IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 20, Issue 1, Ver. II (Jan.- Feb. 2018), PP 09-14
www.iosrjournals.org

Developing Data Warehouse for Student Information System (IIUM as a Case Study)

Hewa Majeed Zangana
Department of Computer Science / College of Computer Science and IT / Nawroz University / Kurdistan Region of Iraq

Abstract: *Intensive research and discussion activities on the data warehouse have been found to take place since the past three decades. In large enterprises, huge volumes of data are generated and consumed, and substantial fractions of the data change rapidly. IIUM as an International University need up-to-date information to make decisions and acquaint quality reports. In IIUM there are three student data servers support ITD in order to obtain a report on students by using a way that takes a long time and efforts for preparation. This paper proposes the integration between the three student database servers by comprehensive data warehouse contains student Information to Up-to-date and quality data to ensured the correct decision making and reports producing..*

Keywords: *Data Warehouse, Integration.*

Date of Submission: 25-01-2018 Date of acceptance: 09-02-2018

I. Introduction

The concept of data warehousing dates back to the late 1980s, when IBM researchers Barry Devlin and Paul Murphy developed the "business data warehouse". In essence, the data warehousing concept was intended to provide an architectural model for the flow of data from operational systems to decision support environments. [1]

The classical data warehouse concept is defined by the father of the data warehouse (Inmon, 1993) as: "A Data Warehouse is a subject-oriented, integrated, time-varient, non-volatile collection of data in support of management's decision making process." [2]

To clear this definition, he defined the terms in the sentence as follows:

- *Subject Oriented:* data that gives information about a particular subject instead of about a company's ongoing operations, such as sales.
- *Integrated:* data that is gathered into the data warehouse from a variety of sources and merged into a coherent whole format.
- *Time-variant:* all data in the data warehouse is identified with a particular time period (More historical data than transaction systems).
- *Non-volatile:* data is stable in a data warehouse. More data is added but data is never removed (Does not change).

A researcher provided a much simpler definition of a data warehouse; data warehouse is "a copy of transaction data specifically structured for query and analysis" [4]. This definition provides less insight and depth than Inmon's, but is no less accurate.

Another definition from another researcher: "A data warehouse is a repository (collection of resources that can be accessed to retrieve information) of an organization's electronically stored data, designed to facilitate reporting and analysis[5]. More simply, a data warehouse is a collection of a large amount of data.

This definition of the data warehouse focuses on data storage. The main source of the data is cleaned, transformed and cataloged and is made available for use by managers and other business professionals for data mining, online analytical processing, market research and decision support [6]. However, the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary are also considered essential components of a data warehousing system. Thus, an expanded definition for data warehousing includes business intelligence tools, tools to extract, transform and load data into the repository, and tools to manage and retrieve metadata.

II. The problem statement:

The way of preparing reports on students in the IIUM does not appropriate with the rapid development of the university; especially with the scientific race between universities. When ITD department wants to make reports on students, it creates a temporary database to collect information from resources (the three servers) and make the reports to the ministry of high education (by following the instruction of the ministry using their format and approach). In this way it faces the following problem: data is not updated daily especially the old records, need a lot of time and resource (admin staff) to update data. The idea of solving these issues is to have comprehensive data warehouse contains student Information

III. Background of the problem

Our case study will be on International Islamic University Malaysia (IIUM). IIUM founded in 1983, now has four campuses with more than 20,000 students from 90 countries studying management, engineering, technology, medicine and the humanities [7]. The Office of Corporate Strategy & Quality Assurance (CSQA) plays an important role in ensuring that the Strategic Plan of the university has been implemented and the Quality Assurance is carried out in a more systematic and efficient way in line with the university's mission. The office of (CSQA) was established on 1 June 2008, and is responsible to develop strategies so as to promote good practices of quality assurance and quality enhancement in the University[13].

CSQA divided into 3 units; namely Quality Assurance & Administration, Corporate Strategy & Planning, and Corporate Performance Management & Reporting.

To get the clear image of students' information in the university, CSQA have to combine different databases from more than three data servers which are: CPS, ANR, Matriculation center.

All of the above has a separated student information system. To combine these data, a temporary data warehouse is made by Information Technology Division (ITD), sent to CSQA, then analyzed and deployed in the office work.

This operation takes a lot time to extract data and standardize it in one form as much as possible. The new database thus contains periodical data, which means the data will not be up-to-date information about the student. Hence, from here the issue of quality of the data arises.

The main objective of the project is to provide the university a system to integrate all student information from all data servers and combine it in on data warehouse that would be available in real time whenever it requires obtaining data and producing reports.

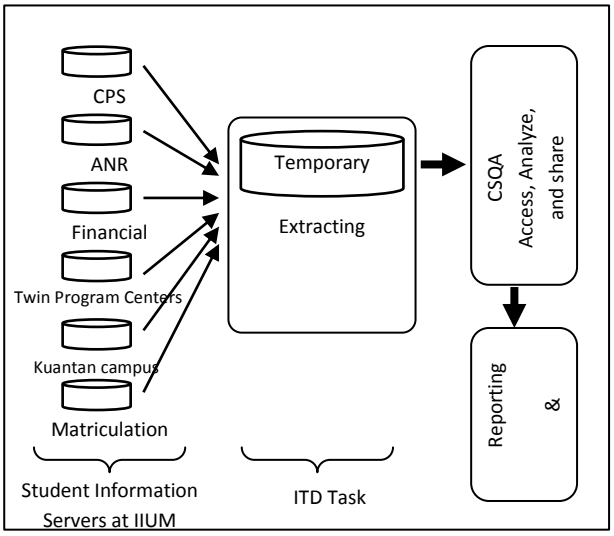


Figure -1- *How CSQA obtain student information*

IV. Statement of purpose

The Business Driver and initiative is geared towards the provision of excellent Services with a strategy to automate the business processes in order to improve efficiency. This will be in a bid to achieve the following goals:

4.1 Business Initiatives and Drivers:

- The need to have a comprehensive data warehouse contains student and staff information.
- Up-to-date and quality of the data must be ensured for the correct decision making and reports.

- To achieve World University Ranking and Rating criteria such as Times Higher Education Supplement (THES-QS) World University Ranking, Malaysian Public University Rating (SETARA).

4.2 Functional scope:

The scope will be to provide opportunities to the ITD and CSQA to achieve its planning success through: the providing accurate information in real time reporting. This is in a bid to achieve the following goals:

- Up to date information of student's information.
- Accurate decision making.
- Obtaining quality data for reporting.

4.3. Business Processes:

- A common interface to link among divisions and departments' databases. (naviCAT)
- A suitable report programming, to produce the required reports.

V. Current assessment of enterprise architecture

The IIUM is one of the universities that are interested of enhancing their fulfilling which is applicable with its vision and mission, on making the periodic reports to evaluate the performance of students and keep up with other universities.

The University has a specialized department to collect reports and evaluate them called The Office of Corporate Strategy & Quality Assurance (CSQA) that evaluates the fulfilling of the students and send them to the ministry of high education.

CSQA gain reports from the ITD, process of making reports is still following the traditional ways that does not fit with the size and demands of the university.

Information is collected from three servers and put in a temporary database, extract reports and send them to CSQA,

and this way have challenge which that: data is not updated especially the old records, need a lot of time and resource (admin staff) to update data.

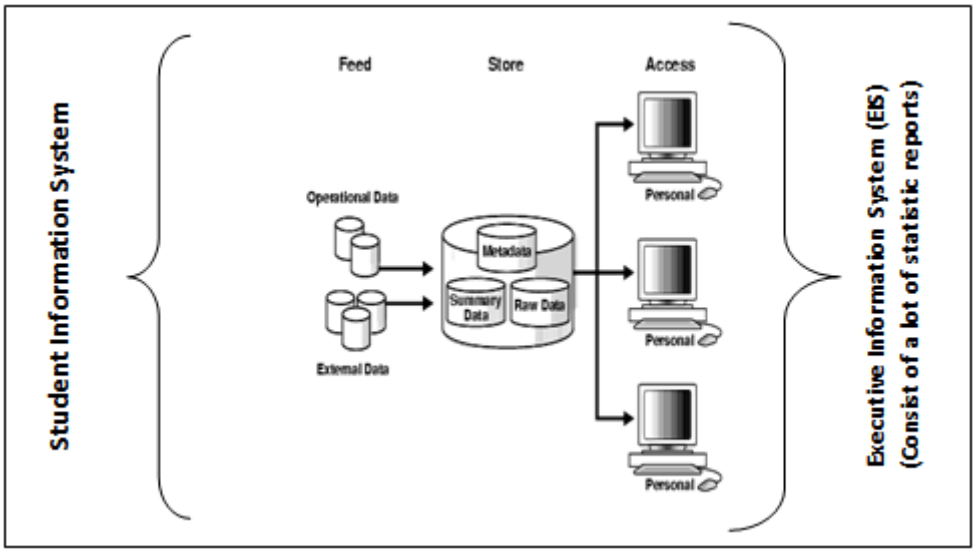


Figure -2*- Architecture of data repository of EIS*

A sample of the some information that are extracted from Operational Data:

- Student Name, student ID number, race, religion, gender, program code, nationality, qualification, address. etc.

VI. Proposed Enterprise Architecture:

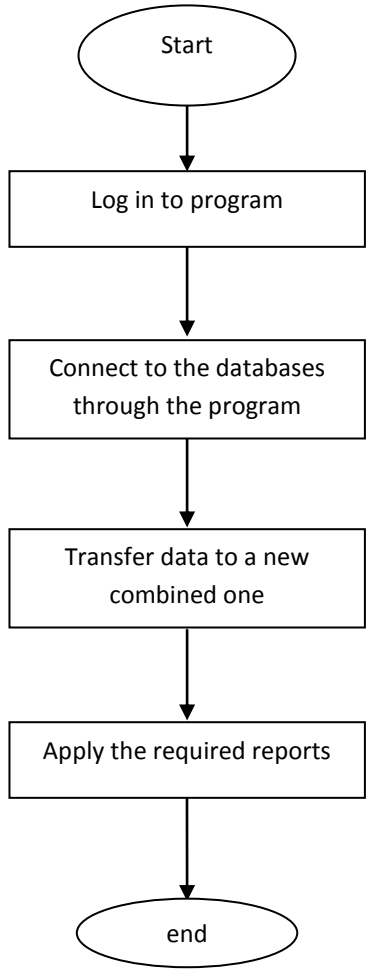
To integrate all servers via linking together the database used in the departments or units involved in the exercise we look up to the need of using the comprehensive data warehouse contains student Information. as a way of letting the systems talk to each other. The proposed flow is as depicted:

When CSQA needs a report

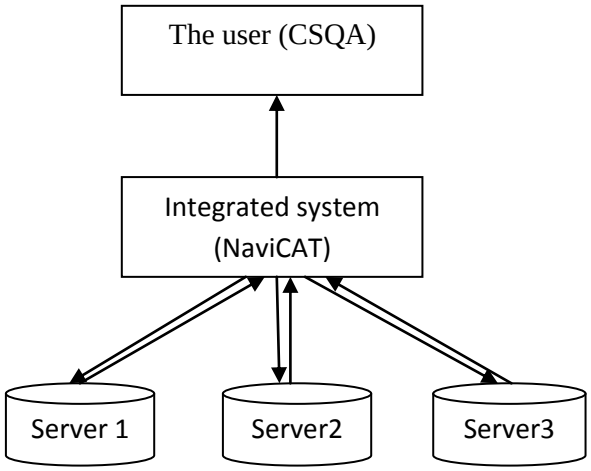
1. First, the user (from CSQA) logs in to the program (NaviCAT).
2. After staff login he/she opens the three servers by program.

3. By connecting to servers a single comprehensive view of student information will be available to the user in the main screen of the program.
4. The user transfer the tables of students' information from all three data bases to the local one using (Data Transfer) and (Data synchronization) tools from the program.
5. The user chooses the appropriate reports that are previously arranged for the requirements of the request.
6. Print reports and end the program.

The current diagram shows the description of these steps:



VII. The Proposed Platform for Development:



VIII. Management and Technical Constraints and Risks for Implementation

1- Project support:

The prime management is responsible on supporting program and developing and the allocation of adequate budget to buy programs and train the employers so must get official approval before starting the project

2- Ensure success of project:

This project should be carried out by the competent authorities with expertise in this area and that the performance of this program must be accepted this Plan subject to evaluation on an ongoing basis with the maintenance

3- Hardware requirement

In order to ensure the success of project, it is needed of hardware requirements such as preparing an additional PC to be the storage for database, this will be the tools that contain the application that links between servers.

4- Software requirements

This project just like any other IT project, require to install the programs and continuous maintenance so that must allocate the trained staff on this system and to gain the support of vendor. The program is a lessened program and must be purchased.

5- Technical skills requirements:

Although the project is easy to implement, but it requires a specialized staff have experience in databases and Oracle in order to overcome any problems that can be happens.

Also it is needed to provide training and workshops to the staff to improve implementation this project.

IX. Important Facts

The following are important facts that we should know about the University Data statues:

The refresh cycle varies for each data collection in the Warehouse. Do not execute queries against a collection in the Data Warehouse when it is being refreshed. Your queries will not be completed because the refresh process involves dropping index pointers from tables.

As previously mentioned, there is a "data delay" between the Warehouse and the transaction system. Keep this time frame in mind when comparing data between the two systems or when analyzing query results--it may help you to understand and account for any data discrepancies found. Always indicate or communicate the date on which your Warehouse results were obtained and identify the Warehouse, and not the transaction system, as the source of the information. For example, include the date of the query in the report header and the source of the information, the University Data Warehouse, in the report footer.

Some University data are more dynamic--change more frequently--than other data or are subject to change at specific times during the fiscal year. Take this into consideration when writing queries against the Warehouse. Examples of data that change are course enrollments during the add/drop period and student grades during the grading period.

Data in the Warehouse is for queries only. That is, you cannot add, change, or delete data in the Warehouse. These activities are done only in the transaction system.

X. Findings and Conclusion

Improving and optimizing the ability to obtain quality and up-to-date student information is one of the necessarily needs in Enterprises nowadays. The present applications in IIUM are heterogeneous data, and it is difficult to present reports that cross-reference from several databases. Many companies have application to solve situations like what we have here, some are too expensive, some are suitable, and others are free. We found that NaviCAT Premium will be a good solution in our case study. It is simple, powerful and can connect to most kind DBMSs. NaviCAT can access remote servers through internet (HTTP) and locally through the exits local network. It can combine all databases to one complete data warehouse, save the data into whatever type of database system you need. Most of the applications in IIUM are Oracle based programming. It can synchronize very easily between servers.

The advantages of deploying the idea is it will avoid the university to pay huge costs on merging databases. Also time spend when requesting data will not be a gap because it will be real-time request, and these data will be the up to date information which will be more trusted data.

References

[1]. Professional Development for the Massachusetts Education Data Warehouse, September 2008, Introduction to the Data Warehouse, Version 2.0.
[2]. SUN, ZHANG, PAN Qidong, and YAN Shaoling, August 02,2010Research and Design of Data Warehouse in Coal Mine Enterprise
[3]. David J. DeWitt, How to Build a High-Performance Data Warehouse
[4]. Weisensee D. Matthews E., McInnis A. (2004) Implementing Data Warehousing and Business Intelligence at McMaster University Using the SAS Intelligence Value Chain 20 Oct 2004 pp125-30.
[5]. Fabio Casati, Malu Castellanos, Norman Salazar, and Umeshwar Dayal, 2007 Abstract Process Data Warehousing.
[6]. Jane Qiong Zhao, 2007, Formal Design of Data Warehouse and OLAP Systems, Massey University, New Zealand.
[7]. About IIUM, available online at IIUM website: (http://www.iium.edu.my/about/about.shtml) accessed date: 29 Sep 2010.
[8]. R. Kimball, M. Ross, The Data Warehouse Toolkit, second edition, John Wiley & Sons, 2002.
[9]. L. Cabibbo and R.Torlone, "A Logical Approach to Multidimensional Databases". In Proceedings of the International Conference on Extending Data Base Technology (EDBT '98, Valencia, Spain, Mar.), pp. 183–197, 1998.
[10]. M. Golfarelli, D. Maio and S. Rizzi, "Conceptual Design of Data Warehouses from E/R schemes", In Proceedings of the 31st Hawaii International Conference on System Sciences (HICSS '98, Kona, Hawai), 1998.
[11]. Cheryl Bach-Scott, Htay Hla, Michele Singleton, Elizabeth Taylor, Paul Teitelbaum, Client/Server Based Data Warehouse at the University of Arizona. The University of Arizona, 2002.
[12]. Ganner Auth and Eitel von Maur, 2002. A software Architecture for XML-based data interchange in DBWS, multimedia engineering - EDBT 2002, Issue 2490, pp 1-14, 2002.
[13]. About CSQA Corporate Strategy & Quality Assurance. Available online at: http://www.iium.e du .my/qau/index.php ? page=about-csqa) Accessed on 27 Sept 2010.

Hewa Majeed Zangana "Developing Data Warehouse for Student Information System (IIUM as a Case Study)." IOSR Journal of Computer Engineering (IOSR-JCE) 20.1 (2018): PP 09-14.