Critical Need of the Data Warehouse for an Educational Institution and Its Challenges

Shaweta,
Assistant Professor,
UIET, Punjab University, Chandigarh, India

Abstract - This paper is going to propose a need of data warehouse for an Educational institute that aims at supporting top level management by providing them different views for the same old piece of data. Various steps for developing and building the data warehouse system are shown, with special focus on how to customize it to the Educational institute. The main concepts of the data warehouse are reviewed first, and then its importance for the Educational institutes. Also this paper addresses issues related to building and maintenance of a data warehouse for the Institutions.

1. INTRODUCTION:
To provide top level management in the institution with information about and going insight into the existing data is the key function of the data warehouse systems, in order to make more informative decisions without interrupting the daily work of an On-Line Transaction Processing (OLTP) system.
A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process [4]. W. Inmon and E. F. Codd observed in the early 1990’s that on-line transaction processing (OLTP) in the operational-level applications and On-line Analytical processing (OLAP) in decision support applications cannot coexist effectively in the same database environment, mainly due to their very different transaction characteristics [1] [16]. Ralph Kimball provided a more concise definition of a data warehouse: A data warehouse is a copy of transaction data specifically structured for query and analysis. Middleware-based database cluster can be used to remove the distinction between OLAP and OLTP [3].

Prior work of common architecture of a Data Warehouse and the prototype development of Data Warehouse has been done in construction Educational [5].

The rest of this paper is organized as follows: Section 2 illustrates the Generic Data Warehouse architecture, Subsection 2.1 a brief about Extraction, Transformation, and Loading processes. Section 3 depicts Need of a Data Warehouse for the Educational institute. Section 4 explains the process of developing the data warehouse system for Educational. And section 5 gives focus on Challenges for building a data warehouses for an Educational Institute. Conclusion is given in section 6 and also a brief description of future work.

2. GENERIC DATA WAREHOUSE ARCHITECTURE:
Different data warehousing systems have different structures. Some may have a small number of data sources, while some may have dozens of data sources. In view of this, it is far more reasonable to present the different layers of data warehouse architecture rather than discussing the specifics of any one system.

A generic data warehouse is shown below in figure 1.

In general, all data warehouse systems have the following layers:

Data Source Layer: This represents the different data sources that feed data into the data warehouse. The data source can be of any format -- plain text file, relational database, other types of database, Excel file, etc., can all act as a data source.

Data Extraction Layer: Data gets pulled from the data source into the data warehouse system. There is likely some minimal data cleansing, but there is unlikely any major data transformation.

Staging Area: This is where data sits prior to being scrubbed and transformed into a data warehouse / data mart. Having one common area makes it easier for subsequent data processing / integration.

OLAP Layer: This is where OLAP cube is built from the data warehouse which is used further as a basis for various kinds of analysis.

Analysis Layer: This layer includes data mining, reporting and various analytical services done on the datawarehouse.

2.1 Extraction-Transformation-Loading (ETL) processes:
Extraction-Transformation-Loading (ETL) tools are a category of specialized tools with the task of dealing with data warehouse homogeneity, cleaning and loading problems.
ETL processes take place on the incoming data from data sources that have to be adapted and formatted before being inserted into the target data warehouse. Data Cleaning aims at finding and removing duplicate records, detecting inconsistent, wrong data, and manipulate sources of errors found in data [1], [7], and [10].

3. THE NEED OF A DATA WAREHOUSE FOR THE EDUCATIONAL INSTITUTE:

In today competitive environment where education has been privatized and cut throat competition is prompt, Educational institutes requires to be more organized and need to take better decisions. 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 Educational institute in a span of years?
- What are the student success ratios of Faculties, Departments and General ratio of the university?
- What is the total number of students that got scholarships from institute in past years?

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” [13]. Today, any reputed Educational Institute’s enrollments count in to thousands. In respect of These factors the real test for the management is cop up with the various needs of students and facing increased complications in academic processes. 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. Top management in the Educational institutes often need timely analysis reports to measure and monitor the performance of students, increase and decrease in students intake, skills of faculty, lab & building maintenance record, industrial visit of students, and the total placement record of students. They also need timely analysis reports to assist in making long-term decisions. It has been observed that most of the reporting and analysis, time was spent on collecting data from the various systems before the analysis can be made. Top Management wants and need more information, but analysts can provide only minimal information at a high cost within the desired time frames [6].

In order to provide information for predicting patterns and trends more convincingly and for analyzing a problem or a situation more efficiently, a data warehouse for this particular purpose is needed.

There are different interfaces, different data, and different data representations, duplicate and a lot of inconsistent information in an institution. The same data can be found in many different systems such as; instructor information can exist in four different systems like library, student information, placements record, fees, and labs and building maintenance record etc. Therefore, combining the information and generating reports or making decision is very difficult in such systems [12]. It’s allowing Institution experts to access to institutes data while maintaining security measures. There are many reasons why the Educational institutes in India considers data warehousing a critical need; the following should be taken into consideration:

- Decisions need to be made quickly and correctly, using all available data.
- Users are Institution domain experts, not computer professionals.
- The amount of data doubles in short time period, which affects response time and the sheer ability to comprehend its content.
- Competition is heating up in the areas of business intelligence and added information value.

So decision makers in the Educational institutes need more and more analytical information to capture the whole picture of their institute’s environment, and it is exactly the role of data warehouse to give them this global view and wide capability for analysis.

4. BUILDING THE EDUCATIONAL INSTITUTION DW

For making the Educational Institution DW, it starts with the sources analysis. Institutional data sources are found at institute in various formats and unrelated structures. 4.1 Analysis of data sources
The existing data sources at institute are described as follows:

a) Data Sources which covers different sectors in Institution such as:

- Record student’s intake in various courses in registrar office of institute.
- Record of industrial visit in TPO office.
- Information about each faculty qualification and skill record.
- Information of various accreditation and affiliation record at different time interval.
- Institution labs and building maintenance record.
- Number of total employees in institution.
- Student’s performance records.
- Data about Placement record.

b) Guest faculty visit records.

c) Data about cultural and technical or Faculty Development Program (FDP) events.

The data about hostels and catastrophic events is found in hardcopies format, so it needs for data entry step before going on its design as a DW cube.

4.2 Data warehouse requirements

The next step in our approach to build Institution DW is initiated by helping users to put their requirements. A set of discussion sessions were held to extract the requirements from the anticipated users. In addition to listing the users’
requirements, I classified them into functional and non
functional requirements to highlight their implementation
priorities.

4.2.1 Functional requirements.
Functional requirements represent what the institutional
DW system is expected to do, which include:

- **Data visualization**: Through the drawing of dynamic
  charts that change according to the change of the
  queries and the resulted values. Different types of
  charts will be used to represent the results of the end
  user queries.

- **Data browsing and dynamic queries**: The aimed data
  warehouse should facilitate the answering for the
  dynamic changes in the user queries in a simple and
  smooth way. This will be satisfied by using OLAP
  tools.

- **Data integration**: One of the main functions of the data
  warehousing is to integrate data from multiple sources.

- **Data validation**: The loaded data to the data warehouse
  should be tested and validated, so the values appear to
  be logic and consistent.

- **Remote access**: Through uploading the data warehousing
  cubes on a web site to enable the end user (top
  management) to access the data warehouse through the
  internet.

- **Security**: The security aspects will be achieved through
  using of different levels of authentication. Each user
  will see and browse the data warehouse according to
  his/her level of authority.

4.2.2 Non functional requirements.
Represents less important requirement, which include:

- **Simplicity**: Using and learning of the data warehouse
  should be simple and direct and doesn’t require any special
  skills.

- **Maintainability and scalability**: The data warehouse should accept the future changes and the expected updates according to the changes in the user requirements.

4.3 Institutional DW architecture
The following figure (2) shows the main steps for building
the data warehouse prototype. This architecture consists of
four phases. The first phase is for data preparation through;
data acquisition to collect the relevant data from different
data sources, data entry to transform the hardcopy data into
softcopy form, data validation to ensure the validity of the
incoming data and discover any potential errors resulting
from data entry process or anomalies appearing in the raw
data from original sources, and data coding to put it in
suitable format for design[18].

The second phase is the design of the data warehouse
through identifying the dimensions and measures, creating
the required ETL process to perform the needed operations
for data integration. The third phase is for building the
multidimensional subjective data cubes, and producing
sufficient number of aggregates to improve the effectiveness.

The final phase is for building the end user tools via which
the Institution experts will access the systems. These tools
include charting tools (to produce different kinds of chart),
OLAP tools (to give user diverse view for the same data
cube) and remote accessing tools (through building internet
web site).

5. CHALLENGES FOR BUILDING A DATA WAREHOUSES FOR
AN EDUCATIONAL INSTITUTE:
Since data generation and use are at its core, achieving the
Dream has developed a range of Materials to help colleges
understand the basic roles of institutional research (IR) and
information technology (IT) and how to align them with the
student’s data at Educational Institutes. A number of issues
must be addressed in building a data warehouse. Some are
associated with the creation of the warehouse and others
with its operation.

In the following, these are the key issues affecting
warehouse creation.

- Database conversion and schema integration is most
difficult task in building and maintaining a data
warehouse for an Educational institute. Also real time
database maintenance is most important factor in
building a data warehouses for an institution [17].

- Making data accessible to a broad audience is difficult.
Data systems can be vast, and data is unclear or not
tailored for specific audiences; thus, finding
information even in a centralized system can require
technical skills possessed by only a few individuals
[11],[12].

- Hiring data people who don’t share student success
goals or have good communication skills. In addition
to possessing essential technical skills, the best data
personnel are mission driven, have strong interpersonal
and broader social scientific skills, and know how to
distill and present data in ways that help internal and
external audiences understand complex issues [12].
• Hiding or ignoring “bad-news” data. When data shows poor program performance or learning outcomes, college leaders may be tempted to sweep the bad news under the rug so the institution will not be portrayed negatively or to maintain morale among their hardworking, underfunded staff. Bad-news data must be handled skillfully, but to put it aside inhibits the very culture of inquiry that Achieving the Dream colleges must create to make a difference for their students [16].
• Heterogeneous data challenge.
• Complexities of the organization;
• Vast number of stakeholders and appraisers.
• Resource limitations:

As operating costs increase, strain on the budget follows as institutional leaders struggle to balance the academic and human resources needs of colleges and universities [14]. Data transformation and integration is another area to be researched further as data warehouse is build up using data from heterogeneous sources therefore we should have efficient tools then available at present. This is one of the most important tasks in data warehousing as different databases have different schemas and format and it’s a prerequisite to convert them to similar format before loading into the data warehouse. The transformation of data with least error and least loss of information is still to go miles ahead. [15]

6. CONCLUSION:

Usually, analysis of the success of the data warehouse project is done considering the financial benefits against the investment. Since most of the educational institutes are nonprofit organizations and service oriented, the evaluation of the usefulness of the data warehouse can be done on the basis of its ability to meet user’s requirements. The academic data which was spread all across different sources has been loaded into single platform. The decision makers can extract information regarding three main components of the institute, namely Employees, Students and Infrastructure. Employee data mart can provide the users with the information such as career growth and attrition rate. Student mart can provide the information related to the student like best outgoing student considering his academic and non academic activities. Information regarding assets such as the investment in a particular financial year can also be accessed. In educational institute, decision makers ask “What are the expected results and benefits?” When making a data warehouse project rather than “What is the anticipated return on investment?” The data warehouse developed has met their expectations.

The scope of the study ends with building a data warehouse. Another useful concept OLAP is the main option for future enhancement. At the present stage, data required from data warehouse is extracted by writing queries. This can be improved by having excellent front end designed for reporting purpose.

REFERENCES: