An Approach to Educational Perspective Management in Computer Training Center: DW and OLAP

May Cho Aung, Myint Thu Zar Tun

Computer University, Maubin

Abstract

Data warehousing and on-line processing (OLAP) are essential elements of decision support, which has increasingly become a focus of the educational database for management in computer training center. A data warehouse, a repository of long-term storage of data from multiple data sources, are organized sources as to facilitate educational perspective management decision making. It provides architecture and tools for private computer center's executives to systematically organize, understand, and use their data to make a strategic approach to educational perspective management. This system supplies the retrieval of educational perspective management concern with the query results from the data warehouse of computer training center information. This system produces a variety of information for students and staffs by using OLAP operations (*roll-up, drill-down, dice, slice) from the data warehouse.*

Keywords: query analysis, management decision making, data warehouse, multiple data sources, OLAP operations

1. Introduction

The educational perspective management in Computer Training Center can be seen from two different analyses. It can mean the use of educational perspective management in commercial enterprises which fosters learning environment and seeks opportunities how educational perspective management can be useful or can complement to knowledge management activities or activities of learning and teaching organization. Second analysis is focused on the use of educational perspective management in all private and public schools and universities. The educational management represents opportunities for the businesses and for the education institutions. Both analyses will be covered since they are very close and may be encouraged for cooperation. However, only in the second analysis, the particular organization will be targeted.

In today, competitive and global would the private and public schools and universities which have to behave like businesses, e.g., private computer training center. One of the tasks of managements of private computer training center is to assess how the center processes are going. This system stores data about the student's progress in an electronic way. Hence, there is no problem to do reporting of e.g., what is the number of students that enroll in private computer training center, and describe the state class this year, what are the teachers that teach in the class, etc... Internally, the private computer center might create reports that show trends in number of students that completed successfully or finished unsuccessfully the studies; number of students applying for some specialization or subject; number of students actually enrolled in a particular school year, etc. Preparing such reports and analyzing such trends enables private computer center to judge rising or falling interest of some specifilizations and might help to decide about the initiation of the process of accreditation of new class to the next year. Such analysis serves for strategic planning of the private computer center and also can be used for comparing the quality of education in different private computer sub branches.

Therefore, management capacity is the process of monitoring and projecting private computer center workload and specifying the most cost-effective computing environment that meets the current and future plan for computer center services. The growing complexity of database system stimulates the development of software tools to probe and predict system performance and to help in system design. This database system is called data warehouse system which is preceded by a highly complex ETL (extract, transform, load) process of integrating the data from multiple systems and bringing it into a consistent state. In recognizing this database, relational online analytical processing (OLAP)

is structured using a star schema consisting of fact tables and dimension hierarchies. [4]

In this system, decision-support functions in a data warehouse, including OLAP, involve hundreds of complex aggregate queries over large volumes of data. It is not feasible to compute these queries by scanning the data set each time. In this paper, the OLAP operations are applied to analyze the data of private training computer center in data warehouse system.

The rest of the paper is organized as follows. Section 2 reviews the concepts of DW and OLAP operations. Section 3 present the system design creation and database construction for computer training center. In Section 4, we apply and implement the OLAP operations by using data warehouse. Section 5 gives the conclusion of this paper.

2. Concepts of DW and OLAP

Data Warehousing is a collection of decision support technologies, aimed at enabling the knowledge worker (executive, manager, analyst) to make better and faster decisions. A data warehouse is a "subject-oriented, integrated, time-varying, non-volatile collection of data that is used primarily in organization decision making". Typically, it is maintained separately from the organization's operational databases. [1]

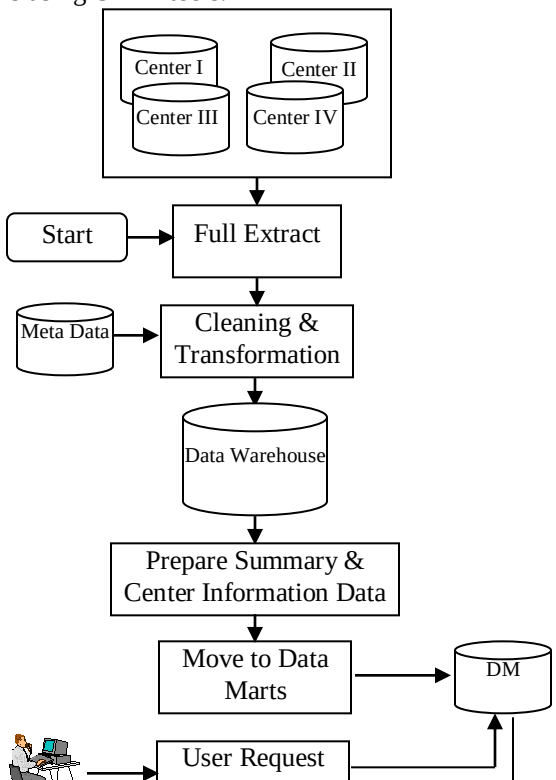
Data warehouse also integrate massive amount of data multiple sources, which is primarily used for decision support purposes. It can store and integrate historical information and support complex multidimensional queries. Data warehouse can be stored in many different types of data bases. Thus, a data warehouse is designed for the query and analysis is rather than transaction processing. It is a centralized repository of all corporate data. Data warehouse provide mangers with information to support the specific decision.

The data warehouse supports on-line analytical processing (OLAP), that are the bread-and-butter day-to-day operations of an organization. Since data warehouses contain consolidated data, perhaps from several operational databases, over potentially long periods of time. The workloads are query intensive with mostly ad hoc, complex queries that can access million of records and perform a lot of scans, joins and aggregate. Query throughput and response times are more important than transaction throughput.

To facilitate complex analysis and visualization, the data in a data warehouse is typically modeled multidimensionally. Often, these dimensions are hierarchical; time of student and staff records may be organized as a day-month-quarter-year hierarchy, centers as a centers-category-sub branch hierarchy. Typical OLAP operations include rollup (increasing the level of aggregation) and drill-down (decreasing the level of aggregation or increasing detail) along one or more dimension hierarchies, slice-and-dice (selection and projection). This paper presents data warehousing with OLAP operations, focusing on the special requirements that data warehouses place on database management systems (DBMSs).

3. Creation of System Design and Database

This system may be created for load balancing, scalability, and higher availability. In such a creation system design, the metadata repository is usually replicated with each fragment of the warehouse, and the entire warehouse is administered centrally. This system uses a variety of full extraction of data, cleaning, loading and transforming data with metadata utilities. After draw to a close transform data, the query engines for OLAP are the multidimensional view of data in the warehouse. Figure1 shows that system design is created for education management of computer training center where data are loaded into the warehouse through ETL processes from operation data stores. Then summarized data are prepared in the data marts and analysis can make query analysis using OLAP tools.



This system includes tools for extracting data from multiple operational databases of four centers such as centerI, centerII, centerIII and centerIV; for loading data into the data warehouse; for cleaning, transforming and integrating these data; and for periodically refreshing the warehouse to refresh updates at the sources and to remove data from the warehouse, perhaps onto slower archival storage. These centers contains enroll data, attend data, staff data, and class data. In enroll data, there are Student_ID, Name, Age, F/M, Class, HigherEducation, StartDate, EndDate, Address, PhNo, PaySalary. Attend_ID, StudentName, ClassName, StartDate, EndDate, Present, Absent, Total, Project and Certificate are contained in attend data. As well, Class_ID, Class, NoofStudents, NumberofClass, OpenDate, CloseDate, Duration, Fees and Project are included in class data. In Staff Data, Teacher_ID, Name, Position, Lecture 1, Lecture 2, Lecture 3, Salary, Address and PhNo are restrained. After these data are loaded, the data mart preparation process starts to make data mart for future analytical reports. This data mart process are used to select the desire field name, field type and field length for building the data warehouse and querying process in OLAP. The summary and centers data marts are prepared for later OLAP processes. Most of the queries over a star schema can be transformed into *N* independent partial queries as shown in Figure 2. In these queries, the computing of the partial results in each database of any center in four centers also needs to access data stored in data warehouse. In this case, the execution of queries is dependent of partial results which are calculated over the data in data warehouse system.

In star schema, there is a set of numeric measures that are the objects of analysis; e.g., a training center star schema. Each of numeric of any center in four centers measures depends on a set of dimensions, which provide the context for the measure. Each dimension of any center in four centers is represented by only one table and each table contains set of attributes. Each table has used one dimensional structure. Training center schema is considered along four dimensions, namely enroll, attend, class and staff. The dimensions together are assumed to uniquely determine the measure. Each dimension is described by a set of attributes. These attributes of a dimension may be related via a hierarchy of relationships. [3] In enroll table, there are six attributes: namely, Student_ID, Name, Age, Class, Higher Edu:, and StartDate. There are also five attributes in attend table such as Attend_ID, Name, Class, Absent and Certificate. The class table consists of seven attributes; class_ID, class, NoofClass, OpenDate, CloseDate, Fees and Project. In staff table, Staff_ID, Name, Post and Salary are described. The schema contains a central fact table for training center that contains the foreign keys to each of the four dimensions, along with measure; sub_branch. The fact table contains dimension identifiers such as Student_ID, Attend_ID, Class_ID and Staff_ID which are system-generated identifiers.

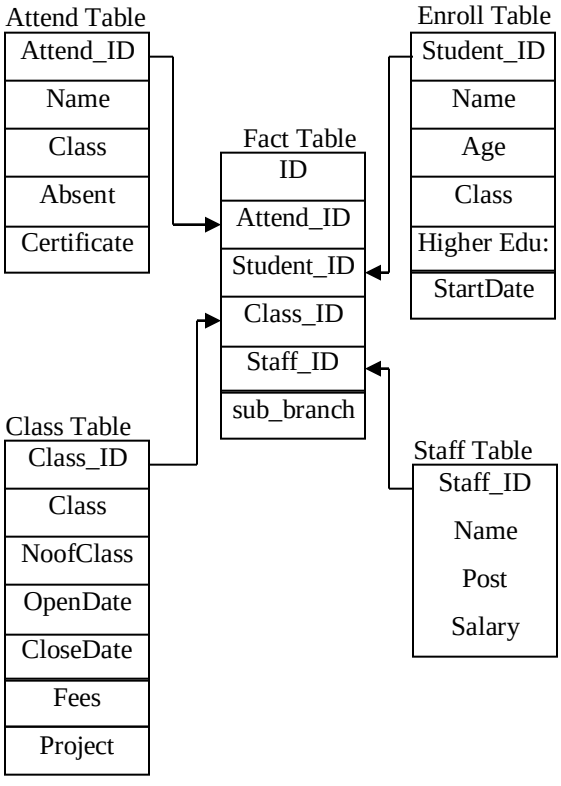


Figure 2. Database Design of the System

4. Implementation of OLAP Operations with Data Warehouse

An OLAP system manages large amounts of historical data that provides facilities for

summarization and aggregation, and stores and manages information at different levels of granularity. Data warehouse and OLAP tools are based on multidimensional data model. Typical OLAP operations in multidimensional data model are roll-up, drill down, slice and dice. This system presents to develop OLAP functions query analysis for computer training center's information data warehouses also implements educational perspective management affect with students and staffs. [2]

In this system, roll-up performs on a data cube based on the four dimensions, namely, enroll table, attend table, class table and staff table, either by climbing up a concept hierarchy for a dimension or by dimension reduction as shown in Figure 3.

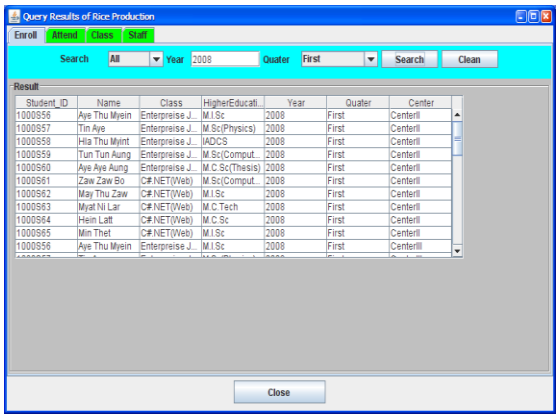


Figure 3. Roll-up Operation in Quarter

Rollup corresponds to taking the current data object and doing a further group-by on one of the dimensions. Thus, it is possible to roll-up the enroll data, perhaps already aggregated on each center. When the manager can search all centers, the system shows quarter in year by using roll-up operation.

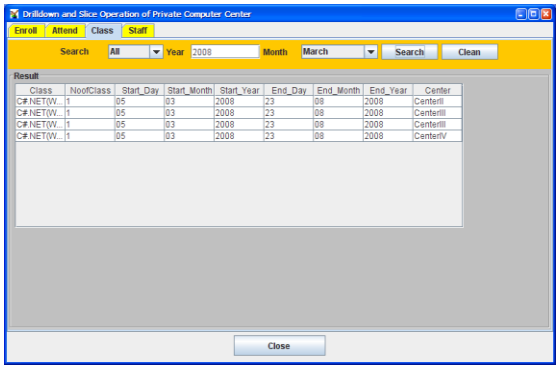


Figure 4. Drill-down Operation in Monthly

Drill-down is the reverse of roll-up. It navigates from less detailed data to data that more detailed. In this system, if the manager selects all centers, all months for any year period from the raw data of database, he can see drill-down operation as shown in Figure 4.

Slice and dice corresponds to reducing the dimensionally of the data, i.e., taking a projection of the data on a subject of dimensions for selected values of the other dimensions. Dice operation defines a subcube by performing a selection or two or more dimensions and navigates from less detailed data to data that are more detailed. Also slice operation performs a selection of one dimension of the given cube, resulting in a subcube. User can select only one attribute for slice operation as shown in Figure 5.

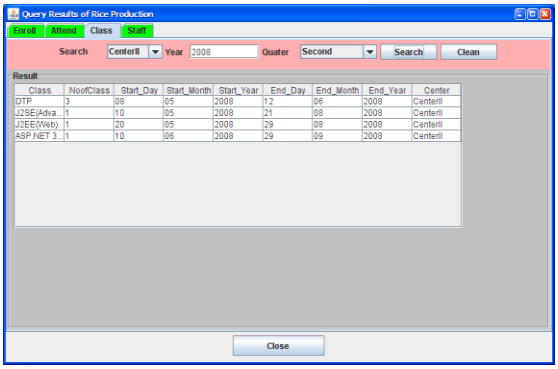


Figure 5. Slice Operation by Each Branch

5. Conclusion

Data warehousing and OLAP operations are essential elements of decision support, which are increasingly become a focus of the database industry. This system is easier, on a regular basis, to query and report an approach from multiple transaction processing systems, from external data sources, and from data that must be store for query purpose only. It also improves the performance of educational perspective management and the advisement of decision making. In this system, we also implement to know the private computer center of branch management for each center and describe the supporting of distributed query processing and analysis from data warehouse.

References

[1]Inmon, W.H., "Building the Data Warehouse"

John Wiley, 1992.

[2]Codd, E.F., S.B. Codd, C.T. Salley, “Providing OLAP (On-Line Analytical Processing) to User Analyst: An IT Mandate.” Available from Arbor Software’s web site http://www.arborsoft.com/OLAP.html

[3]Chatziantoniou D., Ross K. “Querying Multiple Features in Relational Databases” *Proc. of VLDB Conf.*, 1996.

[4]Widom, J. “Research Problems in Data Warehousing.” *Proc. 4th Intl. CIKM Conf.*, 1995.