Implementation of a University Decision Support System using OLAP

Phyu Sin Phwe, Nan Sai Moon Khan
University of Computer Studies, Yangon
psinphwe@gmail.com

Abstract

Data Warehouse and OLAP processing is one of the powerful tools for knowledge discovery in large complex data. Data warehouse provides online analytical processing (OLAP) tools for interactive analysis of multidimensional data which facilitates effective data mining. In this paper, the data warehouse and OLAP technology are used for analysis of students and staffs' data of all Computer Universities in Myanmar. This system can extract specific information of teachers, students, and staffs based on the user request and also produce decision support forms comparing the numbers of these staffs varying with time. This system is implemented by Microsoft C#.Net Programming language.

1. Introduction

The data warehouse is maintained separately from the organization's operational database. Data warehouse are targeted for decision support and historical, summarized and consolidated data is more important than detailed, individual records [4]. The data in a warehouse is typically modeled multidimensionality to facilitate complex analysis and visualization. Due to a lot of tasks, there may be some mistakes. To avoid such situations, a control system is needed. This system uses data warehousing techniques and OLAP tools to deal efficiently with very large set of students and staffs data in all Computer Universities in Myanmar. The primary objective of this system is to implement the data warehouse concerned with students and staffs data by using OLAP technology. This system will also give easy to use and searches based on the user request and implement an OLAP tools to generate decision support system. In this system, users can make different and various kinds of analysis queries and can view the student and staff data forms in different aspects.

2. Background Theory

2.1. Data Warehouse

A Data Warehouse is a subject-oriented, integrated, time variant and non-volatile collection of data in support of management's decision-making process [4]. It is a decision support database that is maintained separately from an organization's operational database. This database is designed based on what kind of information is important in company's decision making process (e.g. sales marketing, health care membership and providers etc.) and it adopts a specialized schema design for maximum efficiency in performance.

Generally a data warehouse is populated by taking data from various legacy sources. Extracting appropriate data from existing operational databases, cleansing or scrubbing the data, demoralizing the data and then loading the data into the database are the typical steps involved in creating a data warehouse. (This data population process is also known as the data transformation process [5].) The data warehouse is shown in Figure 2.1 [2].

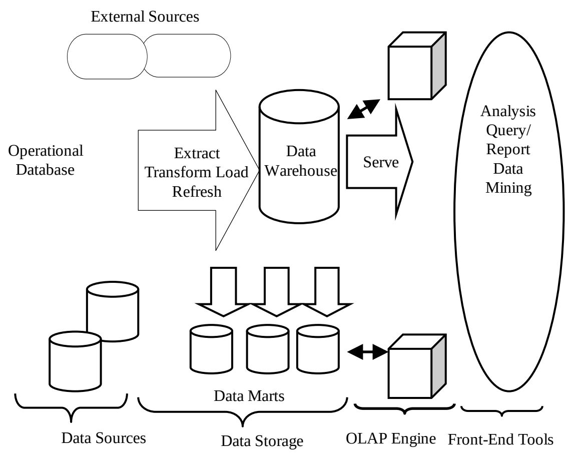

Figure 2.1 Data Warehouse Architecture

2.2. Online Analytical Processing (OLAP)

In the multidimensional model, data are organized into multiple dimensions, each dimension contains multiple levels of abstraction defined by concept hierarchies and provides users with the flexibility to view data from different perspectives. A number of OLAP data cube operations exist to materialize these different views, allowing interactive querying and analysis of the data at hand. OLAP and Data Warehouses are complementary [6]. A Data Warehouse stores and manages data and OLAP transforms Data Warehouse data into strategic information [3].

2.3. OLAP Operations

There are four kinds of OLAP operations:

Roll-up: The roll-up operation also called the drill-up operation performs aggregation on a data cube, either by climbing up a concept hierarchy for a dimension or by dimension reduction.

Drill-down: Drill-down is the reverse of roll-up. It navigates from less detailed data to more detailed-data. Drill-down can be realized by either stepping down a concept hierarchy for a dimension or introducing additional dimensions.

Slice and dice: The slice operation performs a selection on one dimension of the given cube, resulting in a subcube. The dice operation defines a subcube by performing a selection on two or more dimensions.

Pivot(rotate): Pivot (also called rotate) is a visualization operation that rotates the data axes in view in order to provide an alternative presentation of the data[1].

2.4. ROLAP vs. MOLAP

Currently there are two dominant OLAP architectures. They are Relational OLAP (ROLAP) and multidimensional OLAP (MOLAP). ROLAP techniques store data in tables in a traditional RDBMS and provide sophisticated front end to support the complex functions needed by OLAP users. MOLAP techniques on the other hand, provide custom OLAP servers which typically store data as multidimensional arrays or custom storage structures specifically designed for efficient storage and retrieval of sparse multidimensional arrays.

2.5. OLAP Cube

An OLAP cube is a data structure that allows fast analysis of data. OLAP cubes can be thought of as extensions to the two-dimensional array of a spreadsheet. The OLAP cube consists of numeric facts called measures which are categorized by dimensions. Measures are derived from the records in the fact table and dimensions are derived from the dimension tables. Sample OLAP cube is shown in Figure 2.2.

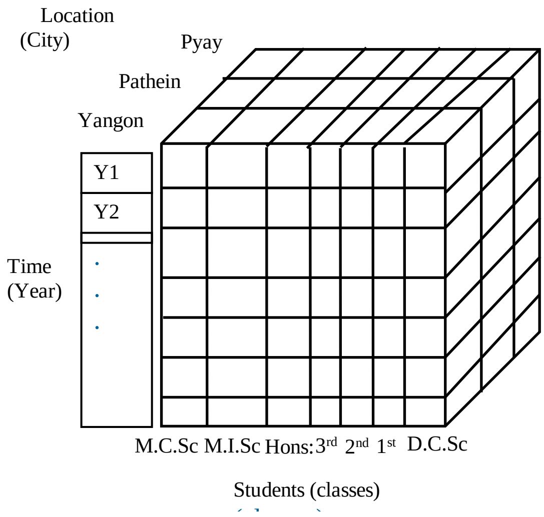

Figure 2.2 Sample OLAP Cube

2.6. Star Schema

Star schema has a single object in the middle radically connected to other surrounding objects like a star. The table in the middle is called “fact table”, which contains compound primary key, with one segment for each “dimension”, and additional columns of additive numeric facts. There will be a single “dimension” table (for each dimension) with a generated key and a level indicator that describes the attribute level of each record.

Since star schema will have only one key from each dimension generating a key for a dimension is necessary. One approach is to form a compound key by concatenating all the keys into a single key but the size of the key will be large and nulls can be confusing.

Star schema is easy to understand as data is organized around subjects.

2.7. Snowflake Schema

The snowflake schema is an extension of the star schema where each point of the star explodes into more points. In this schema, the dimension tables are more normalized as compared to the star schema. The advantages provided by the snowflake schema are improvements in query performance due to minimized disk storage for the data and improved performance by joining smaller normalized tables rather than large demoralized ones. Many consistency problems (due to updates) are also solved because of normalization.

3. Implementation

This system uses data warehousing techniques and OLAP tools to deal efficiently with very large set of students and staff data in all Computer Universities in Myanmar. This system can support various queries based on the user request and users can extract the necessary portion of the data easily and quickly.

3.1. Design Schema for the System

"Star Schema" is used as the design schema for the system. Figure 3.1 shows star schema for Student and Staff.

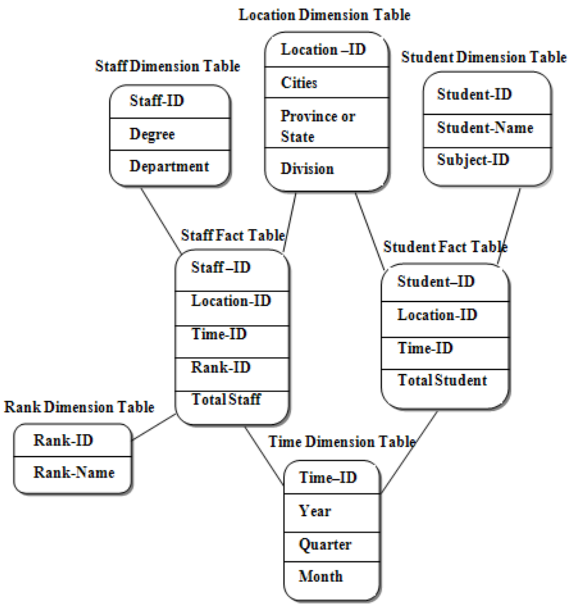

Figure 3.1 Star Schema for Student and Staff

The tuples of whole database is linked by key with each other. When the user wants to extract specific information from historical database, he/she requests to link. Each tuple extract required information form each tables and displays to the user.

3.2. Overall System Design

In this system, user can firstly choose the types of information that they want to view (whether teacher-related information or student or staff) based on location, subtype (means that rector or professor or so on for teacher, major or year or academic year for student and rank for staff).

User can also make various queries such as extracting data based on specific name, year, location, and department and so on. Like this, user can view desired students form based on specific academic year, roll number, major, class, location and so on. And for staff, staff data based on specific department, name and location can also be extracted and viewed.

Figure 3.2 shows the overview design of the system.

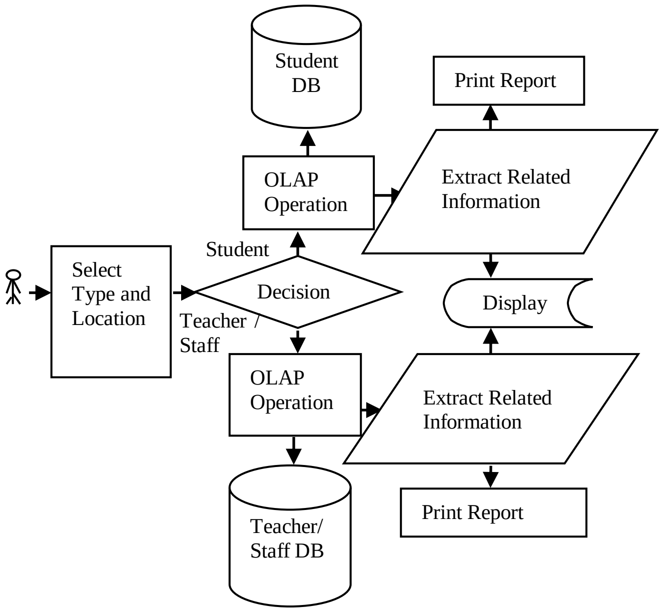

Figure 3.2 Overview Design of the System

The system performs these operations by using related databases such as "Student DB" and "Teacher/Staff DB". And the system performs four OLAP operations slice/dice, roll-up, pivot and drill-down based on the user's chosen information.

The system can also produce the final report (the changing numbers of teacher, student and staff) for the years 2005, 2006, 2007 and 2008 in graph. These reports show the comparison of the varying numbers of teacher, student and staff.

Figure 3.3 shows the report for student list.

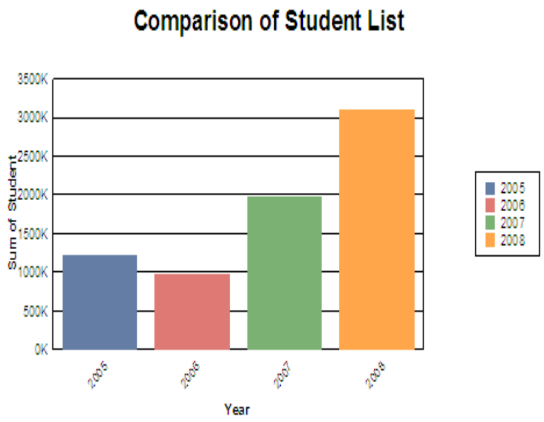

Figure 3.3 Report for Student List

Table 3.1 shows the list of students in all Computer universities for the year 2008.

Location	Year	M.C.Tech	M.C.Sc	…	Module
UCSY	2008	110	125	…	-
UCSM	2008	35	50	…	11
LaShoe	2008	3	3	…	15
Bamaw	2008	3	3	…	10
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
MoneYwa	2008	6	15	…	8
Kalay	2008	4	6	…	6

Table 3.1 Students list for the year 2008

This paper can facilitate to find out the required information based on the user request. In this sample OLAP operation, users can make different and various kinds of queries and can view the teachers' information list.

Figure 3.4 shows the information page to select required information.

Figure 3.4 User Select Required Information Page

Figure 3.5 shows the list of required teachers' information.

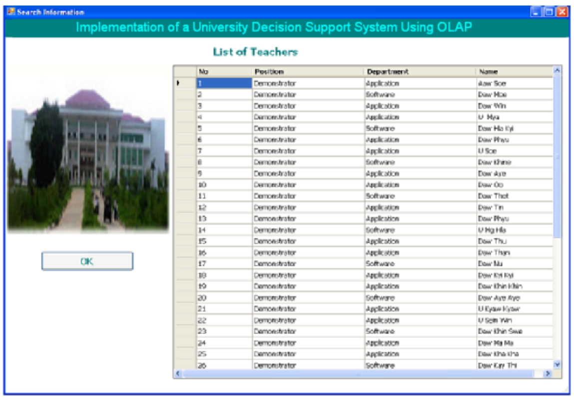

Figure 3.5 List of Required Teachers' Information Page

4. Conclusion

Data warehousing is now one of the hottest topics in the information systems. The goal of the data warehouse used in this system is to provide any university with an easy and quick look at its historical data. Advanced On-Line Analytical Processing (OLAP) tools let Data Warehouse (DW) users generate reports at a click of a mouse and look the university information from various angles. The data warehouse based on extended database technology provides the management to the data store. In this paper, it is implemented for the management of Computer University information to find the complex queries. This data mart can be easily retrieved the analysis data efficiently via on-line analytical processing.

Data warehouse used in this system provides architectures and tools for University executives to systematically organize, understand, and use their data to make strategic decision. In this system, the data for teachers, students and staffs can be extracted as necessary. In this way, the time can be saved. And the complexity of dealing with very large data sets can be reduced.

5. References

[1] J. Han,.M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, A Harcourt Science and Technology Company, 525 B Street, Suite 1900, San Diego, CA 92101-4495, USA.

[2] P. Fazli, "An Overview of Data Warehousing and OLAP Technology", University of British Columbia.

[3] P. Vassiliadis, Y. Song, "Report on the 9th International Workshop on Data Warehousing and OLAP (DOLAP 2006)", Vol. 41 No. 1 June 2007, ACM SIGIR Forum.

[4] S. Chaudhuri, U. Dayal, "An Overview of Data Warehousing and OLAP Technology", ACM Sigmod Record, March 1997.

[5] Wu, M-C., A.P. Buchmann. "Research Issues in Data Warehousing." Submitted for publication.

[6] http:// www. olapcouncil.org