Decision Making for Staff Information using Data Warehouse and OLAP

Kyi Pyar
University of Computer Studies, (Mandalay)
kyipyar.IT@gmail.com

Abstract

In the area of information and communication technology, a wide variety of methods and tools are used for storing, processing and analyzing data. The paper is one in a series describing reference configuration and sizing information for small, medium, and large solutions to meet business data warehouse needs. This system attempts to study OLAP queries which are in terms of a multidimensional data mode. According to the respective roles and privileges, OLAP queries can be retrieved from the data warehouse. The system analyses staff data for technological universities in Myanmar. This system is implemented by using J Builder 9 language and Microsoft Access.

Keywords- Data warehouse, Star schema, and OLAP;

1. Introduction

Data mining process is efficient only the presence of summarized data that is stored in data warehouse [10]. A data warehouse is a semantically consistent data store that serves as a physical implementation of decision support data model and stores information on which an enterprise needs to make strategic decisions. An online analytical processing (OLAP) system manages large amounts of historical data [1].

Organization's decision making requires a comprehensive view of all aspects of enterprise and many organizations have therefore created consolidated data warehouses that contain data drawn from several databases maintained by different business unit; together with historical and summary information. Organizations need to turn their archives of data into a source of knowledge, so that a single consolidated view of the organization's data is presented to the user.

Data Warehouse is a "subject-oriented, integrated, time-variant and non-volatile collection of data that is used primarily in organizational decision making." Typically, the data warehouse is maintained separately from the organization's operational databases. [9].

This paper is organized as follows. Section 2 is introduced the related work of supervised learning. In section 3, we discuss the background theory. Section 4 is represented the system design. The implementation results of the system are described in section 5. Section 6 is presented the conclusion.

2. Related work

Kimball [5] presented a data warehouse is a database that stores information oriented to satisfy decision making requests. The features of data warehouses cause the ones for OLTP's systems. Matrotta [7] presented a set of schema transformation primitives for relational data warehouse design.

E.F. Codd [2] presented Online Analytical Processing (OLAP) based on a dimensional view of data is being used increasingly for the purpose of analyzing very large amounts of data. Widom [10] presented a general research problem in data warehouses. The database community is devoting attention to the research themes concerning data warehouses.

Inmon [3] has been presented a data warehouse is a repository of integrated information, available for querying and modified, the information is extracted from its source, translated into a common model, and integrated with existing data at the warehouse.

Squire [8] presented the data warehousing is a promising technique for retrieval and integration of data from distributed, autonomous and possibly heterogeneous information sources. Data warehouses integrate data from multiple heterogeneous information sources and transform them into a multidimensional representation for decision support applications.

3. Background theory

Data warehouse (DW) is a relational database that is designed for query and analysis rather than for transaction processing. They support information processing by providing a solid platform of consolidated historical data for analysis [4].

Star schema contains a large central table (fact table) and dimension tables. The fact table contains the names of facts, or measures as well as keys to each of related dimension tables [6].

The Online Analytical Processing (OLAP) operates on historical data in both summarized and detailed forms. OLAP system provides facilities for decision makers to access and analyze the data in the data warehouses.

3.1. Data warehouse architecture

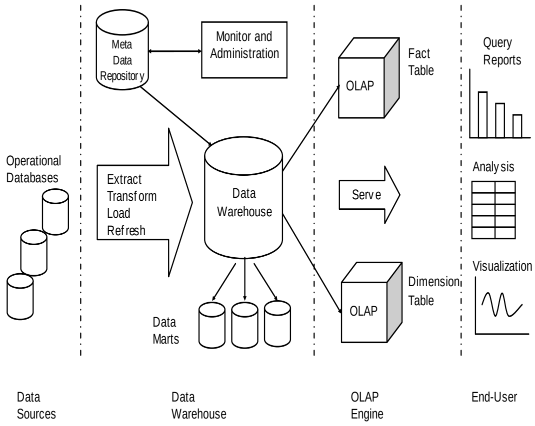

Figure 1. Data warehouse architecture

Fig. 1 shows a data warehousing architecture. It includes tools for extracting data from multiple operational databases and external sources; for transforming and integrating data; for loading data into data warehouse and for periodically refreshing the warehouse [8].

In addition to the main warehouse, there may be several departmental data marts. Data in the warehouse and data marts are stored and managed by one or more warehouses, which present multidimensional view of data to a variety of front end tools: query tools, report writes, analysis tools, and data mining tools.

Finally, there is a repository for storing and managing metadata, and tools for monitoring and administering the warehouse system.

3.2. Star schema

In the star schema, each dimension is represented by only one table and each table contains a set of attributes.

This star schema is presented one fact table and five dimension tables. Dimension tables are location dimension, department dimension, rank dimension, time dimension, and degree dimension. In the schema, the rank dimension table contains the attributes set {Rank-ID, Tutor, Assistant Lecturer, Lecturer, Associate Professor, and Professor}.

A star schema is a key concept for data warehousing. The star schema organizes data into a format that is easy for business users to understand. It allows application developers to standardize ad-hoc queries.

Star schema can be implemented for multidimensional database schema. It has four components: facts, dimensions, attributes and attribute hierarchies [6]. Fig. 2 shows star schema of the system.

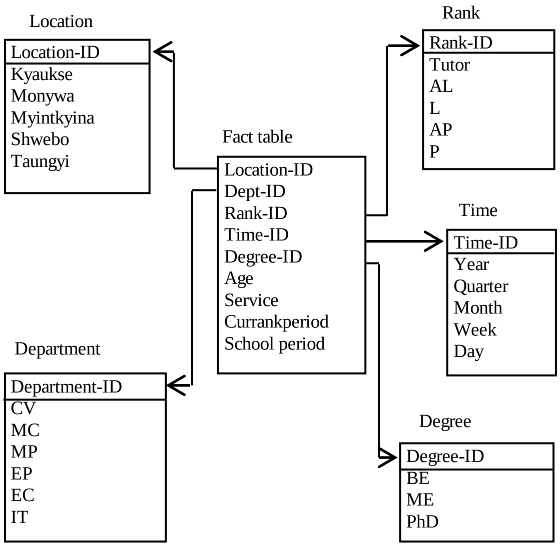

Figure 2. Star schema

Fact tables are the large tables “in the middle” of a dimensional schema [5]. Facts are generally numeric in nature, allowing aggregation of those numeric values into analytical data warehouse reporting structures. Fact tables simply link to more descriptive dimension tables [6].

Dimensions qualify characteristics that provide additional perspectives to a given fact. Each dimension and facts are a collection of attributes. Star schemas tend to perform the best in decision-support applications. Most data warehouses use a star schema to represent the multidimensional data model [9].

Dimension tables contain detailed descriptions about fact values such as time and location details of when and where a particular transaction took place. Dimensions qualify characteristics that provide additional perspectives to a given fact. Each dimension and facts are a collection of attributes.

3.3. Online analytical processing (OLAP)

Data warehouse systems serve users or knowledge workers in the role of data analysis and decision making. Such systems can organize and present data in various formats in order to accommodate the diverse needs of the different users. These systems are known as online analytical processing (OLAP) [9].

Fig. 3 shows a multidimensional data model. The significance of a “multidimensional view” of data is that it allows decision-makers to view many summarized data, historical data records, and groups them by subjects [6].

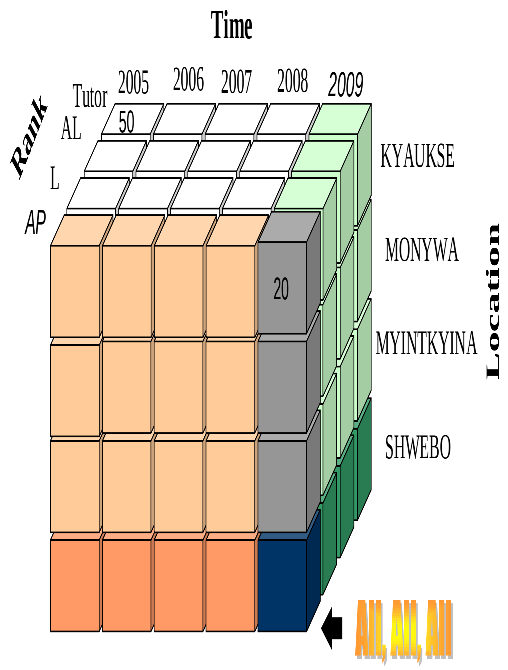

Figure 3. Multidimensional data model

OLAP tools include programs that show multidimensional data, rather like a spreadsheet with six or seven dimensions instead of only two with drill down capability.

While OLTP applications enhance transaction processing by relating atomized data across business processes, OLAP applications and databases provide users with the ability to view multidimensional data thereby enhancing the manager’s ability to create business models.

The significance of a “multidimensional view” of data is that it allows decision-makers to view many summarized data, historical data records, and groups them by subject.

OLAP allows business users to slice and dice data at will. Normally data in organization is distributed in multiple data sources and are incompatible with each other. It would a time-consuming process for an executive to obtain OLAP reports.

Part of the OLAP implementation process involves extracting data from the various data repositories and making them compatible. Data stored by operational systems such as point of sales, created quickly on the database and the results can be recalled by administrator.

4. System design

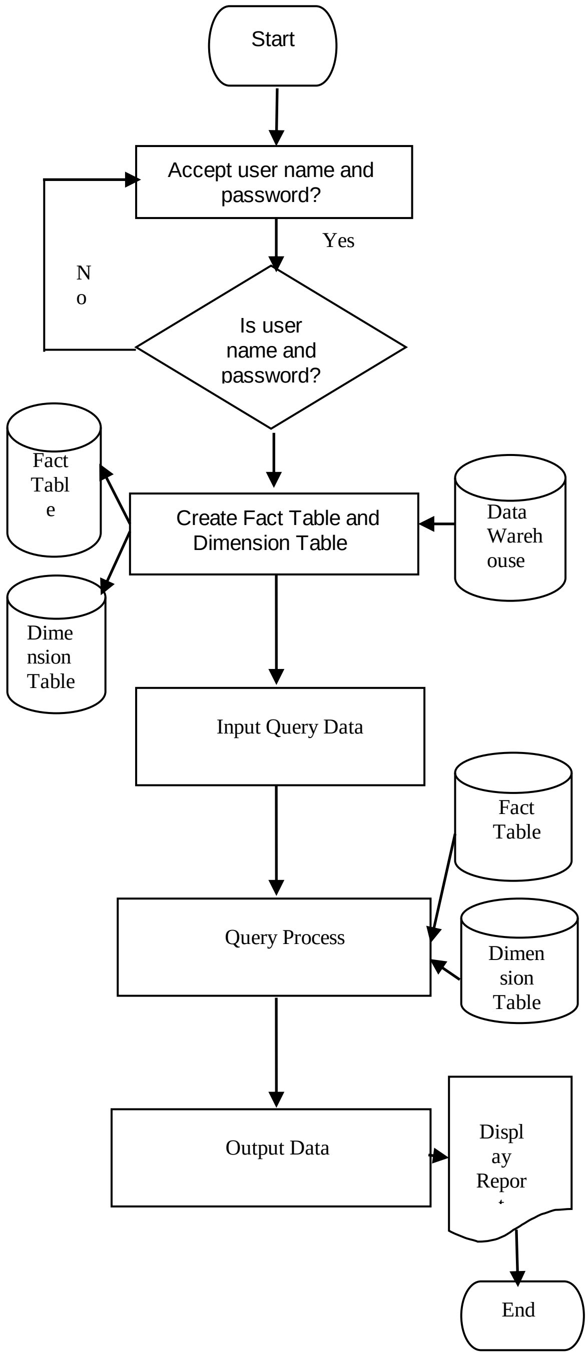

Figure 4. System flow diagram

Fig. 4 demonstrates "how to input into OLTP, how to extract and move into warehouse and how to retrieve the required data from warehouse using OLAP. There are many branch technological universities in Myanmar. Data from those operational sources are extracted and integrated into warehouse. Only authorized users access the staff data from the data warehouse according to their respective rolls.

Firstly, the user name and password are required. If the user name and password are correct, the staff data in technological universities are displayed as a data warehouse. And then, fact table and dimension tables are separately created from data warehouse.

If the user wants to know about someone data, he or she types the input query data. When the query process is started, the data are extracted from fact table and dimension tables. Finally, the system displays the output data to the user as OLAP reports.

5. Implementation result

When the program is run, the user login form appears in Fig. 5. In this form, login process reads the user name and password and checks whether these data are valid or not.

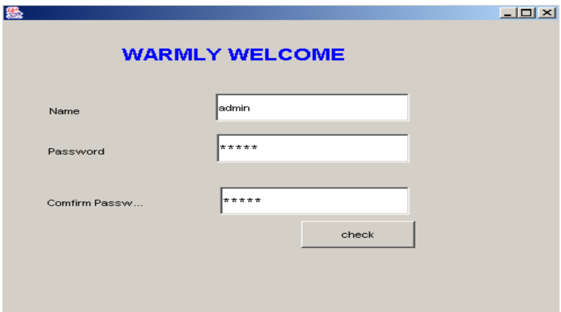

Figure 5. User login form

If the user name and password are correct, the system permits the user to enter the system. If the password is incorrect, the system does not allow the user to enter the system. The user must try again to enter the system until the user name and password are correct.

5.1. Main form

When the user name and password are correct, the main form dialog appears in Fig. 6.

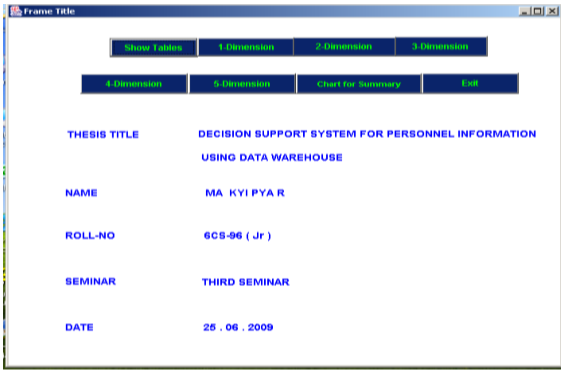

Figure 6. Main form

The main purpose of the system is to analyze the personnel data. The data warehouse and OLAP could produce query processing and maintenance issues. In this system, personnel data are considered into five dimensions such as Time, Location, Department, Degree, and Rank.

5.2. Dimension menus of staff warehouse

Fig. 7 shows staff data by time dimension view. In time dimension, the user can analyze staff data from various viewpoints by monthly or yearly. Fig. 8 shows staff data by location dimension view. In this dimension, the user can analyze location of staff. Fig. 9 shows staff data by rank dimension view.

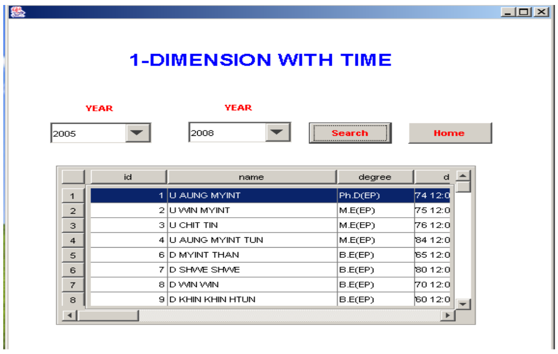

Figure 7. Staff data with one-dimension

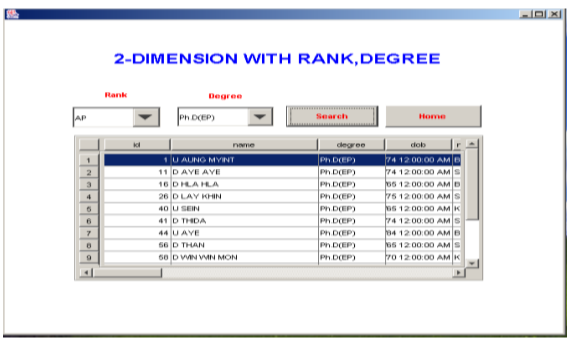

Figure 8. Staff data with two-dimension

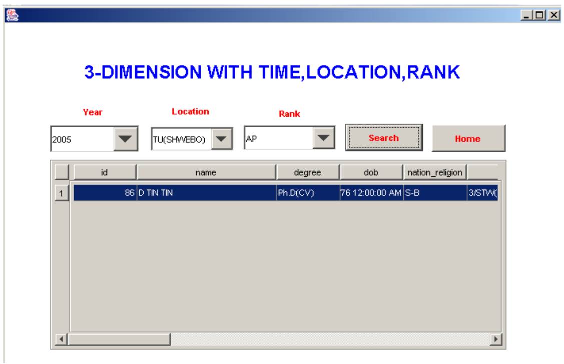

Figure 9. Staff data with three-dimension

Fig. 10 shows staff data by department dimension view. In department dimension, the user can analyze department of staff in each university.

Fig. 11 shows staff data by all dimension views. In all dimensions, user can see analysis all dimension information from various viewpoints by staff's name, age, service, location (university), and degree.

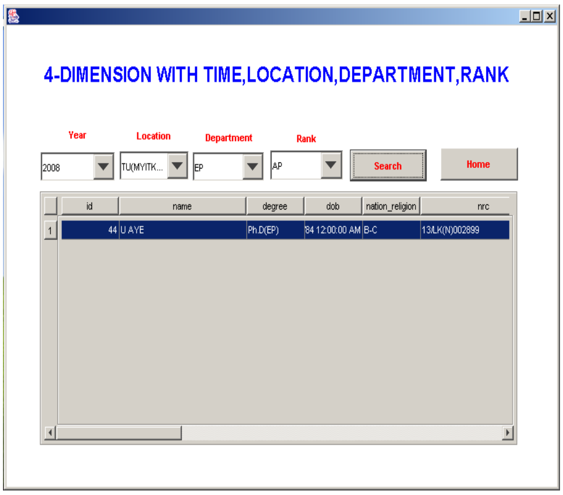

Figure 10. Staff data with four-dimension

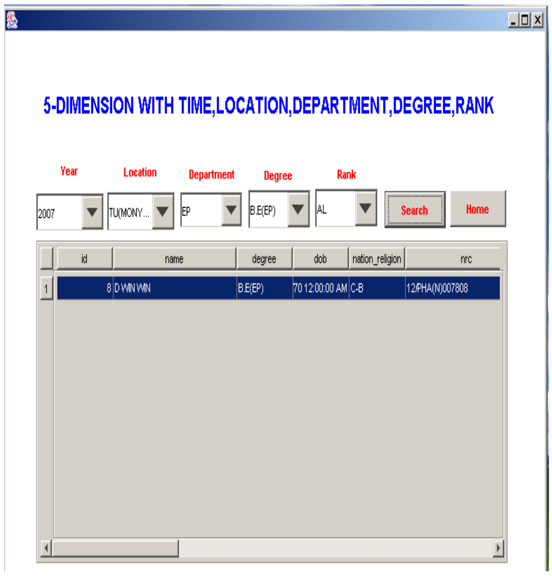

Figure 11. Staff data with five-dimension

5.3. Graphical reports

The following figures are examples of graphical views of report. In Fig. 12, the bar chart illustrates the total teachers in each university.

Fig. 13 shows the total PhD holders in each university are depicted. Fig. 14 shows the total teachers in each department in Technological University (Kyaukse).

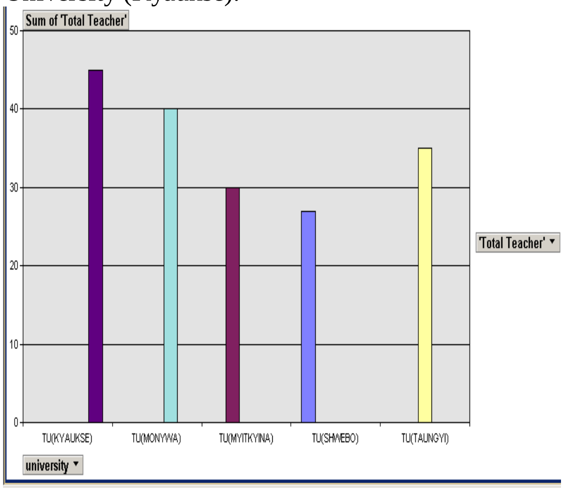

Figure 12. Chart for Total Teachers to each University (2005-2009)

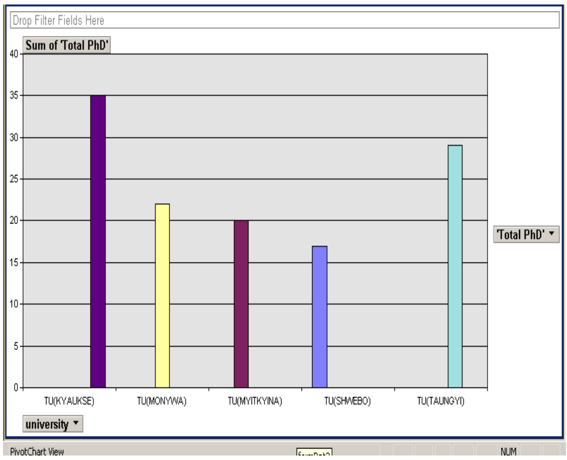

Figure 13. Chart for Total PhD Teachers to each University (2005)

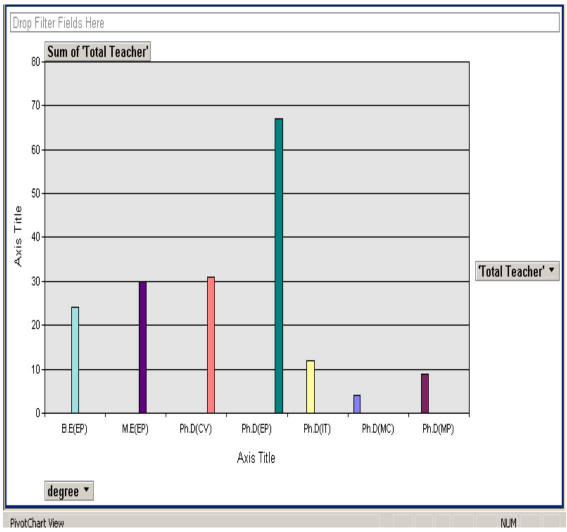

Figure 14. Chart for Total Teachers in KYAUKSE TU (2005)

5.4. Table design

The following Table 1 shows yearly summarized data in five technological universities.

Table 1. Yearly summarized data

Year	University	Total Teachers
2005	KYAUKSE TU	280
2006	MONYWA TU	250
2007	MYINTKYINA TU	200
2008	SHWEBO TU	180
2009	TAUNGYI TU	230

6. Conclusion

In Today's Global economy, IT manager needs the ability to quickly implement robust and scalable. Decision support systems provide the field of query optimization with increasing challenges in the traditional questions of selectivity estimation that can exploit transformations without exploding search space.

The aims of the paper are to understand what the data warehouse and protect the sensitive information stored elsewhere in data warehouse. This system is developed to easy access for frequently needed data, to improve end-user respond time.

The system help to retrieve complex query analysis report. In the paper, we analyze staff information by using Data Warehouse and OLAP tools.

10. References

[1] A. Berson and S.J.Smith, "Data Warehousing, Data Mining and OLAP", MC Graw-Hill Series on Data Warehousing and Data Management.

[2] E. F .Codd, "Providing OLAP (online analytical processing) to user-analysis" An IT mandate. E.F. Codd and Associates, 1993.

[3] W.H. Inmon, Building the data bridge: the ten critical success factors of building a data warehouse. Database Programming and Design,1992.

[4] H. Jiawei and K.Micheline, Data Mining Concepts and Techniques, Margan Kaufmann, 2001.

[5] R. Kimball, "DBMS Online Article", http://ww.dbsmag.com.

[6] M. Lupetin, InfoMaker Inc.,Glenview,Illions, "A Data Warehouse Implementation Using the Star Schema", http://www.microsoft.com.

[7] A. Marotta, A transformation based approach for designing Data Warehouse Internal Report. InCo. Universidad dela Republica, Montevideo,Uruguay, 1999.

[8] C. Squire, Data extraction and transformation for the data warehouse. In SIGMOD, May 1995.

[9] C. Surajit and D. Umeshwar, "An Overview of Data Warehousing and OLAP Technology", pp.2-3, Microsoft Research, Redmond, 2000.

[10] J. Widom, Research Problems in Data Warehousing, in Proc. 4th Int. Conf. on information and Knowledge Management, 1995.