Effective Decision Making in Higher Educational Institutions using Data Warehousing and Data Mining

[1]Alok Singh Chauhan, [2]Akhilesh Latoria, [3]Dr. Yashpal Singh
[1,2]Dept. of MCA, RKGIT, Ghaziabad, India
[3]Dept. of CSE, BIET Jhansi, India

Abstract
One of the biggest challenges that higher education institutions face today is to improve the quality of managerial decisions. Higher educational institute seeks more efficient technology to better manage and support decision making procedures or assist them to set new strategies and plan for a better management of the current processes. One way to effectively address the challenges for improving the quality is to provide new knowledge related to the educational processes and entities to the managerial system. This knowledge can be extracted from historical and operational data that reside in the educational university's databases using the techniques of data mining technology. Recently, there has been a growing trend to use data warehouses to support real-time decision-making about a university's day-to-day operations. This paper presents the capabilities of data warehousing and data mining in the context of higher educational system by proposing a guideline for higher education institutions to enhance their current decision processes and applying data mining techniques to discover new explicit knowledge which could be useful for the decision making processes.

Keywords
Data Mining, Data Warehouse, Decision-Making, Higher Education

I. Introduction

Colleges and universities constantly strive to achieve their institutional missions in the face of increasing competition and limited resources. They make crucial short- and long-term decisions as they do so, however critical questions remain:

- Do institutions have necessary facts to make well-informed decisions?
- Are some decisions being made based on intuition because relevant data is unavailable?
- If adequate data and analysis capabilities are not available to inform decision making, what risks does the institution assume?

Addressing these questions is important for improving decision making across the institution.

According to Luan, (2002), higher education institutions carry three duties that are data mining intensive: scientific research that relates to the creation of knowledge, teaching that concerns with the transmission of knowledge, and institutional research that pertains to the use of knowledge for decision-making [8]. With effective decision making capabilities however, higher education administrators and staff can more accurately identify trends, pinpoint areas that need improvement, engage in scenario-based planning and discuss fact-based decision making options and likely outcomes. These advanced decision making capabilities benefit an institution not only with respect to long-term planning, but also in assisting routine decision making that—when acted upon throughout an institution—can add up to tremendous performance improvements over time [4].

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 effective decision making for higher education institutions [6-7].

II. Related Research

Luan (2002) studied the impact of data mining on higher education. This study helped to gain insights about the existing higher education worldwide and its improvement from data mining perspective [8]. Delavari et al (2004) discussed a new model for using data mining in higher educational system. Waiyamai (2003) suggested that the use of data mining in education can help improve the quality of graduate students.

Ranjan and Malik (2007) proposed a framework for effective educational process using data mining techniques to uncover the hidden trends and patterns and making accuracy based predictions through higher level of analytical sophistication in students counseling process [1]. Talavera (2004) proposed to shape the analysis problem as data mining task. The author suggested that the typical data mining cycle bears many resemblances with proposed models for collaboration management and presented some preliminary experiments using clustering to discover patterns reflecting user behaviors.

Ma et al (2000) visualized that the education domain offers many interesting and challenging applications for data mining. First, an educational institution often has many diverse and varied sources of information. There are the traditional databases (e.g. students' information, teachers' information, class and schedule information, alumni information), online information (online web pages and course content pages) and more recently, multimedia databases. Second, there are many diverse interest groups in the educational domain that give rise to many interesting mining requirements.

III. Data warehousing Overview

Data warehouse is Integrated, Subject-Oriented, Time-Variant, Nonvolatile database that provides support for decision making [9]. A data warehouse is a repository of integrated information, available for querying and analysis. The basic idea behind the data warehousing approach is to extract, filter, and integrate relevant information in advance of queries. When a user-query arrives, the query does not have to be translated and shipped to the original sources for execution. Thus, warehousing can be considered an "active" or "eager" approach to information integration when compared to more traditional "passive" approaches, where processing and integration starts when a query arrives.

Traditionally, data warehouses have been used to analyze historical data. Recently, there has been a growing trend to use

data warehouses to support real-time decision-making about a university's day-to-day operations. The needs for improved query and update performance are two challenges that arise from this new application of a data warehouse. To address these needs, new data warehouse functionality is needed including:

1. Better access to early query results while queries are running and
2. Making the information stored in a data warehouse as fresh as possible.

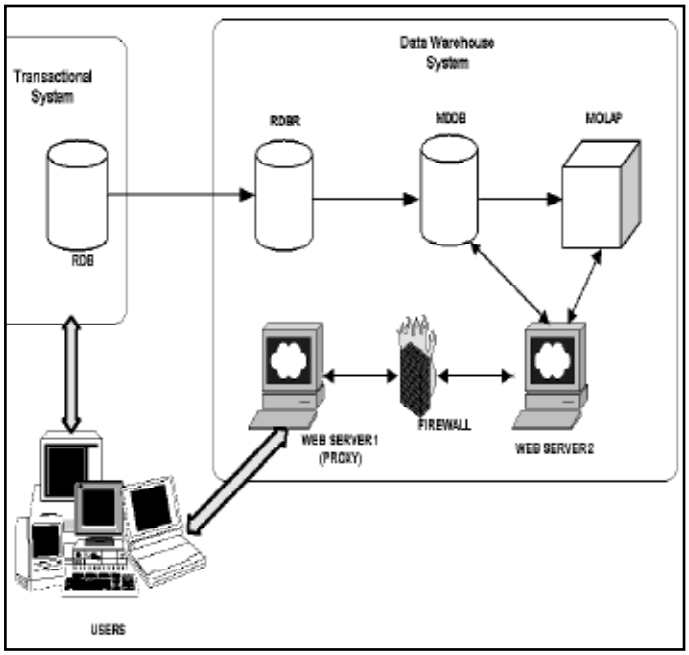

Fig. 1: The Architecture of the Data Warehouse System

IV. Data Mining Overview

Han and Kamber (2006) define data mining as the process of discovering 'hidden images', patterns and knowledge within large amount of data and making predictions for outcomes or behaviors [12]. Data mining methods can help bridge the knowledge gaps in higher educational system.

Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. Data mining uses a combination of an explicit knowledge base, sophisticated analytical skills, and domain knowledge to uncover hidden trends and patterns. These trends and patterns form the basis of predictive models that enable analysts to produce new observations from existing data [5, 9].

V. Data Mining Tasks

A. Associations, Mining Frequent Patterns

Association can be used to track students activities related to discipline programs, specializations and courses. The goal of the association rules is to detect relationships or associations between specific values of nominal attributes in large data sets [12].

Classification and prediction: The classification and prediction models are two data analysis techniques that are used to describe data classes and predict future data classes. The performance levels of the student can be classified as Good, Medium, or Poor. Using prediction, student's choice of specialization and whether the student will opt one particular course or not can be determined. To predict whether a student will show a certain type of behavior implies an assumption that the student belongs to certain type of student group and will therefore show the certain kind of behavior. The models of decision trees, neural networks based classifications schemes are very much useful for academic decisions [10-11].

B. Clustering

It is a method by which similar records are grouped together. Clustering is usually used to mean segmentation. An institution can take the hierarchy of classes that group similar students. Using clustering, students can be grouped based on educational background, age, areas of interest and specialization and so on. [10-11].

VI. Applications of Data Mining in Higher Education

Education is an essential element for the betterment and progress of a country. It enables the people of a country civilized and well mannered. Mining in educational environment is called Educational Data Mining, concern with developing new methods to discover knowledge from educational databases (Galit, 2007) (Erdogan and Timor 2005), in order to analyze students trends and behaviors toward education(Alaa el-Halees, 2009). Lack of deep and enough knowledge in higher educational system may prevent system management to achieve quality objectives, data mining methodology can help bridging this knowledge gaps in higher education system [2]. Some of the most likely places where data miners may initiate data mining projects are:

Alumni: All universities are committed to alumni management, but not all of them are using the best tools. Data mining is this area can potentially help identify those who are most likely to donate or participate in alumni related activities. For big donors, data miners can use the method of treating outliers.

Institutional Effectiveness: This synonymous with assessment of learning, but perhaps larger, is having sweeping impact on higher education. How do students learn best? What courses are often taken together? What learning experiences are most contributive to overall learning outcomes? These intriguing questions are good candidates for data mining [4].

VII. The Decision Making Environment in Higher Education

The decision making environment in higher education is very complex. Multiple entities and stakeholders need reliable access to specific information in order to make sound decisions. Thus, improving decision making capabilities means assessing and overcoming these challenges.

A. Demands for Improved Decision Making Capabilities

Higher education institutions are inundated with demands for information needed to support administrative, academic, planning, regulatory, legislative, research and operational interests. Varying constituents at all levels require dynamic views of information and customized analysis capabilities. Requests for more on-demand, comprehensive capabilities steadily increase as constituents strive to make informed strategic and day-to-day decisions. Each member in this community requires access to relevant information in order to guide and support their decision making. Administrators in particular often have a greater need for information access and analysis capabilities. The institution must be able to respond to the short-term, tactical needs of these constituents as well as meet their long-term, strategic requirements. In the competitive educational landscape, the demands of these and other groups will only continue to increase, with timeliness, collaborative enablement, data usability and multi-point access emerging as

key factors in improving data driven decision making.
The Table 1, below expands on several roles within the spectrum of higher education administration, and provides examples of the types of information that support decision making [6-7], by people within those roles:

Table: 1

Administration	Information that supports decision making
President's Office	• Operating performance metrics for departments or divisions • Financial information (e.g., reports on the use of federal and state funds, grants, status of the endowment) • Admissions and enrollment statistics
Finance	• Budget information (e.g., from the Institutional level down to individual transactions, comparisons of financial commitments with overall budget, projections of budget short-falls) • Endowment funds information (e.g., estimates in the growth of scholarship and/or endowment funds) • Vendor purchasing information (e.g., spending information to support strategic sourcing initiatives)
Provost's Office	• Enrollment information (e.g., monitoring historic enrollment trends and contrasting with current projections, projecting how many course sections will be needed based on specific enrollment figures) • Staffing information (e.g., estimating early retirements, position vacancies and number of new hires required)
Facilities and Infrastructure	• Facilities utilization information (e.g., information for facilities scheduling, maintenance scheduling and planning)
Information Technology	• Service level information (e.g., reviewing network performance, helpdesk responsiveness and mean time to resolution) • Service usage pattern information (e.g., planning for needs in IT environment and services based on institution needs and usage patterns)

B. Challenges for Improving Decision Making Capabilities

As higher education institutions try to meet the information demands of their constituents, their data warehousing initiatives can start off with high aspirations, but often never get off of the ground, or fail to deliver expected services. In order to implement a successful data warehousing initiative, institutions must first consider and reconcile four critical challenges [13] given in Table 2 below.

Table: 2

Challenge	Explanation
Data required for decision making is undefined	Without clearly understanding the users' requirements for information, data sources cannot be identified and utilized for the users' analyses and decision making needs.
Data is everywhere and nowhere	When data resides on multiple systems across many different departments, it's difficult to increase access and make intelligent use of that data.
Users lack the ability to conduct a deeper analysis of data	Institutions are locked into manual, paper-based methods for reviewing performance, meaning users can't drill down to discover cause-and-effect patterns and make decisive improvement decisions.
Demand is high and resources are low	The demand for decision making information exceeds the time, money and IT resources available to meet the requirements.

Data warehousing will provide consolidated and customizable views of institutional and departmental data, structured to optimize analysis, reporting and decision making capabilities [7]. In support of these objectives, data and information will be extracted from multiple sources, making available more effective and efficient queries on a scalable and sustainable platform. Data will be turned into high-quality information to meet analysis, reporting and decision making requirements at all levels. Interactive content will be seamlessly delivered to anyone with information access privileges, in any role in the institution—from deans to department heads, faculty to senior administrators—anytime, anywhere [6]. The next steps in addressing data warehousing challenges are to ensure the development of a scalable, sustainable platform for data warehousing, and a clear roadmap to follow for the development and implementation of improved decision making capabilities.

VIII. A Framework for Improving Decision Making Capabilities

Through a data warehousing implementation, administrators can quickly realize the productivity and efficiency gains they envision. They can readily create the initial analysis tools that will allow users throughout the university to easily access the data that is most important to them for decision making purposes [7]. The five guiding principles below and the program lifecycle that follows can provide the framework that will help institutions plan and implement improved decision making capabilities.

A. Five Guiding Principles for Developing Decision Making Capabilities

1. Provide an Enterprise-Class Data Infrastructure

Institutions need to understand and prioritize the demands made by all of their constituents. This includes not only data access and system responsiveness, but also the stability and security within the infrastructure to operate effectively, efficiently and securely.

2. Design a Comprehensive and Scalable Solution

Higher education institutions need the flexibility of building out their data driven decision making capabilities as budget and resources allow. The solution should be designed to easily provide increased capability as data sources expand and user needs evolve.

3. Utilize Familiar Tools

With limited IT resources to support ongoing initiatives, institutions should design a data warehousing solution that is easy to maintain, requiring minimal user training and on-going support.

4. Create True Closed-Loop Analysis Capability

Closed-Loop Analysis (CLA) is a process intended to set improvement goals, monitor progress, assess impact or effectiveness and realign objectives as required. It becomes a real-time process when data is continuously collected and evaluated. Decision making capabilities can be greatly enhanced with closed-loop analysis processes and enabling tools.

5. Design a Maintainable and Sustainable Solution

Decision making information requirements will evolve over time as users become accustomed to making use of information to improve administrative performance. These new information requirements can place additional demands on an institution's information technology resources, and institutions must be prepared to sustain their solution.

With these five guiding principles in mind, institutions should consider a program lifecycle management framework as they move towards implementation.

B. Program Lifecycle Management

As data and decision making needs change, institutions will constantly cycle through the program lifecycle stages. The lifecycle stages and supporting activities are highlighted below:

1. Assess Current Capabilities and Define Operational Needs

During this stage, the information requirements of each constituent are determined and prioritized. This includes not only the data access and analysis requirements, but also the system performance requirements. In addition, the gaps between the current and desired capabilities are determined. The effort and activities required to close the gaps are captured in a master program plan, which most likely will result in "waved" releases of increased capability.

2. Design Data and System Architecture

The data, system (hardware, software and networking) and security components are architected and the total solution is designed. The design follows a set of standards defined to reduce overall maintenance costs and accelerate future improvement initiatives. During subsequent capability improvement projects, institutions need only refine their standards, not redefine. During this stage, proof-of-concept testing can greatly enhance the ultimate capabilities and potential for success of the project.

3. Develop Solution

During this stage, the individual components are developed, configured and tested. This includes individual components, as well as system-level and stress testing. Integration and legacy interfaces are also created and tested. Security procedures should be well exercised after all other testing is complete and before progressing to the next stage.

4. Implement

The system is taken "live." A phased rollout is recommended to bring increased volume on the system in a steady and controlled manner. Required training and support is coordinated.

5. Evolve

Ongoing maintenance and support begins. In addition, feedback from constituents is captured and retained. Finally, lessons learned during the previous stages of the lifecycle are incorporated into future upgrades and waved releases in the master implementation plan.

By considering the five guiding principles and the program lifecycle stages, institutions can better grasp the journey that lies ahead in developing their data driven decision making capabilities.

IX. Conclusion

In this paper we have discussed that data warehousing and data mining can be useful in various aspects of an institute life cycle. The purpose of this study was to investigate ways in which principals build successful data-driven information systems within their Institutions. Higher education institutions can dramatically improve decision making capabilities by implementing successful data warehousing initiatives that address identifiable challenges, set attainable goals and follow the framework of guiding principles and program lifecycle management. Such activities will definitely guide to better decision making procedures and will improve the quality of instructions.

References

[1] Ranjan, J., Malik, K.,"Effective Educational Process: A Data Mining Approach", VINE, Vol. 37, Issue 4, pp. 502-515, 2007.

[2] Shyamala K., Rajagopalan S. P.,"Data Mining Model for a better Higher Educational System", Information Technology Journal, Vol. 5, No. 3, pp. 560-564, 2006.

[3] Agrawal, R., Imielinski, T., Swami, A.,"Database Mining: A Performance Perspective", IEEE Transactions on Knowledge and Data Engineering, pp. 914-925, December 1993

[4] Delavari N, Beikzadeh M. R, Shirazi M. R. A.,"A New Model for Using Data Mining in Higher Educational System", Proceedings of 5th International Conference on Information Technology based Higher Education and Training: ITEHT '04, Istanbul, Turkey, 2004.

[5] Fayyad, U.M.,"Data mining and knowledge discovery: making sense out of data", IEEE Expert 11, October 1996, pp. 20-25.

[6] Gerardette Furlow,"Integrating Data Warehousing and Data Mining Solutions at Community Colleges to Improve Administrative Functions and Academic Programs", 2001.

[7] Guan, J., Nunez, W., Welsh, J. F.,"Institutional strategy and information support: the role of data warehousing in higher education", Campus-Wide Information Systems, 19(5), pp. 168-174, 2002.

[8] Luan, J., A. Serban, J. Luan,"Data Mining and Its Applications in Higher Education", Knowledge Management: Building a Competitive Advantage fir Higher Education. New Directions for Institutional Research, No. 113. San Francisco, CA: Jossey Bass, 2002.

[9] Berson,"Data Warehousing, Data-Mining & OLAP", TMH.

[10] Berry, J. A, Lindoff, G.,"Data Mining Techniques", Wiley Computer Publishing, 1997.
[11] Bhavani, Thura-is-ingham,"Data-mining Technologies, Techniques tools & Trends", CRC Press.
[12] Han, J., M. Kambler,"Data Mining: Concept and Techniques", Morgon Kaufmann Publishers, pp. 4-27. New Delhi, 2003.
[13] Widom, J. Research Problems in Data Warehousing. Stanford University Database Group; 1995. [Online] Available: http://www.db.stanford.edu/pub/papers/warehouse-research.ps
[14] [Online] Available: http://www.microsoft.com/education

Alok Singh Chauhan obtained his B.Sc. Degree in 2003 from University of Allahabad, MCA Degree in 2006 from U.P. Technical University, Lucknow and is currently pursuing Ph.D. in Computer Application as a part time Research Scholar in Bundelkhand University, Jhansi under esteemed guidance of Dr. Yashpal Singh. His areas of interest are Data Mining and Optimization techniques. He is presently working as Asstistant Professor in Raj Kumar Goel Institute of Technology, Ghaziabad affiliated to Mahamaya Technical University, Noida, UP (INDIA)

Mr. Akhilesh Latoria obtained his BCA Degree in 2004 from JiwaJi University, Gwalior, MCA Degree in 2007 from RGPV, Bhopal and is currently pursuing Ph.D. in Computer Science as a part time Research Scholar in Mewar University, Chittorgarh. His areas of interest are Data Mining and Pattern Recognition. He is presently working as Asstistant Professor in Raj Kumar Goel Institute of Technology, Ghaziabad affiliated to Mahamaya Technical University, Noida, UP (INDIA)

Dr. Yashpal Singh obtained his B.Tech (CS) Degree in 1993 from Bundelkhand University Jhansi, M.E. Degree in 2000 from University of Allahabad and Ph.D. (CSE) Degree in Computer Science and Engineering in 2008 from Bundelkhand University, Jhansi. His areas of interest are Neural Networks, Image Processing and Data Mining. He is presently working as Reader and Head with the Computer Science and Engineering Department, Bundelkhand Institute of Engineering and Technology, Jhansi -284128 (INDIA).