Adoption of Data Warehouse in University Management: Wasit University Case Study

A F Neamah
Computer science and IT faculty, Wasit University, Iraq
ali_alyasree@yahoo.com

Abstract. Presently, the utilization of traditional data instrument and technologies are incapable of managing the load and analytic process of data that can be translated into significant information for top management. It is observed prevalently that Information Technology is becoming an important source for the dissemination of knowledge by the Education in order to retain competitiveness in education systems and for adaptation purposes alongside the dynamic setting of the business. The study of the growth of higher education indicates that the Iraqi higher education has entered a rapid phase of progress. Thus, the optimization of universities is imperative under the circumstances. The main goal of the implementation of a data warehouse is in the conversion of the abundance of data into information that can be used in the improvement of admission, examination, examination results and other processes. In this paper, we highlighted the need of data warehousing for higher education and provided an approach to build the warehouse suitable to Waist University and may be to similar universities in Iraq.

1. Introduction

Describes data warehouse as the cornerstone of data architecture, which serves as the fundamental core of all decision support (DSS) systems development. It is noted that the work of a DSS analyst in the environmental data warehouse is much simpler than in the conventional heritage setting, as there is a single centralized data source (data warehouse), with the ready availability of the granular data sourced from the data warehouse.

A data warehouse entails a storage repository of data which may have been manually or automatically collected from a source or from multiple sources through an integration layer that transforms data to meet warehouse requirements [2]. It may be conceptualized as a large volume of data organized under a traditional, one-stop information center framework. As data is structured and is in a homogeneous format, queries in a centralized data warehouse are easier to write and provide better output to the user than accessing multiple distributed data warehouses [3]. Fig.1 provides an overview of data centers design.
Data warehouses have evolved rapidly in recent years. According to one of the researchers, it is described as such, "a data warehouse is organized according to subject, distributed, is varied across time, and consists of a stable data set primarily utilized in organizational decision-making." Meanwhile, [4] defines data warehouse as "a large, silent storage that extracts data from the separate transactional system, and then stores it on a regular basis. Data warehouse systems will play a major role in academic and industrial communities in coming years. Another interesting feature shared by data warehouses is that as information is applied to existing data, material tends to grow.

Most transactional process schemes have a predefined expiration period during which active data is maintained; the aged data will be stored until the time limit is reached. The expansive data warehouse aspect results in vast and ever-growing quantities of data [5]. In addition, unlike conventional data repositories that concentrate on a single feature or application, data warehouses collect information databases from across the enterprise, allowing an integrated knowledge collection in all areas of the enterprise.

There is an increasing awareness of the role and benefits of data warehousing for public administration. The relationship between data collection and reporting have been improved by data management technologies [6]. Additionally, a data warehouse offers a categoric construction for the variable and static data. Any increase in static variable data may result in an automatic conversion to control the data of the static data. In addition, data processing may require auto-adjustment of the corresponding database. It is imperative for the architecture of the database to possess a multidimensional processing pattern for the department needs as well as for the end-users, which are the people. In light of the universities in Iraq specifically, there are a variety of universities operating independently within the region. The processing of the data may be programmed to be automatic, or established according to specific time by means of an e-governance system. A data warehouse will get all the details required for authorizing the information. Equal distribution of the available information is made accessible through a data warehouse.
2. Need to Dw in Higher education universities

Today, thousands of reputed Educational Institute registrations register. In relation to these factors, the true test for administration is to meet the various needs of students and face rising academic complications. The cost of constructing a data centre is costly for each educational institution as it needs data storage facilities for data storage and the extraction of data from a data storage facility by using data mining software.

This study offers the option of building data warehouses and extracting useful information through open source data storage and knowledge mining software. Top management in educational institutes also requires timely analytical reports to assess and track student results, increase and decrease in admissions, academic qualifications, laboratory and building maintenance records, industrial student visits and overall student placement records. We will require accurate objective reports to help them make long-term decisions. It was found that the majority of reporting and research took time to gather data from the various systems before the study could be carried out. Top management wants and needs more data, but analysts may provide limited information within the required time frames at a high cost. In order to give information more convincingly on the prediction of patterns and trends and to analyze a problem or a situation more effectively, a data warehouse is necessary for this specific purpose. Within an organization, there are numerous interfaces, specific data and various data representations, duplicate, and many conflicting details. Across several different systems, the same details can be found; Teacher Data can be used in four different systems, such as library, student information, placements records, fees and labs, building maintenance records, etc. Combining information and generating reports or making decisions in such systems is therefore very difficult [15].

While security measures are maintained. There are several reasons why data management in Iraqi Educational Institutes find the following important needs to be taken into consideration:

-- All available data have to be used to make decisions efficiently and accurately.
-- Domain experts are agency consumers, not computers
  Skilled. Skilled.
-- In short time, the volume of data doubles, impacting response time and the ability to comprehend its content.
-- Competition heats up in industry areas Value-added intelligence and knowledge.

Decision-makers at educational institutions therefore need more and more analytical knowledge to capture the whole picture of the world of their organization. It is precisely the function of the warehouse to give them this global view and comprehensive analysis capability.

3. Literature Review

3.1. Review of Iraqi higher education development

The world has become progressively dynamic and aggressive in terms of competition for Higher education entities to operate in. In adapting to the national and global economic, political and social changes, they have to compete with other institutions. In addition, various stakeholders expect institutions of higher education to deliver the best response to these demands in a timely manner.

In order to tackle this problem, higher education needs to take the right decisions by reviewing vast data sources generated to cope with these rapid changes.

Most higher education institutions are investing significant resources in information technology for the implementation of data warehouse systems.

The development of data warehouses is a way of collecting useful information from distributed data through a consolidated integrated storage network in certain information systems, and promoting the need for data history [7]. These integrated data can be used for transmitting information and can be verified from various angles; thus, the degree of detail can be set. Today, Iraq's higher education sector will be an integral component in Iraq’s efforts to advance and enhance its economy and community.

Universities in Iraq possess the probability to assist the country in solving its socio-political problems, in addition to offering the faculty and students the opportunity to be agencies of societal transformation.
Despite the consequences of war, more than 10 universities have established higher education entities in fewer than 10 years. Moreover, in spite of the current struggle with inadequate infrastructure and intense socio-political volatility, the higher education sector has an immense role in helming the sustenance of a long-standing civil peace and aims in building a broad base of knowledgeable Iraqi community. However, higher education is highly concentrated in Iraq. It has therefore centralized decision-making to those concentrated areas and decentralized decision-making by making universities more autonomous [6]. The decisions therefore apply to all universities, each of which has specific circumstances.

3.2. Data warehouse VS traditional database

Data centers and databases possess correlational data structures, that were constructed for various intentions. A database is a shared archive that reflects components of worldly realities, while on the other hand, Data Warehouse entails a system of information system which preserves all history of data, as well as communicative data derived from both solitary and numerous sources. The database is utilized for storage of data and the processing of data is done by the center.

Furthermore, the Online Transaction Processing (OLTP) database was established for existing procedures for business, for the purpose of storing present transactions, as well as allow quick passageway and availability to new transactions. A database is designed for the enhancement of the swiftness and effectiveness of data updates (inserted, changed or deleted) to expedite the analysis for the objective of accessing the data. Online Transactional Processing (OLTP) is used by data bases to delete, introduce, replace and upgrade massive amounts of brief electronic transactions.

Other features comprise speedy query processing, data integrity possessing multiple accesses, in addition to the completion per second of a number of transactions. Furthermore, the provision and retention of a common source of present time and precise data are enabled by databases that process OLTP transactions. Due to the joining of a number of tables, it is however problematic to execute the analytical queries and need a database administrator with adequate experience or a developer who knows the application well to input queries resulting in significant analysis [8].

It is important for the response time for the server to be tremendously fast for effective transactions processing. The ultimate critical database feature entails the write process to be documented in the system. A company which offers products to be purchased online could not afford to be too long in company in the event that its database did not document every transaction made.

Information centers and repositories are often partnership information structures developed for different purposes. The warehouse was constructed for huge past data, and enables the rapid and complicated enquiries of the overall data typically used by Online Analytical Processing Server (OLAP). Online analytical processing (OLAP) is used by data warehouses and is constructed for the performance of a small number of complicated queries on large historical data sets aggregates. For the generation of the summary data, multidimensional perspectives, and faster queries response times, the tables are deormalized and transformed Moreover, queries response times are utilized to evaluate the efficiency of an OLAP program.

OLAP helps managers and analysts to select, compile, display and analyze organizational data, being a business intelligence tool, for the purpose of detecting and gaining in depth knowledge into company norms and ascertain potential problems [9].

3.3. Barriers for Developing an Academic Data Warehouse

Businesses began using data centers in the 1970s. Nonetheless, only recently have academic institutions begun to recognize and explore the possibilities and advantages that data warehouses bring. A series of driving forces have been generated by several environmental factors that motivate academic institutional leaders to explore data-ware housing options. These factors include declines in government financial support, faculty supply, and research funding, faculty salaries, in addition to students’, parents’, and employers’ expectations.

Such factors led university administrators to adjust how they view operations in the higher education sector. When managers seek to deal with these environmental variables, their knowledge requirements must be measured in relation to these conditions and available information. Each factor creates drivers
for the details. Drivers can be produced from a single factor or can result from multiple factors that work together [10]. The drivers support the need to build a data warehouse however they are created.

4. Objective of study
It aims to introduce information management system data warehouse to Wasit University, Iraq, in order to support decision-making. The research contribution is to offer a solution to Wasit University leadership for them to ascertain the university’s situation, and the research findings could be applied in other universities that require a support system related to decision-making.

5. Data warehouse in public Universities
Prevalently, the Information Technology has emerged as an important source for the education system to maintain its competitiveness in the education systems, in addition to adjust to the dynamic nature of business setting.

It is imperative for industries such as the education sector that is emerging as a higher-order service sector to put their trust in technology as a means of keeping abreast alongside the global economy which technology has launched. Scholars have also discovered that colleges have been collecting data for a long time: such as grade monitoring, enrollment, sales of textbooks, test scores, meals at the cafeteria, and the like. Yet nothing has actually been done with these materials, whether because of privacy concerns or technological capabilities to improve the learning of the students. There are obviously a lot of opportunities for improved data collection and analysis in education with the introduction of technology in more colleges, alongside with a drive for more transparent data from the government. However, some states have turned to items such as the standardized test score data to gauge the efficacy of teachers, and in turn, retention and promotion for these teachers.

Universities and colleges have prepared various reports that are required for the study of admission processes, for handling exams and in constructing their tests, and others. These reports are for the purpose of their self-determination and for transmission to the regulatory authorities with the assistance of the Management Information Systems in the generation process [11]. Therefore, the term Information Management System is an ancient idea for education systems. Such reports are mostly generated by computers, and can be produced at any time. The use of the terms Data Warehousing and Data Mining is therefore fairly recent. Universities and Colleges’ need for Data Warehousing and Data Mining is for the development of management support systems. Data warehouse solves issues such as lack of content data that may restrict model-based decision-making support and executive information systems by providing real and decision-driven information to facilitate critical success factor control [12].

A data center incorporates vast volumes of business data from various, separate data sources into a centralized repository to access and analyze operating databases. Information storage has become vitally important in the presence of data mining and data warehouse [13].

According to [14], data warehousing for the Iraqi public universities consist of five stages. For this purpose, there are five levels of data warehouse for universities; Firstly, the databases level perceived to be the key sources of data (data and information from faculty and departments); It encompass the data-source and knowledge for the students and staff from which the resource is derived from the data warehouse. Secondly, the ETL level functions in the conversion, extraction and loading of staff and student data and information from faculties and department into common data warehouse. Thirdly, all of these data and information are stored at the warehouse level. Fourthly, the method level is used by OLAP to interpret and display the information in a multidimensional manner. At the fifth-level entails the interface level which enables the data center to gain the information stored by the management interface. Figure 2 shows the flow of data warehouses for management system.
Figure 2. Data Warehouse Architecture for Management System

6. Infrastructure for data warehouse

The infrastructure requires several elements for and technical component such as server hardware, operating system, network software, database software, LAN and WAN, vendor resources, persons, procedures, and training.

Components of the data center system can be broken down into two categories: operating system and physical infrastructure.

**Operational infrastructure**

Part of the base technology is computer hardware and associated applications. People, Processes, Training and Management Software is organizational infrastructure to support and part of the architectural dimension. Those components are as critical to keep the data warehouse running as hardware and software.

**Physical infrastructure**

Any system has to have an overall infrastructure to live on. The infrastructure is made up of the rudimentary hardware parts, the operating system with its utility tools, the network and the applications. In addition to the overall framework, there is a set of software running on the chosen system for the execution of the numerous architectural component purposes and services.

7. Proposed University Data Warehouse Architecture

The following figure illustrates the key steps to create the prototype data warehouse. There are four stages of this architecture. The first stage is data processing; data collection to collect relevant data from different data sources; data entry to convert hard copy data into soft copy form; data validation to ensure the validity of incoming data and to identify any possible errors arising from data entry processes or anomalies occurring in the original raw data. The second stage is ETL is to load, extract and transform information and data from different resources in university. The third stage is warehouse stage; is storage of all information and data.

The fourth stage, which is the final stage is the creation of end-user devices from which the experts of the organization can access the systems. These tools include charting tools (for building different
types of charts), OLAP tools (for providing users with a variety of views for the same data cube) and remote access tools (for building websites).

Figure 3. Proposed Data Warehouse for Wasit University

In traditional data warehouse technology, ETL is the main process which cannot handle unstructured data. In this method, we need a robust ETL process that can deal with many problems of data quality, such as duplicated data, inconsistency data and garbage. The device proposed is shown in Figure 3.

There is a combination of Hadoop and RDBMS in the system. As a data ingestion / staging tool and as a data management and data presentation platform, Hadoop can improve the RDBMS.

As shown in Figure 3, the proposed system architecture has been presented. Structured data are aggregated in our system, whereas unstructured data is volatile and usually in a computer-recognizable format. Unstructured data are free-text, videos, websites, RSS feeds, meta-data and web server logs, for instance. The unstructured data are processed in our proposed framework by categorizing and filtering and then stored in context. The unclassified data is contained in the raw data. First, the data will be stored in similar data via the search process for relationships or trends. Then the relevant data which are already processed and adaptable to predefined structures are stored in exploration data. Eventually, convergence of explored data and aggregate data can be analyzed using OLAP and business intelligence techniques.

8. Conclusion
Most of the growing universities and colleges are seeking alternatives to exploit their data assets to differentiate them from the rest, and posit them in the optimal position. This paper analyzes the manner in which data warehouse applications map with the varying challenges in university systems.

It also explores how the universities can introduce large enterprise data warehouse to provide information on the overall university systems at the atomic level; this is in comparison to the installment of data marts, that can eventually be incorporated into the overall university and college enterprise data center, the starting-up chances of a data warehouse are globalized. Using integrated data store to discover a huge potential stream loss that can be averted and further inform the manner in which courses can be managed, and syllabus grouping for forthcoming time. The dispersed social interactions that learners participate in on a daily basis through the various social media applications are largely being
ignored in the data collection of universities and colleges should be noted. There are many issues that need to be discussed before there can be a widespread adoption of Data Warehouse.

Conflicts of Interest
The author has participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.
This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.
The author has no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript

References
[1] Jaber M M, Abd Ghani M K, Suryana N, Mohammed M A and Abbas T 2015 Flexible data warehouse parameters: Toward building an integrated architecture International Journal of Computer Theory and Engineering 7 (5) p 349
[2] Alhyasat E B and Al-Dalahmeh M 2013. Data warehouse success and strategic oriented business intelligence: a theoretical framework arXiv preprint arXiv:1307.7328 5 (3) pp 169-184
[3] Hopfgartner E, Schuetz C G and Schrefl M 2017 A Case Study of Success Factors for Data Warehouse Implementation and Adoption in Sales Planning pp 1-10
[4] A. Information 2015 Is the Data Warehouse Dead?
[5] Thorat S B, Butey P K and Dhore M P 2018 An approach to establish data warehouse for higher pp 55-57
[6] Mohammed M A, Hasson A R, Shawkat A R and Al-khafaji N J 2012 E-government architecture uses data warehouse techniques to increase information sharing in Iraqi universities 2012 IEEE Symposium on E-Learning, E-Management and E-Services pp 1-5 IEEE
[7] Santoso LW 2017 Data warehouse with big data technology for higher education Procedia Computer Science 124 pp 93-99
[8] Bassil Y 2012 A data warehouse design for a typical university information system arXiv preprint arXiv:1212.2071 pp 12-17
[9] Salvioni D M, Franzoni S and Cassano R 2017 Sustainability in the higher education system: An opportunity to improve quality and image Sustainability 9 (6) p 914
[10] Wierschem D, Mcmillen J and Mcbroom R 2003 What Academia Can Gain from Building a Data Warehouse Educause Quarterly 26 (1) pp 41-46
[11] Maican C I 2009 Integrated university management system based on open source tools 2009 Fourth International Conference on Internet and Web Applications and Services pp 626-631 IEEE
[12] Fredriksson U, Gajek E and Jedeskog G 2009 Ways to Use ICT in Schools to Optimize the Impact on Teaching and Learning Acta Didactica Napocensia 2 (4) pp 21-32
[13] Hadullo K, Obok, R and Omwenga E 2017 A model for evaluating e-learning systems quality in higher education in developing countries International Journal of Education and Development using ICT 13 (2)
[14] Mohammed M A and Anad M M 2014 September. Data warehouse for human resource by Ministry of Higher Education and Scientific Research 2014 International Conference on Computer, Communications, and Control Technology (I4CT) pp 176-181 IEEE
[15] Shaweta 2014 Critical Need of the Data Warehouse for an Educational Institution and Its Challenges 5 (3) pp 4556-4559
Appendix A Difference between Database and Data Warehouse

| Parameter          | Database                                                                 | Data Warehouse                                                      |
|--------------------|--------------------------------------------------------------------------|---------------------------------------------------------------------|
| Purpose            | Is designed to record                                                    | Is designed to analyze                                              |
| Processing Method  | The database uses the Online Transactional Processing (OLTP)             | Data warehouse uses Online Analytical Processing (OLAP).             |
| Usage              | The database helps to perform fundamental operations for your business   | Data warehouse allows you to analyze your business.                  |
| Tables and Joins   | Tables and joins of a database are complex as they are normalized.        | Table and joins are simple in a data warehouse because they are denormalized. |
| Orientation        | Is an application-oriented collection of data                            | It is a subject-oriented collection of data                         |
| Storage limit      | Generally limited to a single application                                | Stores data from any number of applications                         |
| Availability       | Data is available real-time                                              | Data is refreshed from source systems as and when needed            |
| Usage              | ER modeling techniques are used for designing.                           | Data modeling techniques are used for designing.                    |
| Technique          | Capture data                                                             | Analyze data                                                        |
| Data Type          | Data stored in the Database is up to date                                | Current and Historical Data is stored in Data Warehouse. May not be up to date. |
| Storage of data    | Flat Relational Approach method is used for data storage.                | Data Warehouse uses dimensional and normalized approach for the data structure. Example: Star and snowflake schema. |
| Query Type         | Simple transaction queries are used.                                     | Complex queries are used for analysis purpose.                      |
| Data Summary       | Detailed Data is stored in a database.                                   |                                                                     |
### Appendix B application of DW and tradition database

1- Applications of database

| Sector        | Usage                                                                 |
|---------------|----------------------------------------------------------------------|
| Banking       | Use in the banking sector for customer information, account-related activities, payments, deposits, loans, credit cards, etc. |
| Airlines      | Use for reservations and schedule information.                       |
| Universities  | To store student information, course registrations, colleges, and results. |
| Telecommunication | It helps to store call records, monthly bills, balance maintenance, etc. |
| Finance       | Helps you to store information related stock, sales, and purchases of stocks and bonds. |
| Sales & Production | Use for storing customer, product and sales details.                 |
| Manufacturing | It is used for the data management of the supply chain and for tracking production of items, inventories status. |
| HR Management | Detail about employee's salaries, deduction, generation of paychecks, etc. |

2- Applications of DW

| Sector          | Usage                                                                                                                                                                                                 |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Airline         | It is used for airline system management operations like crew assignment, analyzes of route, frequent flyer program discount schemes for passenger, etc.                                                 |
| Banking         | It is used in the banking sector to manage the resources available on the desk effectively.                                                                                                           |
| Healthcare sector | Data warehouse used to strategize and predict outcomes, create patient's treatment reports, etc. Advanced machine learning, big data enable datawarehouse systems can predict ailments. |

| Sector           | Use                                                                 |
|------------------|----------------------------------------------------------------------|
| Insurance sector | Data warehouses are widely used to analyze data patterns, customer trends, and to track market movements quickly. |
| Retain chain     | It helps you to track items, identify the buying pattern of the customer, promotions and also used for determining pricing policy. |
| Telecommunication| In this sector, data warehouse used for product promotions, sales decisions and to make distribution decisions. |