Original Research Paper

Education

A Novel Perspective of Data Warehousing and Data Mining in Education

Mr. P. NAGASURENDRAN

Assistant Professor, Roseline College of Education, Sivaganga

ABSTRACT

Data warehouses standardize the data across the organization so as to have a single view of information. Data warehouses can provide the information required by the decision makers. Developing a data warehouse for educational institute is the less focused area since educational institutes are non-profit and service oriented organizations. A data warehouse is a collection of integrated databases and subject-oriented and designed to support the decision-making function, where each unit of data is relevant to the events at any given time. Data Mining is an emerging technique with the help of this one can efficiently learn with historical data and use that knowledge for predicting future behavior of concern areas. Growth of current education system is surely enhanced if data mining has been adopted as a futuristic strategic management tool. The present study provides an option to build data warehouse and extract useful information using data warehousing and data mining.

KEYWORDS : Data Warehousing, Data Mining, ERP and OLTP

INTRODUCTION

The educational institutes have to generate funds for their research and other operational activities as the government funding has been limited to aided institutes. Utilizing a decision support system is a proactive way to use data to manage, operate, and evaluate educational institute in a better way. Depending on the quality and availability of the underlying data, such a system could address a wide range of problems by distilling data from any combination of education records maintenance system. The data mining from data warehouse can be a ready and effective system for the decision makers. A data warehouse is a subject oriented integrated, non-volatile, and time variant collection of data in support of management decisions . The data from these sources are converted into a form suitable for data warehouse. This process is called Extraction, Transformation and Loading (ETL).

Data mining is useful whenever a system is dealing with large data sets. In any education system, student records i.e. enrollment details, course eligibility criteria, course interest and academic performance may be an important consideration to analyze various trends since all the systems are now computer based information system so data availability, modification and updation are a common process now. Data warehousing may be taken as good choice for maintaining the records of past history. The data warehouse can be easily developed in any education institute with the adaptation of common data standard. Common data standards may eliminate the need of data clarity and modification before loading this for a data warehouse.

An institute with efficient Data Warehousing and Data Mining approach can find out novel way of improving student's behavior, success rate and course popularity. All these effort may finally improve the quality of education, better student intake, better career counseling and overall practices of education system. To face these challenges different systems are used such as ERP, OLTP etc.

ENTERPRISE RESOURCE PLANNING (ERP)

Enterprise resource planning (ERP) integrates internal and external management information of an entire organization like finance/accounting information, manufacturing, sales, service information and most important its customer relationship management. ERP opened new horizons in almost all sectors. In this paper, we are mainly concerned about educational sector. Education is one of India's prime Industries today. Its size is more than domestic software industry or automobile industry. Due to exponential growth of educational sector, educational institutes have now become complex organizations. They are no longer limited to deliver education only, but to manage a large range of activities like marketing of institutes for student admission and corporate student's for placements, managing internal operations like smooth conducting of classes or recruitment and motivation of human resources like faculty and staff, financial and cash flow planning, co-ordination with regulatory and statutory authorities. In addition to that, institutes are also subjected to the vagaries of market forces due to stiff competition and demanding customers (students and corporate).

OLTP (ON-LINE TRANSACTION PROCESSING)

OLTP (On-Line Transaction Processing) has been developed as complete ERP solution for academic institution especially targeted to engineering colleges. The primary objective of the OLTP is to facilitate students and faculty to get a robust and advanced atmosphere of learning. OLTP's academic IT Infrastructure helps us to manage thousands of students from a single integrated management control. It enables students to take up tests on various subjects and get to know their results very fast. The faculty at the same time could easily build their tests and publish them. Student's interactions with online learning environments enable them to access online exercise work, to know their mistakes and to get teacher's comments etc.

DATA WAREHOUSING AND DATA MINING

There are many application areas of data warehousing and data mining like retail sales, telecommunications, transportations, education, customer analytics, agriculture, banking, security applications, mass surveillance, privacy preserving etc. The main concerned area is about data warehousing and data mining applications in educational systems are covered this study.

DATA WAREHOUSING

Data Warehousing is an information delivery system. It is another solution specifically designed for query and analysis of information related to any educational institute. It is relational database approach rather than traditional transaction processing. Data warehouse is categorized based on its data storage. The main source of the data is cleaned, transformed, catalogued and made available for use by managers and other business professionals for data mining, online analytical processing, market research and decision support. In an educational institute it plays a very important role.

Main benefits in an educational institute are listed as follows,

- It provides an integrated and total view of an institute.
- It makes the institutes current and historical information easily available for the decision making.
- It provides the facility to students to get their different subject notes from a web enabled database.
- It provides the information about student's attendance.
- Students can get their results easily and very quickly.
- It helps to provides information about faculty like how many members are their in all the different departments etc.

Overall we can say that data warehousing just simplifies a complex system to a simple and easily accessible system. Data Warehouse maintains its function in three layers: staging, integration and access. Staging is used to store raw data for developers. The integration layer is used to integrate data and to have a level of abstraction from users. The access layer is for getting out for users. which divided mainly into four parts as follows,

Data sources

The data in the data warehouse is derived from the operational sys-

tems that support the basic business processes of the organization. Source data coming into the data warehouse may be grouped into four broad categories like production data, internal data, external data and archived data.

Data staging
In this layer operational data is cleansed and transformed into a form suitable for placement in the warehouse for easy retrieval. Three major functions performed in data staging are data extraction, data transformation and preparing it for loading.

Data storage
In this stage data is stored in the data warehouse. The operational system of an enterprise supports only the current data but in data warehouse historical data is also kept.

Information delivery
In this stage useful information is provided to the wide community of data warehouse users through various systems like online, intranet, internet and e-mail etc.

The ability of a data warehouse that can support the following actions,

Running business data: data is produced by enterprise applications

Integrate business data: to improve and synchronize two or more enterprise applications, even those that do not designed to work alongside each other

Monitor business data: determine the relationship between data, deliver it to end users as reporting tools and support decision making process.

DATA WAREHOUSE STRUCTURE
Data warehouse has some level of detail that is older level of detail (usually found in alternative storage), current level of detail, lightly summarized the data (data mart level), and highly summarized data. The flow of data into the data warehouse from the operational environment. The transformation of data usually come from the operational level to the data warehouse level.

DATA MINING
Data is extraction of interesting patterns or knowledge from huge amount of data. Data mining can be defined in several ways, which differ primarily in their focus on different aspects of data mining. The use of machine learning algorithms to find faint patterns of relationship between data elements in large, noisy, and messy data sets, which can lead to actions to increase benefit in some form viz. diagnosis, profit, detection etc. It is an interdisciplinary field of astronomy, business, computer science, economics and others to discover new patterns from large data sets. The actual data mining task is to analyze large quantities of data in order to extract previously unknown patterns such as groups of data records, unusual records and dependencies.

These patterns can then be seen as a kind of summary of the input data, and used in further analysis. Data mining tasks can be classified as

1. Anomaly Detection
The identification of unusual data records, that might be interesting or data errors which require further investigation.

2. Association Rule Learning
Searches for relationships between variables. For example a supermarket might gather   data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes.

3. Clustering
It is a task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.

4. Classification
It is the task of generalizing known structure to apply for new data. For example, an email program might attempt to classify an email as legitimate or spam.

5. Regression
It attempts to find a function which models the data with the least error.

6. Summarization
It providing a more compact representation of the data set, including visualization and report generation.

FURTHERANCE
Real time ETL refers to the software that moves data synchronously into a data warehouse with some urgency-within minutes of the execution of the business transaction. Implementation of real-time data warehouse reflects a new generation of hardware, software and techniques. Capture, Transform, and Flow (CTF) is a relatively new category of data integration

tools designed to simplify the movement of real-time data across heterogeneous database technologies. The transformation functionality of CTF tools is typically basic in comparison with today's mature ETL tools, so often real time data warehouse CTF solutions involve moving data from the operational environment.

CONCLUSION
Data warehousing and data mining methods are explained in detail. Educational institutes need modern management practices and state of the art technology to manage their internal and external operations. Realizing such demands, software industry started developing automated solutions for educational administration. The data mining from data warehouse can be a ready and effective system for the decision makers. Finally detail evaluation need to be done to make sure the implementation of data warehouse, data mining and dashboard meet Higher Education management needs. The study may help decision makers of educational institutes across the globe for better decisions.

REFERENCES
1. Turban, E., Sharda, R., Delen, D., & King, D, "Business Intelligence a Managerial Approach ", Second Edition. New Jersey: Pearson, 2011.
2. L. Jing, "Data Mining as Driven by Knowledge Management in Higher Education" Keynote for SPSS Public Conference, UCSF, 2001.
3. WEKA : Data Mining Software in Java" (n.d.) Retrieved March 2010
4. Ralph Kimball, the Data Warehouse ETL Toolkit, Wiley India Pvt Ltd., 2006.
5. McMillen and Randy McBroom.,'what academia can gain from building a data warehouse' 2008
6. Bidgoli, and Hossein, The Internet Encyclopedia: John Wiley and Sons, 2004.