
Application of Data Warehousing Analysis in Education Domain

Shabina Kausar*

Deptt. of Computer Science & Engineering
Guru Nanak Institute of Engg. & Technology
Dahegaon, Nagpur, India

Kalpana Malpe

Deptt. of Computer Science & Engineering
Guru Nanak Institute of Engg. & Technology
Dahegaon, Nagpur, India

Abstract

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. 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. The present study provides an option to build data warehouse and extract useful information using data warehousing and data mining open source tools. 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.

Keywords: Data warehousing, Education domain.

****Author for correspondence*** ansarishabikausar@gmail.com

1. Introduction

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. 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. 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 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.

2. 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.

- a. The data is stored in different sources in distributed locations.
- b. Users find difficulty in locating the reports needed by them.
- c. The user interface for the current operational system is not satisfactory and is confusing and hard to use for decision makers.
- d. 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, to the decision makers of the educational institutes.

3. Data Warehouse Applications

In the previous work, there are some different application areas for which data warehouses are built.

Retail Sales

Data is collected at several interesting places in a grocery store. Some of the most useful data is collected at the cash registers as customers purchase products. Modern grocery store scans the bar codes directly into the point of sale (POS) system. The POS system is at the front door of the grocery store where consumer take away is measured. The back door, where vendors make

deliveries, is another interesting data collection point. At the grocery store, management is concerned with logistics of ordering, stocking, and selling products while maximizing profit. Some of the most significant management decisions are on pricing and promotions. Both store management and marketing spend a great deal of time tinkering with pricing and promotions. In such scenarios, data warehouses come to rescue.

Telecommunications

A telecommunications company generates hundreds of millions of call-detail transactions in a year. For promoting proper products and services, the company needs to analyze these detailed transactions. The data warehouse for the company has to store data at the lowest level of detail.

Transportation

In this case, the airline's marketing department wants to analyze the flight activity of each member of its frequent flyer program. The department is interested in seeing which flights the company's frequent flyers take, which planes they fly, what fare basis they pay, how often they upgrade, how they earn. These requirements can be fulfilled by data warehouse.

Education

There are some efforts in the area of data warehouse for building data warehouse for education domain. The paper by Carlo DELL'AQUILA summarizes the experience in designing and modeling an academic data warehouse. Existing facilities and databases affect the chosen data warehouse that brings them together to support decisional activities leading the whole university environment, including administrators, faculties and students. The choice to develop a dedicated system is mainly forced by the peculiar information type that defines the basic information in data warehouse widely different from institution to institution. In the article titled "what academia can gain from building a data warehouse?" by David Wierschem et al., authors have identified the opportunities associated with developing a data warehouse in an academic environment. They begin by explaining what a data warehouse is and what its informational contents may include, relative to the academic environment. Next they addressed the current environment drivers that provide the opportunities for taking advantage of a data warehouse and some of the obstacles inhibiting the development of an academic data warehouse. Finally, the article provides strategies to justify developing a data warehouse for an academic institution.

Disadvantages of the existing system

There are some disadvantages of the current available system, these are given below:

- a. All the available data warehouses are specifically done for some professional and profit based organization. So, there is no system available for education domain.
- b. In education domain, the available system provides the reports in a manner such that they all are different from each other, in between them there is no connection.
- c. The current available system requires more time, and accuracy is very less, and therefore it is not helpful to the decision makers of the educational institute.

4. Proposed Approach

In the previous work, any reputed institute's enrolments count in to thousands. In view of these factors, the challenges for the management are meeting the diverse needs of students and facing increased complexity in academic processes. And also when the consolidated report from two or

more different subject area is required, it is almost impossible. There is no easy way to get assistance. And therefore the complexity of these challenges requires continual improvements in operational strategies based on accurate, timely and consistent information. Developing a data warehouse for educational institute is the less focused area since educational institutes are non-profit and service oriented organizations. After knowing all this fact we decided to work on the current topic known as “application of data warehousing analysis in education domain”.

Need of the system

This system is very necessary because the current system is very tedious and very confusing for work and the report generated by the current system is very time consuming and therefore we think that the new system generated by us is very useful for any type of education domain. The visual lookout of the system is in such a way that it doesn't create any type of confusion for end user due to which to work on the new system is very interesting. 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. The project is preparing in order to extend the usage of current available technology in decision making processes of educational institute.

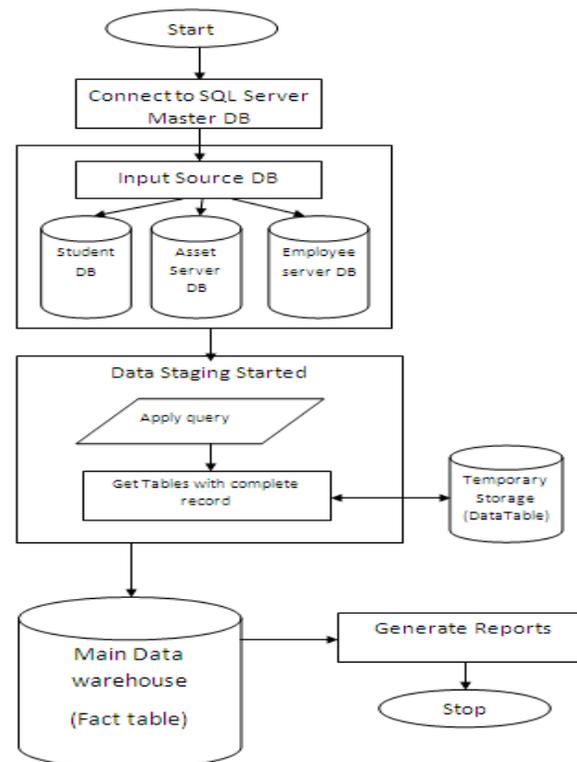


Fig.: Overall Architecture

5. Analysis and Design

Life cycle of the system

Waterfall model is one of the process models used in software development. The waterfall model is a sequential design process, often used in software development processes, where progress is seen as flowing steadily downwards through the phases of Conception, Initiation, Analysis, Design, Construction, Testing, Production/Implementation, and Maintenance. This model is also called as the classic life cycle model as it suggests a systematic sequential approach to software development. The process begins with the communication phase where the customer specifies the requirements and then progress through other phases like planning, modelling, construction and deployment of the software.

Five Phases of Waterfall Model

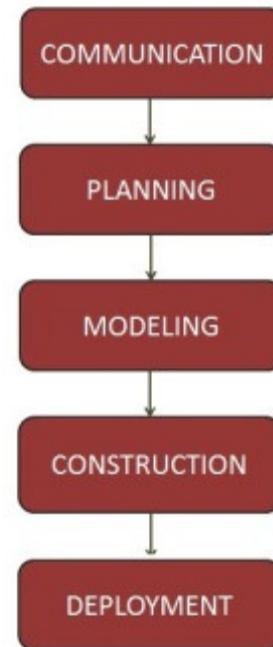
- a. Communication: The major task performed is requirement gathering which helps in finding out exact need of customer. Once all the needs of the customer are gathered the next step is planning.

b. Planning: Major activities like planning for schedule, keeping tracks on the processes and the estimation related to the project are done. Planning is even used to find the types of risks involved throughout the projects. Planning describes how technical tasks are going to take place and what resources are needed and how to use them.

c. Modelling: The architecture of the system is designed in this phase. Analysis is carried out and depending on the analysis a software model is designed. Different models for developing software are created depending on the requirements gathered in the first phase and the planning done in the second phase.

d. Construction: The actual coding of the software is done in this phase. This coding is done on the basis of the model designed in the modelling phase. So in this phase, software is actually developed and tested.

e. Deployment: The product is actually rolled out or delivered & installed at customer's end and support is given if required. A feedback is taken from the customer to ensure the quality of the product.



From the last two decades Waterfall model has come under lot of criticism due to its efficiency issues. So let's discuss the advantages and disadvantages of waterfall model.

Advantages of Waterfall model

It is the oldest and most widely used model in the field of software development. There are certain advantages of this model, which makes it, one of the most widely, used models as yet.

- Being a linear model, it is very simple to implement. The amount of resources required to implement this model are minimal.
- Documentation is produced at every stage of the software's development. This makes understanding the product designing procedure, simpler.
- After every major stage of software coding, testing is done to check the correct running of the code.

Connecting to source database module

This module is nothing but the module where we can connect all the working databases. Because Microsoft SQL server doesn't run or execute any query till the database is not attached in SQL server. Therefore the current stage is necessary and very important. In this module, first of all we don't know the structure of the available database because that database is not created by us. And therefore there is necessity to connect it to Master database of SQL server. Master database will give the complete details about the current attached database.

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.

6. 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] Ralph Kimball. Data Warehouse ETL Toolkit. Wiley India, 2006.
- [2] KV. K. K. Prasad. Data Warehouse Development Tools. Dreamtech Press, 2006.
- [3] W. H. Inmon. Building the Data Warehouse. 3rd ed. Wiley. 2002.
- [4] Alex Berson. Data Warehousing Data Mining & OLAP. Computing McGraw-Hill. 1997.
- [5] McMillen and Randy McBroom. What academia can gain from building a data warehouse? 2008.
- [6] C. F. Naiman, Aris M. Ouksel. (1995). A Classification of Semantic Conflicts in Heterogeneous Database Systems. *Journal of Organizational Computing*, Vol. 5.
- [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. (2010). Analysis of Data Quality Aspects in DataWarehouse Systems. *Inter. Jr. of Com. Science and Inf. Technologies*. 2(1), pp. 477-485.
- [9] Jaideep Srivastava, Ping-Yao Chen. (1999). Warehouse Creation: A Potential Roadblock to Data Warehousing. *IEEE Transactions on Knowledge and Data Engineering*. 11(1), pp.118-126.
- [10] Amit Rudra, Emilie Yeo. (1999). Key Issues in Achieving Data Quality and Consistency in Data Warehousing among Large Organizations in Australia. *In: Proceedings of the 32nd Hawaii International Conference on System Sciences*.