, data analysis, it appears that the turn to the next section for online analytical processing systems processing requirements of data warehouses is introduced.

Data Warehousing and On-line analytical processing (OLAP)

Various tools can be used for data storage, information processing. This is one of OLAP data analysis tools that are currently three different types of OLAP are defined as follows:

Multidimensional OLAP systems (MOLAP):

Because of the special structure of these systems store information faster than other models, and instead of being stored in the data record to the Keys, have designed specific databases as the data into a sorted array of store data. This system is a multi-dimensional tables with indexes and tables of actual data-intensive storage and complex analytical queries are executed in minimum time. This tool will speed up the database size is relatively small.

relational OLAP systems (ROLAP):

These data establish a framework to store and retrieve relational databases are based on the optimization of a relational database is based on the ground that ROLAP systems in the framework of the traditional model Relational OLAP operations are organized and therefore slower than MOLAP systems, need more space.

Hybrid OLAP systems (HOLAP):

The hybrid combination of ROLAP and MOLAP in order to have a large database and has higher efficiency than ROLAP. HOLAP systems are implemented in the form of interstitial portion of the data in the tables are maintenance intensive and take part in a relational space so have the benefits of both models. Online processing systems evolved from the online transaction processing systems (OLTP) are in line with the rapid processing of multidimensional data is used to analyze the data. Using the reporting system of education, student enrollment, finances and the like using a multidimensional data much faster than the system acts OLTP. Online processing systems using a hierarchical structure and a powerful multidimensional data model for organizing data, and reports to simplify calculations provide very fast. The data warehouse systems are associated with online processing of multi-dimensional data cubes that allow quick analysis provides information about the various databases. For example, multi-dimensional database systems as shown in Figure 5 can be annual tuition university education with courses Number of Courses and 'income and education in University of to the cross.

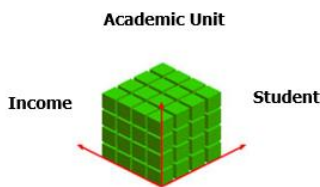


Figure 5. Part of the education system to calculate earnings per academic unit

This intersection is the result of a wide variety of reports, such as place of total revenue derived from fees for university students in a particular academic year, the highest paid and highest paid college course from the academic year of the system is easily. OLAP cubes using comprehensive data models, queries, efficient and fast data presentation and summary tables are the perfect alternative. Complex calculations in cubes are easily and improve the quality of training data analysis. For example, if an OLAP cube can include courses in each discipline and its comparison with the number of courses in the same semester of the previous academic year. The cube structure for storing information in various formats and produce analytical reports to analyze training data is set up. Using a relational database, analytical reports through the summary tables is very difficult to optimize while OLAP systems provide outputs in the form of columns and rows in a matrix of different dimensions and provide an output matrix.

performance of Data Warehousing

OLAP systems continuously copies of data and different data sources are summarized removed, they will sort the data cubes. User can run query on this cube. Complex queries on OLAP systems require much less time than a similar search on OLTP Systems. OLAP is a set of hypotheses generation and analysis, and communication parameters for the query, the database sends them to confirm or deny, in fact, on-line analytical processing system operations, data processing, and data mining mechanism to bring together the analytical database. While the differences are between the two instruments are complementary to each other. The use of a relational database application analysis and validation is faced with a set of queries against data in the form of a hypothesis. For example, an analyst may want to "factors that contributed to the success of an academic unit's inability" to analyze. In analyzing the data using the OLAP may be assumed that the factor "a small number of faculty members' academic unit confirms the hypothesis fails (or refuses). If this hypothesis was not authenticated by the data analyst may be a factor, "too many" students to look at as a result of risk factors. If this factor also may not approve of the "old and authentic college" is as best reflecting the risk of failure of the academic unit to be considered. In other words, OLAP, an analysis of a set of assumptions and parameters used for the communication with the database queries and a method to confirm or deny the process of inference is a necessity. However, data mining models faster confirms the hypothesis and using the same data to discover the patterns and mechanisms for dealing with processes of inference is necessary. For example, suppose someone is going to risk factors for the success of the academic unit using data mining and analysis to identify. Low number of factors may discover a data mining tool for faculty, students and old too low number of units, but it should be noted that this type of analysis is the possibility to ignore important and effective in some factors. As an example, the average age of faculty members and students could possibly be the decisive factors in the success of a college course. That's where the use of parallel data mining and OLAP can complement each other. OLAP carefully and focus on the important factors and interactions can identify exceptions to decipher.

An example of implementing Data Warehousing in education system of University

In this part of the article we analysis the process of creating a database instance to discuss with OLAP processing. The educational system, for example, the following information fields are available to students:

- Gender (male, female)
- Type a period (day, night, open ...)
- Name of School (Science, Engineering, Literature ...)
- Degree (BA, MA, PhD.)
- Input (901,902,911 ...)
- Degree (computers, literature, agriculture ...)
- Overview (studying, withdrawal, dismissal, change of course)

According to the requirements specified in this example, four reports based on students' grade point average and the number of research papers to separate the above fields is shown in Figure 6.

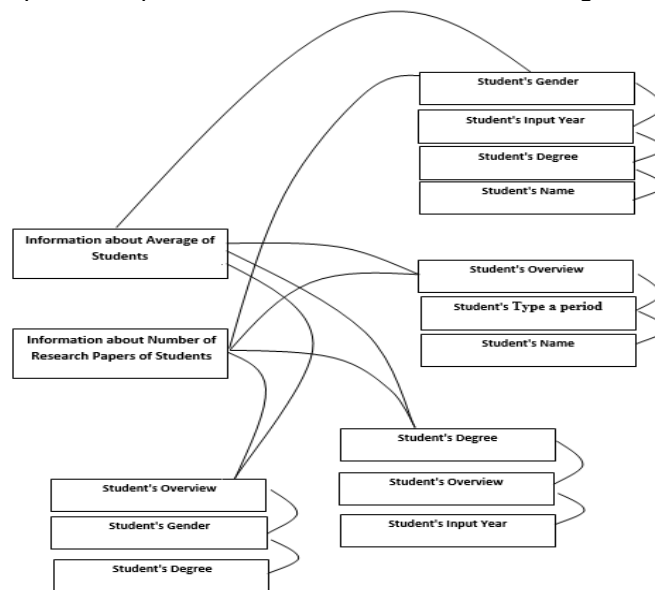


Figure 6. Example of a multi-dimensional reporting

The output of the logical design phase is an ERD diagram specifies the data storage structure. Figure 7 is an example of the typical structure of a data warehouse star schema template that has been created in order to produce the observed training reports. After determining the requirements for the rational design ERD diagrams must be drawn to determine what data is present in the system and can meet the user's needs and how the data should be stored in the data storage and the absence of some data fields, how to acquire them.

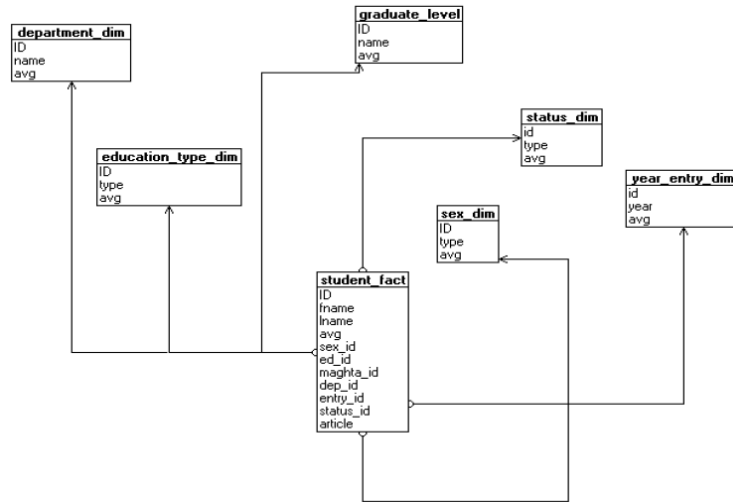


Figure 7. Diagram ERD related to the rational design

As seen in Figure 7 is a table outlining the facts in logic design which includes information such as the number of articles, name and foreign keys define its relationship with other aspects of tables. After this stage, the exact fields, turn to the physical design, structure, tables, constraints, keys and the like, and what data should be transferred to the structure of the data flow, it where have to be saved, and in what should be used to platform.

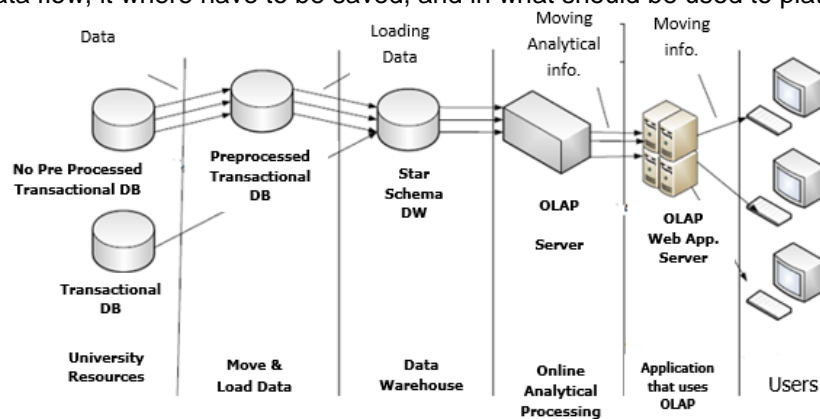


Figure 8. Data flow architecture

As is shown in Figure 8, after adjusting for the review and analysis of multidimensional data storage on data cubes OLAP tools are employed. Along OLAP tools is for users to visualize information at different levels of information when used according to requirements. The dashboard is also for users who need high levels of knowledge of the overall situation around the University of detail used for reporting.

CONCLUSION

The study how to make data storage and processing system that also pay online were made to collect data from various systems, such as students' education system and integrated them into the desired form provided to teachers, principals and other parties, provide learning centers. Using the integrated data as well as data warehousing and

online analytical processing application procedures, automatically or semi-automatically provide suggestions to improve teaching and learning for teachers and students. In this paper, we model the process of creating a data warehouse in academic environment and performance was researching online analytical processing tools that the data warehouse provides the ability to store data in an autonomous, analytical and operational data stored for each time and access to them should be completely independent from the operational database and so effective step to be taken in order to improve quality and productivity reports rapid analysis of educational systems.

REFERENCES

- Chen G, Liu C, Ou K and Liu B. 2005. "Discovering Decision Knowledge from Web Log, 2005." Portfolio for Managing Classroom Processes by Applying Decision Tree and Data Cube
- GISMO. 2007. <http://gismo.sourceforge.net/>
- MacLennan J, Tang ZH and Crivat B. 2009. "Data Mining With Microsoft Sql Server 2008",2009.
- Pendse N. 2008. " the OLAP report:market shareanalysis" ,2008.
- Polvises S. 2005. "Data warehousing with an OLAP system: a case study for thesis operation of the school of mahidol University", (2005).
- Romero C and Ventura S. 2008. "Data Mining Algorithms to Classify Students",2008.
- Thomsen E. 2002. " OLAP Solutions: Building Multidimensional Information Systems",2002.
- Zorrilla ME. 2009. "Data Warehouse Technology for E-Learning", Springer-Verlag Berlin Heidelberg 2009.