DEVELOPING A DATA WAREHOUSE FOR DISTANCE REMOTE LABORATORY

Guzin Turkmen Nergiz Ercil Cagiltay
Atilim University Atilim University
Ankara, Turkey Ankara, Turkey

ABSTRACT
Data warehouse is an important contemporary issue for many organizations and is relatively a new field in the realm of information technology. As data warehousing, e-learning is also a new field. Little research has been done regarding the special considerations and characteristics of academic data and the complexity of analyzing such data. Educational institutions measure success very differently from business-oriented organizations and the analyses that are meaningful in such environments cause unique problems in data warehousing. Especially for the educational purposes, data warehouses offer several benefits.

This paper discusses the use of Data Warehouse and Decision Support resources to aid in the assessment of Distance Remote Laboratory Environment.

Keywords: Data warehouse, e-learning, DSS

I. INTRODUCTION
It is difficult to analyze the data in a large database system. Since the data is stored in multiple formats and different locations, the data changes constantly. In addition to this, it may be a challenge to access the historical data. Data warehousing provides a solution to these problems by consolidating data and making it available for analysis.

A. Definition of Data Warehousing
Over the last few years, institutions have increasingly turned to data warehousing to improve information flow and decision support. A data warehouse can be a valuable asset in providing easy access to data for analysis and reporting but building and maintaining an effective data warehouse has several challenges. The term “data warehouse” has been defined in various ways. For they bring different aspects of the concepts, several definitions will be given to provide the reader a better idea.

The first definition had been made by Bill Inmon (Dal, 2000), known as the father of the data warehouse concept as, “a Data Warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management’s decision making process” (Kimball, 1998).

A data warehouse is a central repository for all or significant parts of the data that an enterprise’s various business systems collect. W. H. Inmon defined the term. Inmon & Hackathorn stated that;

A data warehouse is a separate store of data extracted from one or more production databases to produce an authoritative source for decision support [1]. It can be considered a decision support system. Reporting data is another important aspect of data warehousing because the main output from data warehouse systems are either queries with minimal formatting or formal reports.

B. Structure of the Data Warehouse
Data contained in a data warehouse contains five types of data: data currency, older detailed data, data summarization (lightly and highly summarized data), and Meta data.

The following diagram is from "Building a Data Warehouse" 2nd Ed, by W.H. Inmon, Wiley '96

1) METADATA
Tannenbaum defined metadata as;
The detailed description of the instance data; the format and characteristics of populated instance data; instances and values depend on the role of the metadata recipient. Examples of metadata include filenames, data element definitions, data element names, lengths, program names and so on. (Tannenbaum 2002, p. 461)

An important component of the DW environment is metadata. Metadata or data about data is a critical component in a Warehouse environment. Metadata has many uses. Users need to know what data is available, how it’s organized and what it means. It allows the end user to navigate through the possibilities. In other words, when a user approaches a data warehouse where there is
no metadata, the user does not know where to begin the analysis.

Metadata typically contains:

- Structure of data as known to the programmer and to the DSS analyst
- Source data
- Transformation of data
- Data model
- DW
- History of extracts [2]

2) DATA CURRENCY

Current detailed data is at the heart of the data warehouse and reflects the most recent happenings. This data is stored according to a data model. Older detailed data accumulates over time as fresh data enters the application. Generally, in most decision support systems, data obtained from detailed records on transactions cleaned and then loaded. [3]

3) DATA SUMMARIZATION

Data is lightly summarized and stored in the data warehouse. The data might be summarized over different units of time and different attributes. Data could also be found in a highly summarized form for quick retrieval. The highly summarized data could be physically stored within or outside the data warehouse but is logically considered a part of the data warehouse. [4]

C. Data Warehouse Modeling Tools

Building a data warehouse from independent data sources is a difficult process. This process involves extracting, converting, cleaning, integration and transformation of the data. In order to do these operations, it is needed to use an ETL (extract, transform, load) tools. The key steps that need to be undertaken to transform raw operational data to a format that can be stored in a data warehouse for analysis are: [5]

- Extraction from a number of systems adds loading the data warehouse environment periodically. The goal of the data extraction step is to bring data from different sources into a database before modification.
- Converting the data into a format that is suitable to the data warehouse.
- Cleaning of the data. Data entry errors and differences in schema formation can cause for example student dimension table to have several corresponding entries for a single student. This is a very important step in data warehousing for believability of the data.
- Integration of the different datasets to suit the data model of the data warehouse.
- Transformation of the data through summarization and creation of new attributes. It is a set of rules and scripts typically handles the transformation of data from an input schema to the destination schema.

II. BENEFITS OF DATA WAREHOUSING

There are many benefits of data warehousing. The most important are listed below.

- Improves access to administrative information for faculty and staff.
- Can get data quickly and easily to do analysis. We can work with better information, make decisions based on data. Data Warehouse increase productivity of corporate decision-makers.
- Data extraction from its original data sources into the central area resolves the performance problem, which arises from performing complex analyses on operational data.
- Data in the warehouse is stored in a specialized form, called a multidimensional database. This form makes data querying efficient and fast.
- A huge amount of data is usually collected in the data warehouse. Compared with relational databases that are still very popular today, data in the warehouse doesn't need to be in normalized form. In fact, it is usually de-normalized to support faster data retrieval.

III. DATA WAREHOUSING IN EDUCATION

To stay competitive in today's rapidly changing environment, an organization needs to an effective database management. The concept of collecting data into separate, multidimensional repositories to handle complex decision making-activities can be used for educational purposes as well. Educational institutions need this capability to ensure quality data management for strategic decision-making. A review of the literature reveals that data warehousing becoming an increasingly popular way to store and retrieve data. Many educational institutions are creating a data warehouse to provide integrated administrative information for planning and reporting purposes. Today, data warehousing helps higher education enterprises understand who their customer base is, what they do, and what types of courses and services that we offer affect their learning and satisfaction [3].

Educational institutions are improving decision making capabilities by implementing successful data warehousing applications that address identifiable challenges and set attainable goals and follow the framework of guiding principles and program lifecycle management. Many administrators now recognize the importance of having accurate and accessible data in order to support the decisions that improve day-to-day operations, as well as long-range strategic planning. These data driven decision making capabilities can be greatly enhanced within institutions through the use of data warehousing.

Data warehousing collects and organizes data from multiple sources so that it can then be easily analyzed, extracted and used. Because of this, the data warehouse is a core component for enabling data driven decision making for educational institutions. The use of a data warehouse presents several potential advantages for educational institutions, including timely access to evaluate data. They can determine strengths and
weaknesses within themselves. The use of a data warehouse allows an educational institution to use this information in making appropriate decisions, and that is the desired end goal of the data warehousing process. [5] By using a data warehouse educational institutions could overcome this kind of problems. Not only that they set operational data free from complex retrieval, there are also many other benefits, concerning the management. For example, consider these few questions, which could be answered easily, using a data warehouse in a higher education institution:

• What is the total number of professors in Faculty of Engineering in a range of years?
• What are the STUDENT success ratios of Faculties, Departments and General ratio of the university?
• What is the total number of student that gets FF from COMPE 103 in years 2006-2007?

Note that those questions are not very easy to answer just by using a relational database. The main reason behind this problem is a simple fact namely,” relational databases are good at retrieving small number of records quickly, but they do not retrieve a large number of records and summarize them on the fly” [6]

Data Warehousing in Higher Education is a very young, growing industry, because decision makers in higher education need accurate and easily accessible information. Many Higher Education Institutions starting to see the value that integrated, standardized, clean and easy to access data has in terms of providing the end users an opportunity to solve their business problems. With the usage of data warehousing, management decision making and other reports can be done in a simpler way.

Universities use data warehouse in areas like Financial, Human Resources, Student grades, Financial Aid, Course, Student Fees, Research, Student Enrollment & retention, faculty workload and administrative.

Data warehouse in a remote laboratory environment also potentially provides several benefits for better system management and decision support. For example getting information about the most popular experiments for different groups of users among different time periods, how often each experiment is performed by different user groups in different time intervals, how often people require to look for the equipment instructions and does this situation changes in different time intervals. Similarly, such a data warehouse can also be supported by different remote educational systems to provide comparative analyses.

IV. CONCLUSION

To conclude; implementation of data warehouses in different educational platforms will support the managers and educators in their decision support.

REFERENCES