Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A FRAMEWORK FOR EDUCATIONAL DATA WAREHOUSE (EDW) ARCHITECTURE USING BUSINESS INTELLIGENCE (BI) TECHNOLOGIES

[1]AZWA ABDUL AZIZ, [1]JULAILY AIDA JUSOH, [1]HASNI HASSAN, [1]WAN MOHD RIZHAN WAN IDRIS, [2]ADDY PUTRA MD ZULKIFLI, [2]SHAHRUL ANUWAR MOHAMED YUSOF

[1]Faculty of Informatics and Computing, Sultan Zainal Abidin University (UniSZA), Tembila Campus, 22200 Besut, Terengganu, Malaysia

[2]Faculty of Design Arts and Engineering Technology, Sultan Zainal Abidin University (UniSZA), Gong Badak Campus, 21300 Kuala Terengganu, Terengganu, Malaysia

E-mail: {azwaaziz, julaily, hasni, wmrizhan, addyputra, shahrulanuwar}@unisza.edu.my

ABSTRACT

Business Intelligence (BI) is the process of getting the right information to the right decision makers at the right time and in the right format. It is a platform that supports analysis, reporting and decision making. Educational Business Intelligence (EduBI) architecture utilizes BI technologies to integrate various sources of academic data into a single repository (Educational Data Warehouse – EDW). BI is a useful tool since it reinforces the process of performance and analysis evaluation that is required within all levels of educational environment. Nowadays, there are various BI products ranging from simple reporting technologies to sophisticated BI platforms. Selection of a BI tool deemed appropriate for a particular task may turn out to be difficult; hence careful considerations must be made in the selection process. This paper proposed an EDW architecture that employs the integration of proprietary and open source BI tools.

Keywords: *Data Warehouse, Educational Business Intelligence, ETL, Educational Intelligence*

1. INTRODUCTION

Managing an institution involves access to information in order to monitor the institution actions [1]. Growing capacities of client populations can bring difficulties to an institution due the complexity in the process of handling various data. In an educational setting, there are a lot of data to be recorded into various sources systems, on a daily basis. While some data are not fully utilized to perform an analysis, some are isolated from other systems; leading to the 'island of information' dilemma.

Educational Data Mining EDM is an emerging discipline for developing methods to explore unique types of data from educational context [2]. In fact, EDM is an application of data mining techniques implemented in the area of education to obtain better comprehension on students' learning processes and acknowledging the ways they participate in it, in order to improve the quality of the educational system [3]. Previous research related to EDM focus on applying data mining algorithm for educational purposes such as Merceron & Yasef [4] who used the Association Rules (ARs) technique to analyse learning data. In another study, Zorrilla et al. [3] analysed Learning Management System (LMS) data using Decision Support Technique to understand the pattern of students' behaviour.

Most of the prior researches obtained similar outcomes when studying the nature of academic data. In conventional teaching environments, educators are able to obtain feedback on student learning experiences via face-to-face interactions with students, enabling continuous evaluation of their teaching programs [5]. The evaluation records contain data from student' assignments, participation in classes such as quizzes, exams, test, and presentations, and other methods of evaluation.

Nowadays, most face-to-face interactions have been replaced by virtual learning system. Students' data are recorded on a web platform using LMS platform such as Moodle system. Web-based learning environments are able to record most learning behaviours of the students, and are hence able to provide a huge amount of learning profile [2]. This serves as an opportunity for data mining researches where academic data can be analysed and the findings could be used as a contribution to improve an educational system. Since in Malaysia the government has included E-Learning as one of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the key area of National Higher Learning Institution Strategic Plan, it offers a huge opportunity to analyse educational data in our country.

In order to perform critical analyses on various types of data, the effort to integrate those data into a single repository for information is imperative. Data Warehouse (DW) presents itself as the best solution to be applied in academic area. DW is the process of taking data from legacy and transaction database systems and transforming it into organized information in a user friendly format to encourage data analysis and support fact- based business decision making [6]. DW has proved its usefulness in traditional business analysis and decision making processes [7]. It consists of a combination of technologies and techniques to produce comprehensive reports for decision makers. In academic context, decision makers can be lecturers, deans, counsellors and other authorities responsible in the attempt to provide the best solution to improve students' academic achievements.

However there are two critical issues to be considered prior to the implementation of DW in an educational setting. First, it relates to the amount of data in an educational institution. Academic data is dynamic in nature where it changes from time to time. In addition, the growth of data is hard to measure and the process of filtering to get the right information is critical. As an example, one lecturer may have 60 students per semester. There are a lot of data regarding the students such as class attendance, results on coursework, tests and final examination, and also on co-curriculum activities that might contribute to students' final results.

The second issue is regarding the cost of having a DW in an academic institution. DW projects required various tools to be implemented such as BI tools, Extract, Transform, Load (ETL) tools and Database Management System (DBMS) of DW itself. Organizations need to spend a lot of money to implement DW technologies in their places hence it becomes an impediment to non-profit organizations such as Public Higher Learning Institution and schools. To overcome this problem, we proposed a solution for the implementation of DW architecture in academic environment using several technologies that is combination of open sources tools and proprietary software.

2. RELATED WORKS

Introduction to Educational Intelligence

Two major components to ensure the success of DW projects are BI and ETL tools. BI can be defined as a discipline that combines services, applications, and technologies to gather, manage, and analyse data; transforming it into usable information to develop the insights and understanding needed to make informed decisions [8]. On the other hand, although BI systems are widely used in industry, research about them is limited [9].

BI provides accurate and easy data access that can help educational organisation to gather data. Consequently, results from the data analyses could be used to support future planning for the organization. In practice, a BI technology is usually a single version of historical, current, and predictive views of data that are normally utilized in business area to support reporting, analyses and decision making. A BI platform covers all the steps of BI processes (data modelling, data loading, analytical presentation and statistical elaboration, navigation across information, delivery, security) and all the analytical requirements (Query and reporting, OLAP analysis, Dashboard, data mining). BI tools provide aggregation, analyses and reporting functions on the organizations data. BI facilitates the achievement of organizations' missions and objectives by providing required information or the intelligence for the decision makers with regard to the evaluation and control of predefined metrics. In the context of education, academicians can monitor the performances of their students and can predict their interests based on the data that have been gathered.

There are still limited studies for the application of BI in academic environment. Piadade and Santos introduced a concept of Student Relationship Management (SRM), an adaptation from Customer Relationship Management (CRM) concept. SRM was used to promote students success and to closely monitor their academic activities [10]. The CRM systems support the translation of customer information into customer knowledge. Consequently, this knowledge could be used to provide adequate strategies and actions to manage, develop and maintain stronger and profitable relationships between the enterprise and its customers [11]. The SRM system suggested by Piadade and Santos is based on this principle and mainly supports activities related with the students

Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and those particularly associated with the teaching and learning process. BI technology has been proposed as the main pillar technology to support SRM and Figure 1 is shows the framework of SRM fundamental concept.

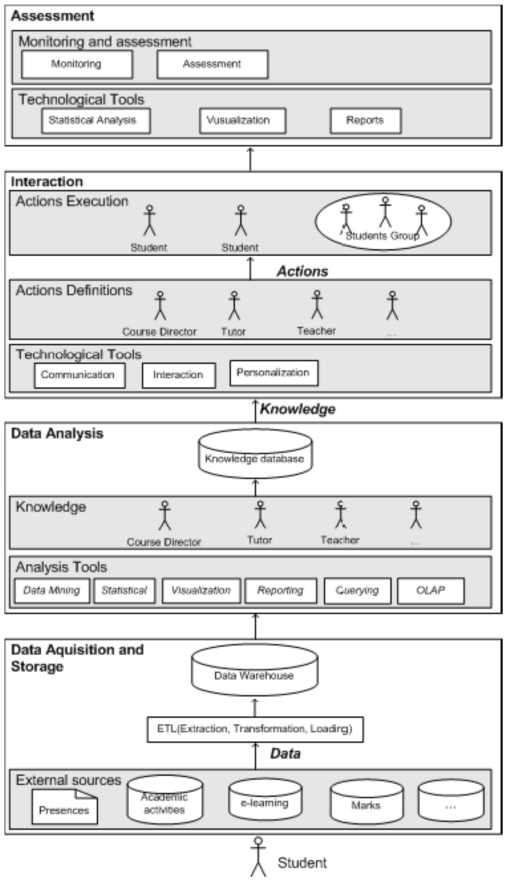

Fig. 1 The Srm System

In our past research, we have proposed an Educational Intelligence (EI) Concept to effectively manage data in academic environment [12]. EI model combine technologies, tools, and technique, people (students, academia, and other stakeholders) to achieve constructive knowledge that helps improve teaching and learning process. Figure 2 illustrates the general idea of EI fundamental framework where Educational Data Warehouse (EDW) is the main component to support EI concept.

In 2012, the EI concept has been further expanded by Chen in his master research thesis entitled 'Applying Business Intelligence in Higher Education Sector: Conceptual Models and Users Acceptance'[13]. Chen has suggested several modifications to EI concept to enhance its relevance of use in Higher Education Institution (HEIs). He also investigated factors that determine potential users' acceptance of an EI system.

Extraction is the first step in the process of gathering the data into the DW environment. Extracting means reading and understanding the source data and copying the data needed for the DW into the staging area for further manipulation. Once the data has been extracted into the staging area, there are numerous potential transformations, such as cleansing of the data (correcting misspellings, resolving domain conflicts, dealing with missing elements, or parsing into standard formats), combining data from multiple sources, reduplicating data, and assigning warehouse keys [14].

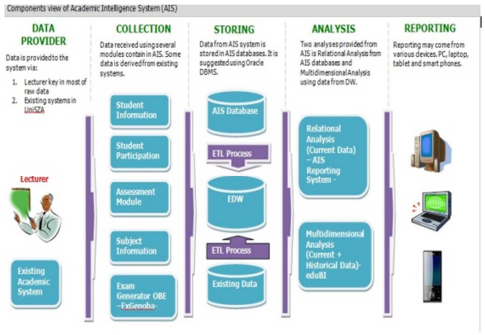

Fig. 2 Educational Intelligence (Ei) Framework

A DW obtains its data from multiple sources in various formats and applies a set of transformations for cleansing and modifying the data in order to be stored into the DW. This chain of activities is known as the ETL process which is normally applied using ETL tools. ETL is the foundation of any data warehouse, data mining and business intelligence [15].

Despite their costs, the deployment of ETL tools is imperative in the implementation of a DW since it helps reducing the overall cost of a DW project. Among the advantages of ETL tools are the Metadata Repository (that can synchronize metadata from source systems, target databases and other BI tools) and a scheduler that makes the ETL process manageable. Examples of some commercial ETL tools are SAS ETL, Informatica and IBM Datastage for proprietary software whereas OSS ETL, Talend Open Studio (TOS), Pentaho Data Integration and b Clover ETL are among the famous open sources solutions.

The process of selecting the right tools and technologies is important in DW development based

Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

on total budget and the critical level business requirements. Now, users have the choice either to use proprietary software or open source solution. The following table shows a comparison on proprietary and open sources BI software in terms of seven aspects that are : support for the end-users, Total Cost Ownership (TCO), innovation, interoperability, compatibility and integration, security and administration; and documentation for the application (software) based on MAIA Intelligence white paper on 2010.

Table 1. Perspectives of comparison on proprietary and open source BI software [7]

Perspective	Open Source	Proprietary BI
End user support	Poor support, consultancy and training due to improper documentation of developed modules from various sources outside the parent company	Availability of timely helpdesk & technical support that is needed for big and stable enterprises
Total Cost Ownership (TCO)	Higher cost of installation and customization leading to higher lifestyle cost	Low TCO primarilybecause of ease of use and hence low training costs and reduced implementation time
Innovation	Work on old- fashioned code written by different developers around the world	Well-built architecture provides all sort of business feature for various industries
Interoperability	Adopted within most of the smaller companies, there are obstacles slowing their adoption, particularly around interoperability	Committed to interoperability in multi-vendor environments, hardware, software, network, data warehouse, ETL, analytics
Compatibility & Integration	Not free and easy to integrate with other application	Excellent integrations with different technologies
Security & Administration	Lack of security support, administration and metadata layer	1KEY BI supports all sorts of user security using LDAP or Active Directory (ADS) and easy administration with 1KEY Console
Application Documentation	Unavailability of documentation – functional or technical	Provide better documentation. 1KEY Agile BI Suite comes with the help manuals for developers' users of desktop and web

Based on the table, although proprietary software seems to be promising in many aspects; the factor of cost is usually a hindrance to its adoption among organizations with lack of financial support. However, realizing the potential of proprietary software and trying to adopt the availability of open source software, we proposed a hybrid approach that integrates both types of technologies in a framework.

EDW Framework

EDW is a framework to develop DW in educational domain to analyze students' performances. It is the back end of an EI framework. The front end of an EI framework is known as Educational Business Intelligence

Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(EduBI). EduBI utilizes BI technologies to integrate various sources of academic data into a single repository (Educational Data Warehouse – EDW).

Taking Universiti Sultan Zainal Abidin (UniSZA) as our case study, the patterns on students' academic behaviors and performance were analyzed using EduBI. A set of data from students' examination results in relational database was extracted into multi-dimensional model (EDW) to support OLAP query processing. Next, an analysis on students' academic performance using BI technologies such as Analytical Reporting, Dashboard, and OLAP was performed. EduBI plays a significant role by optimizing the time to perform current and historical data analysis on students' results hence serves as a platform to identify the strengths and weaknesses of the students. Based on the findings, appropriate recommendations can be provided to the students as an aid to enhance the students' academic performance. Figure 3 presents the framework of EDW and EduBI that utilizes on two layer model.

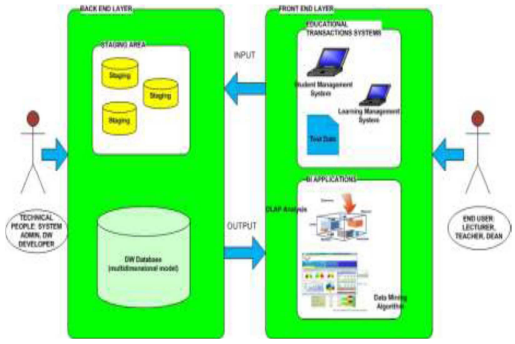

Fig. 3 Back-End (EDW) And Front-End (Edubi) Framework

Ramos & Olievera introduced data warehousing competencies in the undergraduate degree of Informatics Engineering at Polytechnic Institute of Leiria (IPL) [16]. The paper focused on some aspects related to the adaptation, difficulties and challenges during the implementation process and the solutions adopted towards and effective acquisition of competencies by students. Figure 4 shows the architecture of DW suggested based on three layer model.

3. EDUCATIONAL DATA WAREHOUSE (EDW)

Most of the BI literature came from within the business world, the IT industry, and vendors while academic research especially within the field of Information Retrieval is still at its infancy.

We have proposed an EDW in our previous research paper [6] that described the conceptual framework to form DW applications in educational environment. This paper explains the implementation of our proposed framework.

The EDW allows timely access to critical information based on the monitoring metrics in order to measure students' achievements in certain programs and curriculum. A set of data regarding students' examination results from various sources could be extracted, transferred and loaded into multi-dimensional model to support OLAP query processing. Next, an analysis on students' academic performance using BI technologies such as Analytical Reporting, Dashboard, and OLAP could be conducted. The application parts/segments are referred to as Educational Business Intelligence (EduBI) where it serves the purpose to analyse and interprets students' academic data.

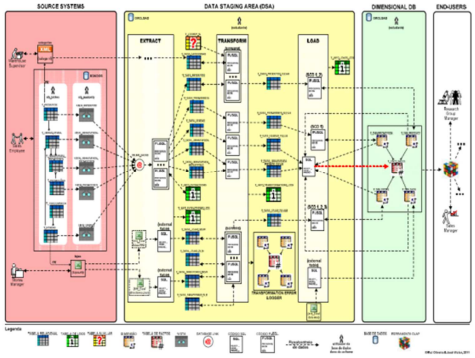

Fig. 4 Ipl Dw Architecture

Designing a multidimensional model for an EDW is not an easy task since it contains fact and dimension tables. Fact tables contain numeric values that represent a measurement of business rules. A row in a fact table corresponds to a measurement at the same grain. As an example, the cost per unit is a sample of fact. From this cost per unit, we can derive sum per unit, average per unit, percentage and so on. These are called derived facts. Dimension tables are attributes (descriptions) of fact tables such as Geographical Dimension, Time Dimension and Product Dimension. A combination of dimensions and fact tables can be grouped into several specific subject areas. The physical view of the subject areas is known as data mart.

For our proposed EDW design, students' results become one fact table with several dimensions

Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

attached to it. Figure 5 illustrates the EDW multidimensional model.

Typical DW serves the purpose to summarize and aggregate the various data. On the other hand, the core issue in designing multidimensional model of educational data is it requires the details of students' data. In order to predict and analyse students' performance, every detailed data needs to be recorded. As an example, one particular student might have a large number of records including marks of assignments; attendance, final exam results and other related information. The eduBI proposed in our research aims to adopt and adapt BI technologies in educational settings.

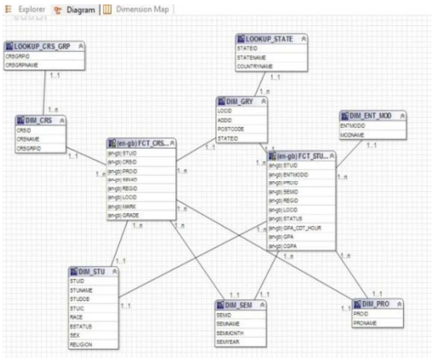

Fig. 5 Edw Multidimensional Design

4. IMPLEMENTING HYBRID TECHNOLOGIES TO EDW FRAMEWORK

As our main objective is to implement BI tools to perform various analyses on students' performances, two factors that we considered were:

a. The capability of every component in the framework.

b. The cost of having those components.

Therefore, we integrated Open Sources Software (OSS) with proprietary software to be implemented in EduBI architecture. Figure 6 shows the framework of EduBI.

5. EDUBI KEY COMPONENTS

Data sources for EDW are usually in heterogonous formats. However, most of the data are Oracle and MySQL data while a there are also text file data in Microsoft Excel format. The hybrid technologies used in this framework is illustrated as in Figure 7.

The key components of eduBI framework are:

a. ETL – Talend Open Studio

Talend Open Studio (TOS) for Data Integration operates as a code generator, allowing and transformation of data scripts and underlying programs to be generated in Java. Its GUI is made of a metadata repository and a graphical designer. The metadata repository contains the definitions and configuration for each job but not the actual data being transformed or moved. The information in the metadata repository is used by all of the components of TOS for Data Integration. The product contains several key components and is based on Eclipse Rich Client Platform (RCP) where TOS is an open sources solution for ETL.

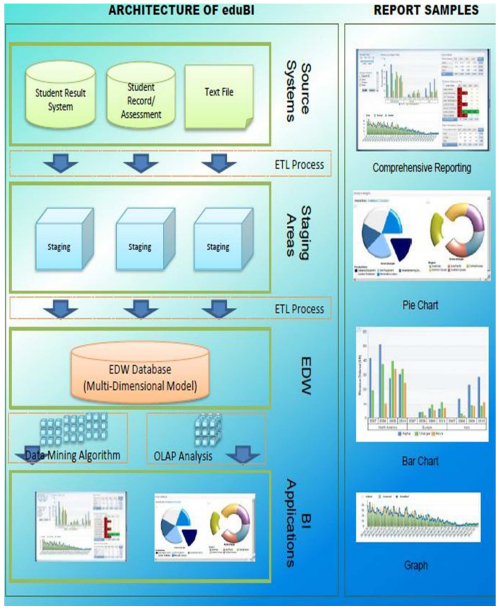

Fig. 6 Edubi Framework

Fig. 7 The Technologies Used In Edubi

Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 8 shows the extraction job to populate the data from academic data source (in MYSQL DBMS) to EDW (Oracle DBMS) using TOS job mapping.

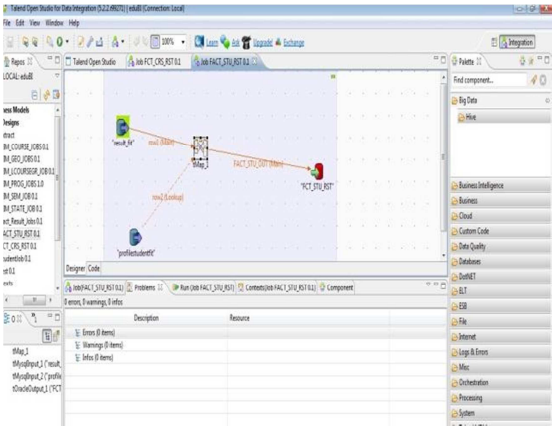

Fig. 8 Job Mapping Process

b. BI - IBM Cognos

Cognos is IBM's Business Intelligence (BI) and performance management software suite. The software is designed to enable business users without technical knowledge to extract corporate data, analyse it and assemble necessary reports.

Cognos is composed of nearly three dozen software products. Since Cognos is built on open standards, the software products can be used with relational and multidimensional data sources from multiple vendors, including Microsoft, NCR Teradata, SAP and Oracle.

The Cognos 8 BI suite which was first introduced in 2005 is IBM's leading performance management software. Cognos 8 BI's capabilities include reporting, analysis, dashboarding and scorecards, delivered through a web-based service-oriented architecture (SOA).

IBM's Cognos products are used by over 23,000 companies around the world, and are consistently ranked in the "leaders" category of Gartner Magic Quadrants for Business Intelligence. Cognos has several components that are used for specific purposes such as Cognos Map Manager, Cognos Framework Manager and Dynamic Quary Analyzer.

Figure 9 shows the Cognos Framework Manager that maps EDW to Cognos applications for the analyses of students' results. The Framework Manager acts as a metadata tool to map the physical databases to the BI front end where it maps EDW databases to Cognos architecture. Each component of Cognos will refer to the Framework Manager to view data.

Once data mapping was accomplished using Cognos Framework Manager, there are several features that can be used to perform desired analyses and to produce the corresponding reports. Figure 10 demonstrates an example of a report for students' results that was obtained using Cognos Query Studio (a web based system). The system contains various useful analysis tools to produce graphical reports and to create a prompt for filters.

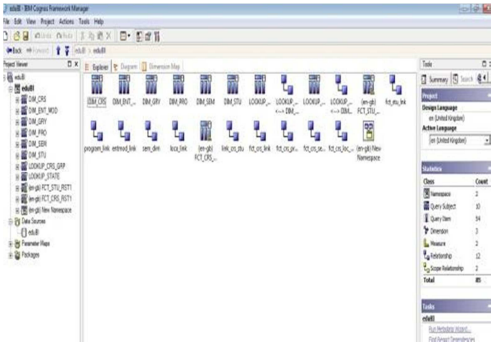

Fig. 9 Cognos Framework Manager

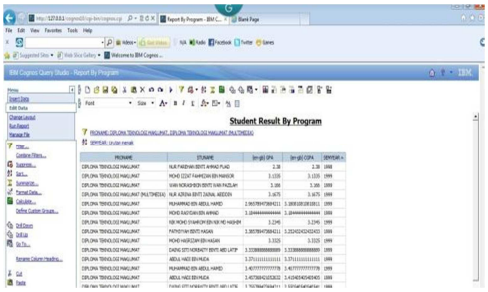

Fig. 10 An Example Of A Report

c. EDW DBMS - Oracle

Oracle is the world's leading supplier of software for information management and best known for its sophisticated relational database products (notably Oracle9i), which are used in Fortune 1000 corporations and by many large Web sites. Oracle's relational database was the world's first to support the Structured Query Language (SQL), now an industry standard.

Multidimensional Model requires high level of DBMS operations that are mostly not available on OSS solution. Reason being, the queries and analyses made are more complicated than relational query hence the justification of the choice of Oracle as our EDW DBMS.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Cognos metadata uses Oracle DBMS for storage. Normally, Cognos prefer DB2 as their main DBMS application since it is a product from the same vendor that is IBM. However, Cognos metadata is also capable of being installed in others recognized DBMS although the steps to install and configure the metadata are usually different for each DBMS.

d. Web Server – Apache

The Apache HTTP Server is a web server application notable for playing a key role in the initial growth of the World Wide Web. It is an open source software with its initial development in early 1995. Apache is developed and maintained by an open community of developers under the auspices of the Apache Software Foundation. Most commonly used on a Unix-like system, the software is available for a wide variety of operating systems, including UNIX, FreeBSD, Linux, Solaris, Novell NetWare, OS X, Microsoft Windows, OS/2 and TPF. In 2009, it became the first web server software to serve more than 100 million websites [9]. To date, Apache still lead the market share to 52.19% from overall market.

EduBI uses Apache as a web server to run the Cognos applications. A set of configuration techniques need to be applied to enable Apache to manage all Cognos Features. Cognos has provided supports via FAQ in their official web to connect every leading web server available in market.

6. EXPECTED RESULTS

Results of EDW framework have been successfully obtained using Cognos BI Reporting. Examples of reports are students' result by subject, geographic information, time dimension and personalized students' achievement. There a still lot more of comprehensive analyses of students' data that can be done using EduBI (front-end) based on EDW multidimensional model (back-end) that has been developed. Individual students' performance can be obtained by analysing detail performances of particular students for all semesters. Group analysis such as detail achievement of students by subject's group could also be accomplished using this framework while various comprehensive analyses could be conducted using BI features such as dashboard, slice and dice, OLAP analysis.

7. CONCLUSIONS

This paper has provided detailed technical architecture of EDW and EduBI which is the main pillar to support the EI Concept. The architecture has been accomplished by integrating proprietary and open source BI software. The need for a proper method to manage academic data that grow rapidly in time has necessitated the proposed framework. Once the data is structurally organized, many data mining algorithm can be applied effectively to obtain the desired results that could further be used to improve students' performances. Based on our research, we concur that the proposed framework is suitable to be adapted in educational settings to help achieve the desired outcomes of any educational organizations.

REFERENCES

[1] Aziz, A., et al., Integration of Heterogeneous Databases in Academic Environment Using Open Source ETL Tools. in The International Conference on Informatics & Applications (ICIA,) 2012. Terengganu, Malaysia.

[2] Romero, C. and Ventura, S., Educational Data Mining: A Survey from 1995 to 2005, ExpertSystems with Applications, 33(1), 135-146 , 2007.

[3] Zorrilla, M., Garcia, D., and Alvarez, E., A Decision Support System to Improve e-Learning Environments, In Proceedings of the 2010 EDBT/ICDT Workshops, DOI:10.1145/1754239.1754252 ,2010.

[4] Merceron, A. and Yacef, K., Interestingness Measures for Association Rules in Educational Data, 1st International Conference on Educational Data Mining (EDM08), Montreal, Canada, 2008.

[5] Sheard, J., Ceddia, J., Hurst, J., & Tuovinen, J. , Inferring student learning behaviour from website interactions: A usage analysis, Journal of Education and Information Technologies, 8(3), 245–266, 2003.

[6] Kimball, R., The Data WarehouseToolkit. 2006: Wiley and Sons.

[7] Nebot, V. and R. Berlanga. Building Data Warehouses with Semantic Data. in Proceedings of the 2010 EDBT/ICDT Workshops. 2010: ACM.

[8] Turban, E., et al., Business Intelligence : A Managerial Approach. 2008: Pearson Prentice Hall Upper Saddle River.

[9] Negash, S., Business Intelligence. Communications of the Association for Information Systems, 2004. 13(1): p.177-195.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[10] Piadade, M., B. and Santos, m., Y., Business intelligence in higher education, Information Systems and Technologies (CISTI), 2010.

[11] A. Payne. Handbook of CRM. Achieving Excellence in Customer Management. Elsevier-BH, 2006.

[12] Aziz, A., et al., Intelligent System for Personalizing Students' Academic Behaviors - A Conceptual Framework, International Journal on New Computer Architectures and Their Applications (IJNCAA) 2(1): 138-153, 2012.

[13] Chen, M., Applying Business Intelligence In Higher Education Sector: Conceptual Model and Users Acceptance, Master Research Thesis, University of Bedfordshire, 2012.

[14] El Akkaoui, Z. and E. ZimÃ¡nyi. Defining ETL Worfklows using BPMN and BPEL. in Proceedings of the ACM 12th International Workshop on Data Warehousing and OLAP. 2009: ACM.

[15] Hanlin, Q., Xianzhen, J. and Xianrong, Z., Research on Extract, Transform and Load(ETL) in Land and Resources Star Schema Data Warehouse , Fifth International Symposium on Computational Intelligence and Design, 2012.

[16] Ramos, J., V., and Oliveira, R., Data Warehousing in the Context of a Bologna, Undergraduate Degree, IEEE Global Engineering Education Conference (EDUCON), 2013.