Business Intelligence in E-Learning

(Case Study of Iran University of Science and Technology)

Mohammad Hassan Falakmasir[1], Jafar Habibi[2], Shahrouz Moaven[1], Hassan Abolhassani[2]
Department of Computer Engineering
Sharif University of Technology
Tehran, Iran
[1]{Falakmasir, Moaven}@ce.sharif.edu
[1]{Jhabibi, Abolhassani}@sharif.edu

***Abstract*— Nowadays, e-learning platforms are widely used by universities and other research-based and educational institutions. Despite lots of advantages these educational environments provide for organizations, yet there are many unresolved problems which cause instructors and training managers with some difficulties to get proper information about the students' learning behavior. On one hand, lack of tools to measure, assess, and evaluate the performance of learners in educational activities has led the educators to fail to guarantee the success of learning process. On the other hand, strict structure of learning materials prevents students to acquire knowledge based on their learning style. Consequently, developing tools monitor and analyze the learner's interaction with e-learning environment is necessary. Business intelligence (BI) and On Line Analytical Processing (OLAP) technologies can be used in order to monitor and analyze the learner's behavior and performance in e-learning environments. They can also be used to evaluate the structure of the course content and its effectiveness in the learning process. This article investigates the use of business intelligence and OLAP tools in e-learning environments and presents a case study of how to apply these technologies in the database of an e-learning system. The study shows that students spend little time with course courseware and prefer to use collaborative activities, such as virtual classroom and forums instead of just viewing the learning material.**

Keywords- E-Learning; Educational Data Mining; Business Intelligence; Data Warehouse; OLAP; Intelligent Data Analysis

I. INTRODUCTION

Today, most universities and educational institutions use e-learning platforms to offer their educational material benefiting from their advantages such as work "any-time, anywhere", use of collaborative tools, support different styles of learning, etc. However, these e-learning platforms do not cover all teaching aspects since they do not usually provide teachers and instructional designers with tools which allow them to monitor and assess all the activities performed by learners [1,2,3,4].

These environments provide the professor with access summary information such as the date of the first and last connection, the number of visited pages according to the category specified by the platform (not by the professor), or the number of messages read/sent by each learner; the total number of page visits, the average time spent on each page but, globally, per page not by learner. As can be deduced, this information is not enough to analyze the behavior of learners and their evolution. Even more, when the number of students and the diversity of interests are high, the professor has serious difficulties to extract useful information.

Business Intelligence techniques applied to the database of Web-based Educational Systems could help instructors and other educational experts to generate statistics, analytical models, and uncover meaningful patterns from these huge volumes of data. In this article, a framework for applying business intelligence in e-learning environments has been proposed, which increased both flexibility and performance of e-learning environments. Hence, on one hand, the proposed environment enables educational technologists to identify, analyze and monitor relevant aspects of instruction, such as different style, paths, and strategies of learning. On the other hand, such parameters may be used to adapt the learning process to each individual learner and improve the performance of the learning process.

The reminder of this paper is organized as follows. Section 2 briefly describes business intelligence principles. A brief history of the related work is represented in the third section. A description of the e-learning system and the proposed architecture are represented in section 4, moreover, three steps to deploy the data warehouse for the e-learning environment are demonstrated and some valuable results are presented. Finally, the last section contains the conclusion and the future work.

II. BACKGROUND

Nowadays, the role of business intelligence as an important strategic tool for business management and making efficient decisions is significant. Additionally, because of prompt progress of business competitors, the ability to obtain information in real-time and reduce decision-making cycle times has become increasingly critical in recent years. Taking advantage of business intelligence in companies can help them to improve business performance by giving them the opportunity to gain insight into their business and make better decisions [5].

In short, business intelligence helps companies to gain a comprehensive and integrated view of their business and facilitate better and more effective decision-making and other

benefits such as having access to the summarized and distributed relevant information on time. Moreover, companies are provided with a framework capable of introducing and measuring business key performance indicators while analyzing its process and understanding their behavior. In order to carry out these tasks, Business Intelligence uses a wide range of techniques and technologies: the data warehouse as an integrated repository of strategic information, the OLAP (On-Line Analytical Processing) technology for the exploration of information under different perspectives, dashboard, scorecard and reporting tools for the analysis and visualization of information and trends, and data mining techniques to discover meaningful patterns and rules in large volumes of data by automatic or semi-automatic means. Before continuing, it should be said that some authors, such as Kimball [6] consider that the data warehouse is the platform for business intelligence; however other authors such as Inmon [5] consider that the data warehouse is simply the database where the business data are consolidated and stored. In this work, we follow Kimball's idea.

III. Related Work

Lots of works have been performed in the literature in order to satisfy mentioned problems in e-learning environments and overcome their weaknesses. Student's usage logs are the main starting point to perform such analysis. Usage statistics can be extracted by standard tools designed to analyze web server logs or specific tools developed to satisfy educational needs. In this way, some tools, like GISMO [7] were proposed which monitors activities of students in a popular Learning Management System (LMS) called Moodle [8]. The tool extracts tracking data from Moodle log file and represents results in graphical format. Moreover, it provides professors to the information in several different ways such as reports and graph representation.

As another tool, CourseVis [1] can be mentioned which can track students' data that is collected through web log files of the LMS web server. Sinergo/ColAT [9] is another tool that offers interpretative views of the activity developed by students in a group learning collaborative environment and resembles the learning process of students. Another tool presented [10] in order to show the tutor-student interaction in a hierarchical representation.

Finally, the most recent tool is MATEP [11] which presents a web interface for the instructor providing a set of reports according to his requirements. According to instructors' opinion, this tool helps them to gain a more accurate knowledge of what is happening in their courses since it allows them to analyze and visualize data with different level of detail and perspectives, discovering student behavior patterns' and understanding how their courses are used. That means that they have the quantitative and qualitative information available to take improvement actions about their courses.

IV. The Platform

The E-Learning Department of the Iran University of Science and Technology (IUST) started its services in the spring semester of 2004 with about 700 students and is currently serving about 1,800 students in two Bachelors' and three Masters' programs. The instructional plan in this department is designed in a way that main the learning materials are developed in the form of multimedia courseware and can be accessed by students in a weekly manner. In addition, the teacher can evaluate the process of learning by giving the students with assignments and quizzes. Finally, the students, having gained proper perception about the course concepts, participate in a virtual class session and collaborate with the teacher and other students on the problems and learning materials. The teacher can also provide them with complementary information about the content and get some feedback about the level of each student's knowledge. For supporting the "any-time, any-where" promise of e-learning, all the sessions are recorded and archived for the students who cannot participate the online classes.

After three years of using a commercial e-learning system and in response to a perceived demand for more reliable and flexible delivery of courseware, the E-Learning Center of IUST started to use Moodle as its LMS since March 2006. Moodle is one of the most popular and widely used e-learning platforms in Iran and all around the world. The transition from commercial platforms to open source systems, such as Moodle, is a growing trend in all around the world and the spread of these online learning environments are under continuous evolution. Moodle has been used as a platform for sharing useful information, documentation, and knowledge management in a lot of research project, yielding important benefits to researchers [12,13]. Data mining techniques have also been used as complementary system to Moodle, where the results where the results are achieved through the use of association rule mining, classification, clustering, pattern analysis, and statistical methods [15].

Moodle keeps detailed log of all activities that students perform [12]. It logs every click that students make for navigational purposes and has a modest log viewing system built into it. Log files can be filtered by course, participants, day and activity, but they are not suitable for analytical applications. Fortunately, Moodle does not store logs as text files. Instead, it stores the logs in a relational database. There are about 145 interrelated tables in Moodle database but not all of them contain the information we need for analyzing the data and building up the data warehouse.

V. Main Idea

Although Moodle presents several reports about the students' activities, they are not flexible enough to satisfy the instructors' needs to observe their interactions with the system. The teacher has access to a summarized report about students such as the date of the first and last connection of students and the number of visited pages. The information about each learning activity is also available, but according to the categories specified by the system, not the professor. Consequently according to some interviews with instructional technologists and training managers of the IUST E-learning Center a list of data elements and analytical dimensions were defined and corresponding information is extracted from the Moodle database to answer their questions.

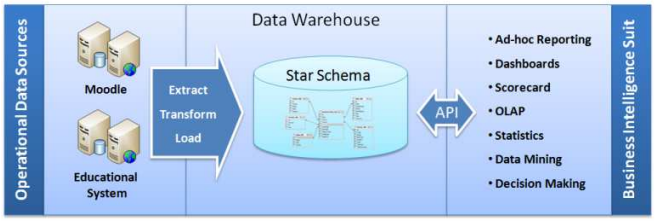


Figure 1: Proposed Architecture

Figure 1 illustrates the architecture of the proposed solution. As one can observe, the data from multiple sources such as educational system, LMS, and other legacy databases is integrate in the form of dimension and fact tables according to star schema and can be accessed and analyzed through the usage of API functions.

The first step to create a BI solution is the identification of business requirements and their associated values. As it is mentioned before, in this step according to some interviews with instructional experts and institutional managers, a list of data elements and desired reports that would help to answer the analytical questions were gathered and recorded as business requirements document. This document consists of questions such as: When do students connect to the system? How long do they spend viewing learning materials? How often they use collaborative tools? Which learning activity they prefer to participate? Which learning resource they prefer to visit?

The second step is designing a single, integrated, easy-to-use, high performing information model that gathers the identified business requirements and building a dimensional schema [5]. Dimensional schema is made up of a central fact table and its associate dimensions. It is also called star schema because it looks like a star with the fact table in the middle and the dimensions serving as the points on the star. According to Star Schema, in proposed solution two fact tables have been designed:

- CourseActivity_Fact: This fact table contains the details about students' learning activities in a course in each semester such as the number of: resource view, forum read, forum post, message read, message write, virtual classroom participation, recorded session review, etc.
- VirtualClassroom_Fact: Having focus on virtual classroom sessions, this fact table contains the information about how students participated in the virtual classroom sessions. It contains details about students' activities in the virtual classroom such as the number of: posted text message, raise hands, change status, broadcast audio, etc.

As can be observed, all these measurements or facts are numeric and additive, meaning they can be summed up across all dimensions. CourseActivity_Fact contains the records of the all activities performed by a learner in a course. VirtualClassroom_Fact gathers the information about every virtual classroom session conducted in the e-learning environment. The level of detail or grain of this fact table is a row for each completed learner session. A learner session is defined as the time spent by a student since he or she connects to a certain virtual class until he leaves it. There are also four dimension tables designed to hold the attributes that describe fact records:

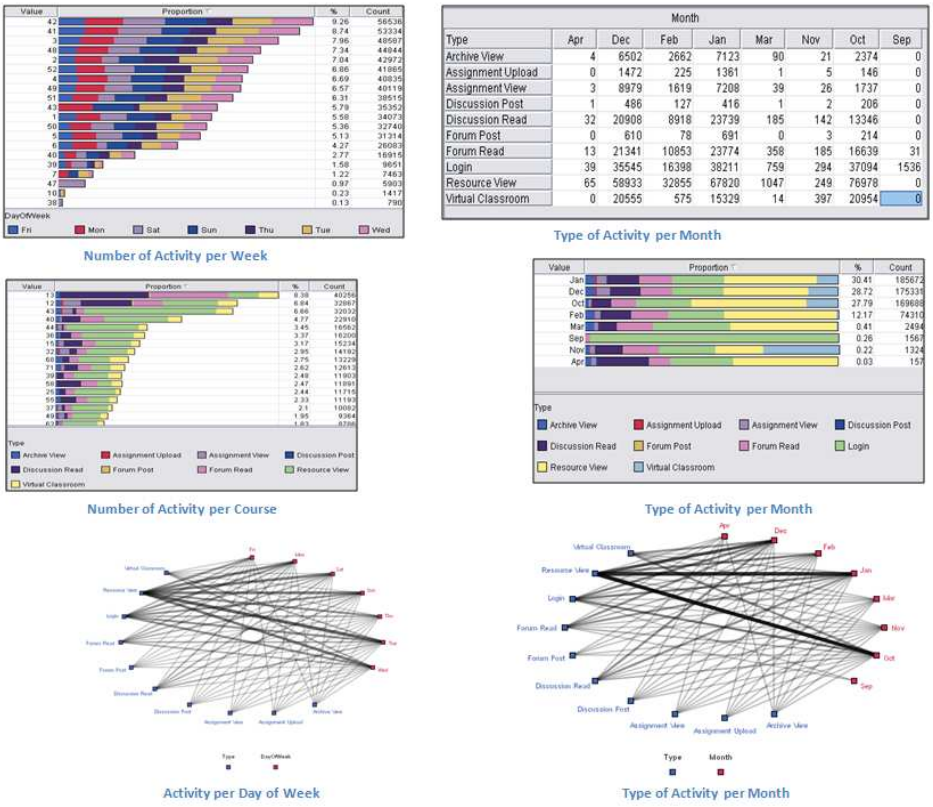
- Date_Dim: It gathers each day of year with all characteristics such as number of day, month, day of week, week, year, and so on.
- Learners_Dim: it collects the students' information, such as name, gender, major, city, field of study, etc.
- Session_Dim: it stores an identifier per each session, containing IP address and its relevant location.
- Courses_Dim: it contains the information about each course such as name, type, semester, etc.

These tables are the foundation of dimensional modeling, containing descriptive information relevant to analyze the fact table attributes from different perspectives.

Once the dimensional schema is designed, the data stage and ETL processes must be defined and programmed. The ETL process of proposed solution was quite simple because all the data is stored in relational database. However some preprocessing steps have to be performed for transforming data into suitable shape (e.g. star schema). ETL process normally is a manual process in which the administrator has to apply a number of general data preprocessing tasks such as, data cleaning, user identification, session identification, path completion, transaction identification, data transformation and enrichment, data integration, data reduction [6]. In our case, data preprocessing of LMS is a little more simple due to user authentication in which logs have entries identified by users.

Although the amount of work required in data preparation is less, it is necessary to build some new tables according to star schema and convert the data from relational into multi dimensional format. For this reason, we have to transform the transactional data from several tables of Moodle database into fact and dimension tables and integrate the most important information for our objectives. After these steps, data is prepared for multi dimensional analysis in the BI environment. The only thing remaining to do is using several analytical tools such as dashboards, scorecards and dynamic reporting tools for investigation and analysis of the behavior of students in e-learning environment.

There are several products to implement a BI solution. In the past, while the BI market was strictly dominated by closed source and commercial tools, the last few years were characterized by the birth of open source solutions: first as single BI tools, and later as complete BI platforms. An Open Source BI platform provides a full spectrum of BI capabilities within a unified system that reduces the overhead for the development and management of each application. Oracle, SQL Server, Sybase or DB2, to name some, are commercial closed source Database Management Systems which support data warehousing and OLAP technology. On the other hand, there are some open source platform (namely JasperSoft, Pentaho and SpagoBI) which allows educational institutes to begin immediately to deploy the architecture without any need to purchase commercial tools, but it requires a fair amount of customization.



Month								
Type	Apr	Dec	Feb	Jan	Mar	Nov	Oct	Sep
Archive View	4	6502	2662	7123	90	21	2374	0
Assignment Upload	0	1472	225	1361	1	5	146	0
Assignment View	3	8979	1619	7208	39	26	1737	0
Discussion Post	1	486	127	416	1	2	206	0
Discussion Read	32	20908	8918	23739	185	142	13346	0
Forum Post	0	610	78	691	0	3	214	0
Forum Read	13	21341	10853	23774	358	185	16639	31
Login	39	35545	16398	38211	759	294	37094	1536
Resource View	65	58933	32855	67820	1047	249	76978	0
Virtual Classroom	0	20555	575	15329	14	397	20954	0

Figure 2: Sample Reports Generated by the Proposed Solution

VI. Results

The study examined the activity logs of 1,300 students, in nearly 100 courses over the 16 week fall semester of 2008. The key findings showed that 80% of all students accumulated less than one hour a week with course materials in LMS and most looked at large portions of their content for less than one minute. Contrarily, about 65% of students participated in virtual classroom sessions where they can directly interact with their teacher via instant messaging and voice chat.

Around 60% of students spent less than one hour a week in LMS within the first 10 week, but the percentage shrinks to 20% in remaining 6 weeks. There are some active students (about 5%) that spent more than 400 hours in LMS within the 16 week semester and some passive students (about 7%) that spent less than 50 hours.

Figure 2 demonstrates some of the reports generated using the proposed solution. Using such information, instructors can adopt some decisions to improve performance of students and avoid accumulation of students' works for the end of semesters which avoids students to mostly focus on their exam and review their courses. For example, suitable and uniform distribution of assignments during the semester and forcing students to upload their assignments exactly at the prescribed time can be considered as an appropriate decision for achieving the mentioned goal.

VII. Conclusion and Future Works

In this paper, in order to overcome shortcomings of traditional e-learning environments in making appropriate decisions, the application of Business Intelligence and Data Warehousing tools in the field of e-learning was presented. Moreover, the architecture proposed for these systems, tries to solve the lack of analytical and subjective reporting tools in a widely spread, open-source, learning management system and, moreover, provide instructors with detailed reports about progression of students and give them the ability to track and assess the student performance and evaluate the design of their virtual courses in order to take suitable managerial decisions.

The study shows that students spend little time with course materials online and about 80% spend less than two hours a week viewing multimedia courseware and they just have participated in virtual classroom sessions. It is also inferred that the students prefer to use collaborative activities, such as instant messaging and forums, instead of just viewing the learning material.

However, expanding the period and the number of students involved in analytical study and applying data mining methods and association rule mining techniques to extract other unresolved meaningful patterns and rules among the student's data usage can be considered as future work. Moreover, applying clustering and classification algorithms in order to predict the performance of new students according to their behavior is other future work.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the management and educational technologists in the E-learning Center of the Iran University of Science and Technology for providing the access to the database of e-learning environment and giving valuable suggestions and comments.

REFRENCES

[1] R. Mazza and V. Dimitrova, "CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses," *International Journal of Human-Computer Studies*, vol. 65, no. 2, pp. 125–139, 2007.

[2] Cristobal Romero and Sepastian Ventura, *Data mining in E-Learning. Advances in Management Information*.: WIT Press, 2006, vol. 4.

[3] M. E. Zorrilla, D. Marin, and E. Alvarez, "Towards virtual course evaluation using Web Intelligence," in *EUROCAST 2007 LNCS*.: Springer, Heidelberg, 2007, pp. 392–399.

[4] O. Zaïane and J. Luo, "Web usage mining for a better web-based learning environment," in *Proceedings of conference on advanced technology for education*, Banff, Alberta, 2001, pp. 60-64.

[5] W. H. Inmon, *Building the Data Warehouse*. Chichester: Willey & Son, 2002.

[6] R. Kimball and J. Caserta, *The Data Warehouse ETL Toolkit.* Chichester: John Wiley & Sons, 2002.

[7] (2010) GISMO. [Online]. http://gismo.sourceforge.net/

[8] (2010) Moodle. [Online]. http://www.moodle.org

[9] N. Avouris, V. Komis, G. Fiotakis, M. Margaritis, and G. Voyiatzaki, "Logging of fingertip actions is not enough for analysis of learning activities," in *Proccedings of Workshop Usage Analysis in learning systems*, Amsterdam, 2005.

[10] J. Mostow et al., "An Educational Data Mining Tool to Browse Tutor-Student Interactions: Time Will Tell!," in *Workshop on educational data mining*, 2005, pp. 15–22.

[11] M. E. Zorrilla and E. Alvarez, "MATEP: Monitoring and Analysis Tool for E-learning Platforms," in *8th IEEE International Conference on Advanced Learning Technologies*, Santander, Spain, 2008, pp. 611-613.

[12] J. Cole, *Using Moodle*.: O'Reilly, 2005.

[13] A. Uribe-Tirado, L. M. Melgar-Estrada, and J. A. Bornacelly-Castro, "Moodle learning management system as a tool for information, documentation, and knowledge management by research groups," *Profesional de la Información*, vol. 16, no. 5, pp. 468–474, 2007.

[14] Cristobal Romero, S. Ventura, and E. Garcia, "Data mining in course management systems: Moodle case study and tutorial," *Computers & Education* , vol. 51, no. 1, pp. 368–384, 2008.

[15] W. H. Rice, *Moodle e-learning course development. A complete guide to successful learning using Moodle*.: Packt Publishing, 2006.

[16] J. Mundy, W. Thorthwaite, and R. Kimball, *The Microsoft Data Warehouse Toolkit: with SQL Server 2005 and the Microsoft Business Intelligence Toolset.* Chichester: Wiley Publishing Inc., 2006.

[17] C. J. White, "Corporate Performance Optimization Guide.Intelligent Business Strategies," Oracle Corporation, January 2003.

[18] E. Thomsen, *OLAP Solutions: Building Multidimensional Information Systems*, 2nd ed. Chichester: John Wiley & Sons, 2002.

[19] Cristobal Romero and S. Ventura, "Educational Data Mining: A Survey from 1995 to 2005," *Expert Systems with Applications*, vol. 33, pp. 125-146, 2007.