

By Applying Data Mining Technique for Taking Quick Decision in Educational Domain

Shabina Kausar
M. Tech. (Cse)
Gniet, Nagpur, Maharashtra, India

Prof. Kalpana Malpe
Assistant Professor
Gniet, Nagpur, Maharashtra, India

Abstract- The present study provides an option to build data warehouse and extract useful information using data warehousing and data mining open source tools. 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. Data-driven decision support systems, such as data warehouses can serve the requirement of extraction of information from more than one subject area. 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. In present day scenario where education has been privatized and cut throat competition is prevailing, institutes needs to be more organized and need to take better decisions. Institute's enrollments are increasing as a result of increase in the number of branches and intake. Now a day, any reputed Institute's enrollments count in to thousands. The cost of building a data warehouse is expensive for any educational institution as it requires data warehouse tools for building data warehouse and extracting data using data mining tools from data warehouse. In this paper we have explored the need of data warehouse / business intelligence for an educational institute, the operational data of an educational institution has been used for Experimentation. The present study provides an option to build data warehouse and extract useful information using data warehousing and data mining open source tools.

I. Introduction

This data base contains data about data-description of source data, target data and how the source data has been modified into target data. The client software will be used to generate reports. Data warehouse obtains the data from a number of operational data base systems which can be based on RDBMS/ERP package, etc. The data from these sources are converted into a form suitable for data warehouse. This Process is called Extraction, Transformation and Loading (ETL).

In addition to the target database, there will be another database to store the metadata, called the metadata repository.

In present day scenario where education has been privatized and cut throat competition is prevailing, institutes needs to be more organized and need to take better decisions. Now a day, the educational institutes have to generate funds for their research and other operational activities as the government funding has been limited to aided institutes. By applying data mining on data warehouse is a proactive way to use data to manage, operate, and evaluate educational institute in a better way. Data-driven decision support systems, such as data warehouses can serve the requirement of extraction of information from more than one subject area. 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.

II. Objective

1. The main aim of our project is to help decision makers of educational institutes across the globe for better decisions.

2. It was used to investigate current system of information delivery and proposing a better system for timely, accurate, consistent information delivery to the decision makers of the educational institute.
3. The project is preparing in order to extend the usage of current available technology in decision making processes of educational institute.

III. Motivation

Motivation for building data warehouse for the educational institute is from two sources, internal sources like inability of current operational systems to provide required information for parameter driven analysis and external sources like competitive factors. A survey is carried out by visiting several educational institutes to gather information regarding the current practices the institutes have implemented as decision support systems. The findings are summarized below.

- 1) The data is stored in different sources in distributed locations.
- 2) Users find difficulty in locating the reports needed by them.
- 3) The user interface for the current operational system is not satisfactory and is confusing and hard to use for decision makers.
- 4) When the consolidated report from two or more different subject area is required, it is almost impossible.
- 5) There is no easy way to get assistance.

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 purpose of this paper is to investigate current system of information delivery

and propose a better system for timely, accurate, consistent information delivery to the decision makers of the educational institutes.

IV. Proposed system

1. Connecting source databases:

This module is used to connect all the databases which are used as a source databases, they are in any format such as SQL server database, MS access database, CSV files, etc. This module connects all the working databases because Microsoft SQL server doesn't run or execute any query till the database is not attached to SQL server. Therefore the current stage is necessary and very essential.

In this module the developers don't know the structure of available databases therefore there is necessity to connect it to Master database of SQL server. Master database will give the complete details about the current connected databases.

Working procedure for first module

1. First start the software.
2. After that there appear a main form which having a menu bar named as file and it consist of two tabs connect and upload.
2. Data staging:

This module is used as an intermediate database between source databases and data warehouse; here we use Data Table class of C# .Net. Data Table class is not the part of the SQL server or MS Access; it is the part of the Microsoft .NET system. It is present in system Data library.

In data staging we first apply data mining for comparison we used dictionary variables.

There are approximately 50 different types of dictionary variables & all these variables 4 to 5 are synonyms. After comparison we are applying pipe sort algorithm in that pipe sort joints no. of tables through different path & create a single fact table in this way we can create three different fact tables which is stored in temporary memory.

3. Data warehouse generation:

This is very important, because this module is used to generate data warehouse of the project. Pipe sort methods of OLAP system and generate a data warehouse for perfect decision making.

- 1) In data warehouse generation the data is particularly stored in fact tables from temporary memory from hard disk.
- 2) This can be achieved by using a command builder of SQL server which gives a facility to execute any types of queries.

Query can be executed by using SQL data adapter after execution of any type of query, the data adapter adapts the executed data & finally it stored in data warehouse fact table by using one function that is known as update.

4. Visual output system:

Module is used to generate any type of reports and decisions from the generated data warehouse. It provides both complicated query writing system as well as complete visual lookout system where user can select the reporting type and many a required field and generate reports.

There are two approaches to show output of a system.

1. Technical approach

Technical approach gives a freedom to write any type of query, but it requires complete knowledge of SQL query.

2. Non technical approach

Non technical approach, have some criteria that criteria can be searched by using data mining concept.

Data mining takes the criteria to find and scan the process of fact table for one single output.

V. Dataware house Environment

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. The data from reputed engineering college namely R V College of Engineering, Bangalore, Karnataka, India, has been considered for this study. Fig. 1 shows the data warehouse architecture of RV College where source systems are smart campus, asset management server and csv files, the information is spread across diverse platforms, data from different sources is collected and then consolidated to produce required report.

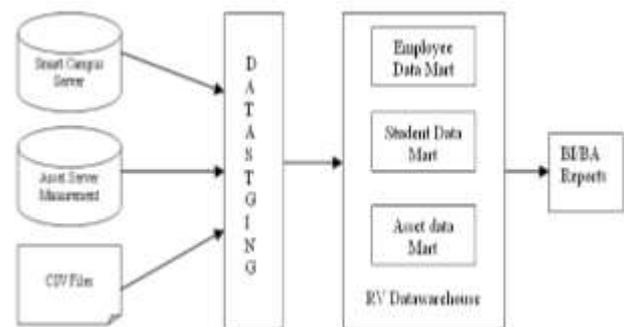


Fig. 1. Engg_Data warehouse architecture

Data warehouse enables the decision makers with benefits listed below.

- 1) Phenomenal improvements in turnaround time for data access and reporting
- 2) Standardizing data across the organization so that there will be one view of information.
- 3) Merging data from various source systems to create a more comprehensive information source.
- 4) Reduction in costs to create and distribute information and reports.
- 5) Encouraging and improving fact-based decision making.

Technology Used :

Data Mining

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data

mining is the process of finding correlations or patterns among dozens of fields in large relational databases

Pipe Sort :

The Pipe Sort method combines the optimisations share-sorts and smallest-parent to get the minimum total cost. Here a number of statistical procedures for this purpose. The input to the algorithm is the search lattice, which is a graph where a vertex represents a group-by of the cube. A directed edge connects group *i* to group-by *j* whenever *j* can be generated from *i* and *j* has exactly one attribute less than *i* (*i* is called the parent of *j*). Thus, the out-degree of any node with *k* attributes is *k*.

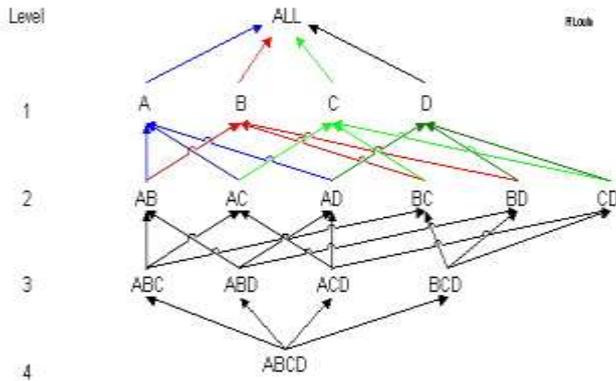


Fig2: Pipe sort algorithm

Viterbi algorithm :

The Viterbi algorithm is a dynamic programming algorithm for finding the most likely sequence of hidden states – called the Viterbi path – that results in a sequence of observed events, especially in the context of Markov information sources and hidden Markov models.

Forward algorithm

The forward algorithm in the context of a hidden Markov model is used to calculate a 'belief state': the probability of a state at a certain time given the history of evidence. The process is also known as filtering. The forward algorithm is closely related to but distinct from the Viterbi algorithm.

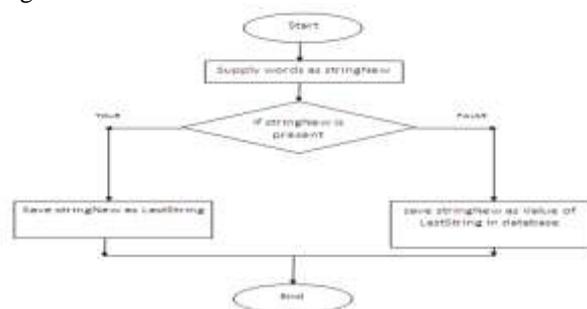


Fig3: Forward algorithm

Working procedure

- First of all start the software.
- And then start the connect screen.
- And logon to the SQL server as you know such that if you want to connect local SQL server then you have to just select windows authentication and click on connect button, otherwise you have another choice to

select SQL authentication in which you have to provide username and password to connect it to server machine SQL server.

- After login the software next click on the link upload database.
- After that there is a button known as browse database click on it we have maximum three database choice.
- After that we can check the tables present in the current database by clicking on tables button below the database.

VI. RESULTS

Once the data warehouse is deployed, it invariably becomes a mission-critical application. Users depend on the data warehouse to provide them with the information they need to function properly. To make certain that the ETL process runs and completes, it must be actively monitored and supported. Some of the results observed after querying the data marts are documented below. The results are cross checked with the requirements specified by the different types of users. The requirements with regard to asset information were to extract the information regarding the number of assets of each type in the Institute. The different data marts are queried using SQL query. The results returned by the queries are found accurate and meeting users demands. The sample screen shots of queries and the results are shown,

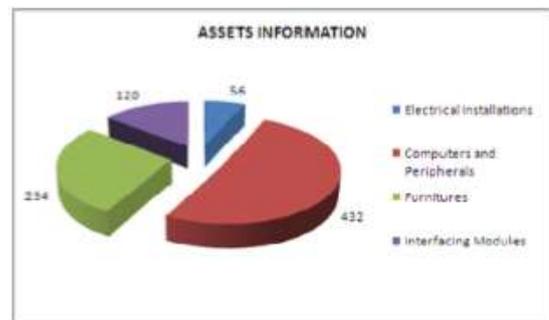
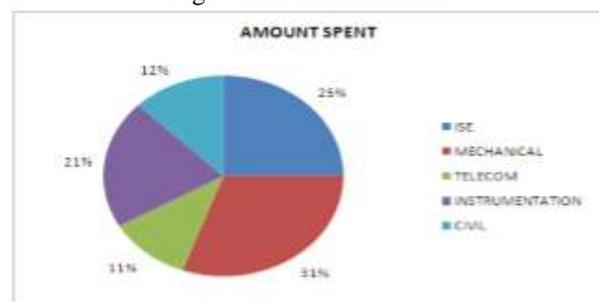


Fig1. Assets Information



Fi2g. Amount spent

Fig. 2 gives the asset information of the institute like electrical installations, computers and peripherals, furniture’s and interfacing modules. Fig. 3 gives amount spend on each department viz EE, MECHANICAL, TELECOM, INSTRUMENTATION and CIVIL. Fig. 2 and Fig. 3 are output from asset mart.

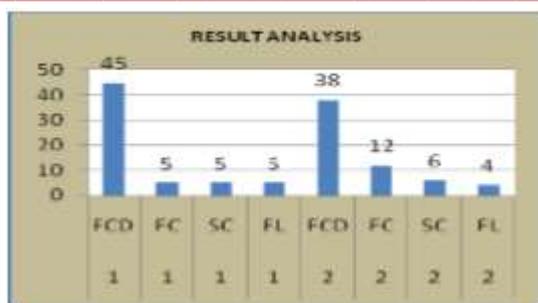


Fig. 4. Result Analysis

Fig. 4 shows the detailed result analysis which shows number of students who have obtained different classes; this is the output from student mart.

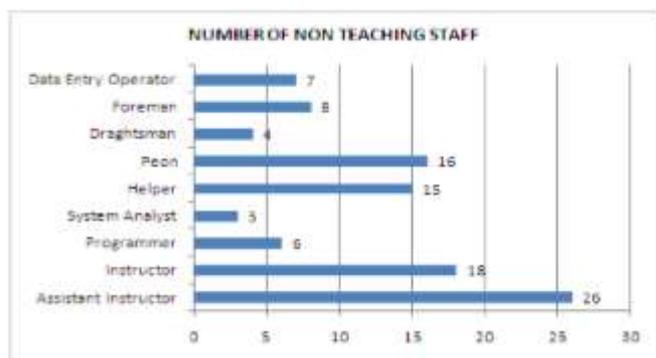


Fig. 5. Staff details

Fig. 5 gives the information regarding was the number of employees in each cadre of two particular departments. This is the output from employee mart.

VII. Conclusions

The newly generated systems visual lookout becomes very standard and easy for working. Institute can take micro-level decisions in a timely manner without the need to depend on their IT staff. They can perform extensive analysis of stored data to provide answers to the exhaustive queries to the administration cadre. This helps them to formulate strategies and policies for employees and students. This helps students

and Employees in making decisions. They are the ultimate beneficiaries of the new policies formulated by the decision makers and policy planner’s extensive analysis on student and employee relate data.

References

- [1] Alex Berson, Data Warehousing Data Mining & OLAP, Computing Mcgraw-Hill, November 5, 1997.
- [2] Ralph Kimball, the Data Warehouse ETL Toolkit, Wiley India Pvt Ltd., 2006.
- [3] KV. K. K. Prasad, Data warehouse development Tools, Dreamtech Press, 2006.
- [4] W. H. Inmon, Building the Data Warehouse. Wiley; 3rd edition March 15, 2002.
- [5] McMillen and Randy McBroom., ‘what academia can gain from building a data warehouse’ no.1,pp.41-46.2008
- [6] Channah F. Naiman, Aris M. Ouksel "A Classification of Semantic Conflicts in Heterogeneous Database Systems", Journal of Organizational Computing, vol. 5, 1995.
- [7] John Hess, "Dealing With Missing Values In The Data Warehouse" A Report of Stonebridge Technologies, Inc-1998.
- [8] Manjunath T.N, Ravindra S Hegadi, Ravikumar G K."Analysis of Data Quality Aspects in DataWarehouse Systems", (IJCSIT) International Journal of Computer Science and Information Technologies, vol. 2 (1) , 2010, 477-485.
- [9] Amit Rudra,Emilie Yeo (1999) "Key Issues in Achieving Data Quality and Consistency in Data Warehousing among Large Organizations in Australia", Proceedings of the 32nd Hawaii International Conference on System Sciences – 1999.
- [10]Jaideep Srivastava, Ping-Yao Chen,"Warehouse Creation-A Potential Roadblock to Data Warehousing", IEEE Transactions on Knowledge and Data Engineering January/February 1999 (vol. 11, no. 1) pp.118-126.