American Journal of Mathematical and Computer Modelling
2022; 7(4): 55-60
http://www.sciencepublishinggroup.com/j/ajmcm
doi: 10.11648/j.ajmcm.20220704.12
ISSN: 2578-8272 (Print); ISSN: 2578-8280 (Online)

Building a Secured Data Warehouse for a University Staff Management System: A Case Study of Gombe State University, Gombe

Muhammed Kabir Ahmed[1, *], Ahmadu Bappah Muhammad[2], Abubakar Adamu[2], Aishatu Yahaya Umar[1]

[1]Department of Computer Science, Gombe State University, Gombe, Nigeria

[2]Department of Mathematics, Gombe State University, Gombe, Nigeria

Email address:

mkahmed@gsu.edu.ng (Muhammed Kabir Ahmed), bapduk@gsu.edu.ng (Ahmadu Bappah Muhammad), abusiham@ymail.com (Abubakar Adamu), aishatuyu@gmail.com (Aishatu Yahaya Umar)

[*]Corresponding author

To cite this article:

Muhammed Kabir Ahmed, Ahmadu Bappah Muhammad, Abubakar Adamu, Aishatu Yahaya Umar. Building a Secured Data Warehouse for a University Staff Management System: A Case Study of Gombe State University, Gombe. *American Journal of Mathematical and Computer Modelling*. Vol. 7, No. 4, 2022, pp. 55-60. doi: 10.11648/j.ajmcm.20220704.12

Received: December 11, 2022; **Accepted**: January 4, 2023; **Published**: January 17, 2023

Abstract: Universities are among the several organizations with complex activities regarding staff records, payroll, and staff promotion. This work intends to use Data warehouse as a solution to simplify these complex activities. Due to their ability to combine heterogeneous data from several information sources in a single storage location for querying and analysis, data warehousing is gaining importance in terms of strategic decision making throughout time. Data collection is a long-standing practice among organizations. Building a massive data warehouse enables them store all important and relevant information. This information is available, but very few organizations have been able to use it to make informed decisions. Due to the significant role played by the Data Warehouse (DW) and Data Mining in strategic decision making, this work developed a Data Warehouse system that can be used academic staff promotion in a University. The prototype which demonstrated the benefits of Data Warehouse and sensitizes universities in Nigeria to start binding such facilities into their staff management system for the purpose of establishing effective administrative system. The system was implemented under oracle Data Warehouse Builder and is meant to serve as a repository of data for data mining operations. The system offers high degree of accuracy in predicting the case of promotion, staff-students ratio as well as the budget projection.

Keywords: Data Warehouse, Oracle, Management System

1. Introduction

A collection of data collected to aid in decision-making is known as a data warehouse (DW). Additionally, it serves as a database of recent and old information that administrative managers within the company can find useful. Data is often organized so that it is accessible in a form that is ready for analytical processing activities (e.g. Online Analytical Processing [OLAP], Data Mining, querying, reporting and other decision making supportive applications). Because there isn't enough historical data in OLAP for establishment, it can't be used to give in-depth details about how the firm is run. On the other hand, DW offers a central archive of historical data that serves as an integrated platform for data analysis going back in time. "With a Data-Warehouse and Online Analytical Processing (OLAP), users can perform better data analysis and gain better knowledge from the repository data" [1].

Data Warehouse (DW) is a group of decision-supporting technology designed to help administrative workers (executive, manager, and analyst) make choices more quickly and accurately. An Extraction, Transportation, Transformation, and Loading (ETL) solution, an Online Analytical Processing (OLAP) engine, client analysis tools, and other applications that control the process of gathering

data and delivering it to business users would be included in a Data-Warehouse environment in addition to a relational database. "Data Mining, or Knowledge Discovery in Databases (KDD) as it is also known, is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data" [1]. This "encompasses a number of different technical approaches, such as clustering, data summarization, learning classification rules, finding dependency networks, analyzing changes, and detecting anomalies" [2]. According to Nafeez, and Rikita, lecturers and students in Malaysia's Institutions of Higher Education (IHE) are increasingly embracing online learning. 90% of Malaysian IHEs currently have an online learning policy, and nearly 70% have made it mandatory for both professors and students [3]. But there aren't many strategies for systematically gauging the success of online learning deployment. In order to systematically track the use of a blended learning system, the study displays the integration of several online learning system data sources. To track the lecturers' use of blended e-learning technology, the business intelligence and data warehouse approach has been employed to collect, process, integrate, and evaluate the blended learning data. The outcomes have demonstrated that instructors and university administration can successfully monitor the use of blended learning.

Data is now a major component of digitization and a valuable resource for many businesses. Across a heterogeneous application ecosystem, there are numerous data sources that are isolated from one another in different domains and across multiple organizations. To combine data into one system, well-known centralized solution technologies like data warehouses and data lakes exist, however they don't necessarily scale well. As a result, reliable and decentralized methods of data management can offer businesses higher value and give them a competitive advantage over a single central repository. In this work, Majid, A *et al.* explained when and why decentralizing a monolithic data storage for increased scalability is appropriate, as well as how to do it [4]. Data is now a major component of digitization and a valuable resource for many businesses. Across a heterogeneous application ecosystem, there are numerous data sources that are isolated from one another in different domains and across multiple organizations. This study is aimed at developing a University Data-Warehouse that will provide effective automation of staff promotion and welfare, predict or determine the required number of staff to be recruited in some time to come as well aid in budget preparation and projection.

2. Literature Review

Traditional data warehouses fall short of the modern enterprise's expanding demands for the integration and analysis of a wide range of data produced by social, mobile, and sensor sources. The fact that so many businesses have switched to using NoSQL databases for data storage is astonishing. NoSQL databases must be studied as a data source for the modeling and implementation of data warehouses because of the significant role that these types of databases play. [5] Suggests a technique for creating the data warehouse schema from NoSQL databases, which are schema-free databases. As an example of a NoSQL database, the approach begins with the extraction of schemes from document-oriented databases. The MapReduce paradigm was used to do this extraction. The structure was then defined.

Large businesses and hospitals rely on data in the current environment, and the database server will manage this data. In order to obtain the patient's records, it is a difficult process to handle the clinical data. The examination of patients' medical records, illness analysis, population-wide disease research, treatment decision assistance, and enhancement of current healthcare practices all depend on a centralized clinical data repository. Pravin and Manoj; Senda, *et al.;* as well as Azman, *et al*. presented a data warehouse in a medical domain how it can be useful in clinical practice [6], [7, 8]. Pravin and Manoj presented a number of unique difficulties for creating a consolidated clinical data repository because the Bangladesh's existing Electronic Health Records (EHR) are kept in disparate, disconnected sources without a consistent, unique patient identity [6]. The key obstacles to creating a uniform and interoperable centralized clinical data repository are data integration with secure record linking, privacy preservation, quality control, and data standardization. A anonymized National Clinical Data Warehouse (NCDW) platform to support research and analysis based on the results from our prior studies was created. The outcome of the analysis was satisfactory. Thus, Senda, *et al.* suggested the Medical Information System that can conveniently manage the daily patient records using data mining techniques to overcome this kind of problem. For collecting daily clinical data and carrying out ordinary hospital tasks, these databases are helpful [7]. There is intense rivalry in the industry to implement data mining and warehousing concepts to analyze clinical data in order to develop new methods and find more effective ways to carry out the daily operations of hospitals. Similarly Azman, *et al*. indicated that clinical Text Analysis and Knowledge Extraction System are a solution for processing notes for research can be done using natural language processing (NLP) tools like the clinical Text Analysis and Knowledge Extraction System, but maximizing their efficiency for a clinical data warehouse is still difficult [8]. Utilizing the clinical Text Analysis and Knowledge Extraction System, a high throughput NLP architecture was created and a use case for a prediction model was also demonstrated. Over ten years, 1 103 038 patients made up the CDW. The Hadoop data repository was used to build the architecture, and three large-scale symmetric processing servers were used for NLP. The Unified Medical Language System notion unique identifier was linked to each named entity mention in a clinical document (CUI).

Antti *et al*. pointed out that organizations especially universities that use the transactional data have gathered over time to create a more accurate picture of their operations can

make better decisions in today's data-driven world to further their objectives and interests [9]. However, because there are so many of these gathered transactional data, it is difficult to use them simply and immediately for reporting and analysis needs. In order to facilitate management decisions, a data warehouse is required to house the accumulated data that has been gathered from many sources within an organization. In this study, a financial data warehouse with a multidimensional construct that divides time, finance unit, account, and temporal dimensions was created. This warehouse is updated on a regular basis with the accumulated transactional data. The outcome of the result revealed that the query time was greatly reduced.

Even though Antti *et al*. presented a University Data Warehouse and it visualization tool, however it is mainly for financial data [9]. Other aspect of the University Management System were totally excluded especially the aspect of academic activities. However, the desire of this study is to further pursue this issue.

As pointed earlier, there is a Data Warehouse model, designed from a university information system as presented by Youssef. Additionally, it is built on converting an operational database into an information warehouse that decision-makers may utilize for data analysis, forecasting, and prediction. Data extraction, data purification, data transformation, and data indexing and loading are the four stages of data transfer that the concept is based on. The entire system was developed using MS Access 2010 and is designed to act as a data warehouse for data mining tasks [10].

Another system in operation is the On-Line Transaction Processing (OLTP) model (operational database), as designed by Akintola et al. It provides an overview of the creation of data cubes as well as the use of OLAP and data mining tools for data analysis. The goal of the study is to educate enterprises in Nigeria about the advantages of data warehouses and encourage them to begin integrating these features into their enterprise resource management systems in order to make smart business decisions that will help the companies develop quickly.

Similarly, on the observed data as well as a collection of OLAP operators for restructuring and granularity change, OLAP analyses are based. The research looked into a deep integration where the operators of the fundamental data warehouse take into account both data and expertise. The objective was to find hidden patterns in the data. Sadly, this strategy also has its roots in analytic training. The addition of knowledge to data warehouses results in enriched analysis contexts that explicitly represent, handle, and depict objects and their relationships. In order to analyze online usage, the article employs the knowledge data warehouse concept [11].

"The past couple of decades have seen a dramatic increase in the amount of data (information) being stored in electronic format. This is because every enterprise uses a database to store its vital information" [12]. In this regard, internal databases are used as data storage and management systems by dynamic websites, accounting information systems, payroll systems, and stock management systems. "The competition in the market place has led business managers and directors to seek a new way to increase their profit and market power, and that by improving their decision making processes. Hence, the idea of Data-Warehouse and data mining was born" [13].

Another write up by Singh and Chauchan provided an overview of how Artificial Neural Networks are applied to Data Mining process and its suitability for application in "business and economic systems that traditional quantitative tools in statistics and econometrics cannot quantify due to the complexity in translating the system into precise mathematical functions" [14]. Thus they are introducing a concept of Artificial Neural Networks in Data Mining which is "a promising field of research especially given the ready availability of large mass of data (*Data Warehouse*) sets".

3. Justification of the Study

Since processing data results in the extraction of significant information, and studying data and information results in the extraction of knowledge, as well as information. Knowledge play significant roles in many human endeavors. Large data repositories, such as data warehouses, are required due to the issue of managing, storing, and analyzing the massive volumes of data that are generated constantly by diverse sources. Due to the aforementioned, data warehousing has gained a lot of scientific and industry attention (DW). Numerous research over the past few years have shown various problems and difficulties in the field of data warehousing [15].

Presently, almost all the Data Warehouse design and utilization for universities (as established by Youssef; and Akintola *et al*. focused more on their application to students activities and welfare, such as payment of fees; classroom activities; course allocation; etc [10] and [1]. These however, exclude the aspect of Employee Management which is handled by the Registry Department. Of course, this is also crucial in the smooth running of a university, as it acts as the central administration of any institution. Therefore, there is need to come up with Data Warehouse that can support decision making, both at the Strategic and Tactical level of a University, especially regarding the staff management. This will provide enough information for effective management activities.

4. Research Methodology

In this study, a multidimensional modeling approach was employed. It calls for specific design approaches similar to conventional database design approaches as indicated in table 1.

Table 1. *Database Design Method.*

Step	Input	Output
Analysis of Operational System	Information regarding the operational systems	Database schemes
Requirements Elicitation	Database scheme	Specifications for data warehouse
Conceptual design	Database scheme and Specifications	Conceptual schema
Logical design	Conceptual schema	Logical schema
Physical design	Logical schema	Physical schema

The following five stages make up the database design:

1) The analysis of operational systems, which is the initial phase, aims to gather data on the staff promotion, obtaining number of required staff to be recruited and Budget projection.
2) The second stage involves assembling and sieving the system requirements; at this phase the prospective users of the proposed system were also involved. The guidelines of the promotion for academic staff was studied with users and a template was adopted based on the condition of service of the university. Table 2 show the promotion template use in Gombe State University. The promotion criteria is based on obtaining minimum required point for the rank to be promoted.

Table 2. *Gombe State University Guideline for Promotion and Appointment of Academic Staff.*

S/N	Criteria	Professor		Reader		Senior Lecturer		Lecturer 1		Lecturer II		Asst. Lecturer		GA
		APPT	PROM	APPT	PROM	APPT	PROM	APPT	PROM	APPT	PROM	APT	PROM	APPT
1	Qualifications	10	10	10	10	10	10	7-10	7-10	≥10	7-10	7	7	4
2	Publications	≥70	≥60	≥60	≥45	≥35	≥25	15	≥8	≥5	≥5	-	-	-
3	Teaching experience	12-15	12-15	10-15	9-15	7-15	7-15	3-6	-6	≥3	≥3	-	-	-
4	Academic leadership / Administrative experience	2-5	1-5	1-5	1-5	1-5	1-5	1-5	1-5	1-5	0-5	-	-	-
5	Membership of professional bodies	2-3	1-3	1-3	1-3	1-3	1-3	0-3	0-3	0-3	0-3	-	-	-
6	Postgraduate supervision	2	2	2	2	1-2	1-2	-	-	-	-	-	-	-
7	Community / Public service	1-5	1-5	1-5	1-5	1-5	1-5	0-5	0-5	0-5	0-5	-	-	-
8.	Academic related activities	1-3	1-3	1-3	1-3	1-3	1-3	-	-	-	-	-	-	-
9	Ability to attract grants	1-3	1-3	1-3	1-3	1-3	1-3	-	-	-	-	-	-	-
10	Interview performance	7-10	-	7-10	-	7-10	-	5-10	-	5-10		5-10	-	5-10
	Total points obtainable	126	106	116	91	91	71	54	37	46	31	17	7	14
	Minimum points required	100	89	94	71	66	48	31	19	25	15	12	7	9

Next is the Algorithm design which will be transform into various activities and procedures that will be implemented in the Data Warehouse. Three different algorithms were developed (Algorithm 1: Staff Promotion, Algorithm 2: Required staff prediction and Algorithm 3: Budget Projection) for the Data Warehouse operation.

Algorithm 1: Staff Promotion

Input: Qualification, Publication, Other_Point, Total_Point_Obtained, Required_Point

Output: Qualified, Not-qualified, Recommended, Not_recommended

ForEach application:

Compute point for Qualification and Publication

Obtain Total_Point_Obatained

If Point >=Required_Point Then Status "Qualified" and "Recommended"

Else Status "Not_qualified" and "Not_recommended"

Return Status

End

Algorithm 2: Required_Staff Prediction

Input: Staff_No, Students_No

Output: Balanced, Bottom_Heavy, Top_Heavy

ForEach Department:

Compute Staff-Students ratio

Obtain Staff_Mix

Get Dept_Status

If Dept_Status = "Balanced" Then "No Recruitment"

ElseIf Dept_Status = "Bottom_Heavy" Then "Recruit Senior Lecturers"

Else "Top-Heavy" = Then "Recruit Junior Lecturers"

End

Algorithm 3: Budget Projection

Input: Salary_Scale, Rank, Required_Staff

Output: Additional_Budget

ForEach Required_Staff

Get Rank and Salary From Salary_Scale

Compute Total_Salary for all Required_Staff

Additional_Budget = Total_Salary

Return Additional_Budget

End

At this phase the ER diagram overall system was developed from the three algorithms. The ER diagram is

shown in Figure 1.

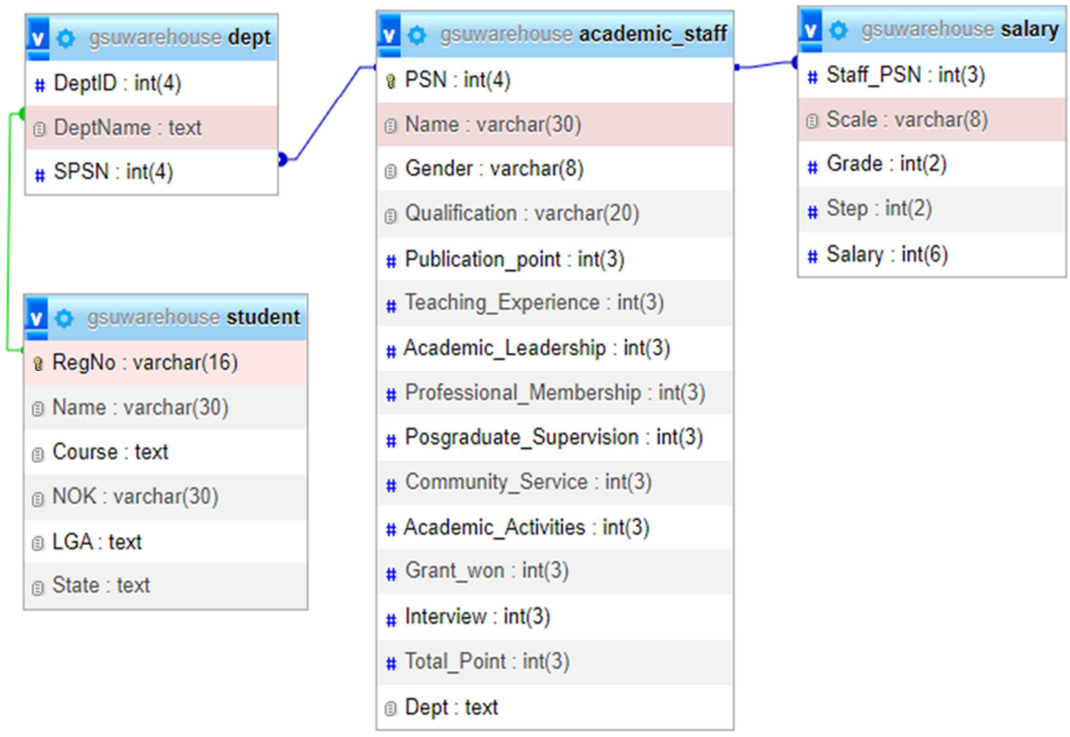

Figure 1. *ER Diagram.*

Finally, there is a phase of physical design that considers difficulties unique to the implementation tools chosen, such as indexing and allocation.

5. System Implementation and Experimental Setup

Oracle Data Warehouse Builder was used in the implementation and performance evaluation was done after the system deployment. A two hundred (200) curriculum vitae of academic staff was extracted from the data warehouse. The data comprises of one hundred (100) staff that are due and qualified for promotion while the second part of the data too, contain one hundred (100) staff that are not qualified for promotion. The essence is to demonstrate how the system can identify those that are qualified for promotion. The outcome of our prediction was tested and evaluated based on the following parameters.

Accuracy: Accuracy is the level of correctness to calculate the performance of the software system. Accuracy=TP+TN/(TP+TN+FP+FN).

Sensitivity: It is an evaluation of how a software system properly identifies a qualified promotion case. Sensitivity = TP/(TP+FN).

Specificity: It is an evaluation of how a software system properly identifies a not qualified promotion case. Specificity = TN/(TN+FP).

The table 3 present the Academic Staff Prediction result.

Table 3. *Academic Staff Prediction Result.*

Parameters	Values
TP	89
TN	86
FP	11
FN	14

6. Result and Discussion

This research is intended to show the design and implementation of a data warehouse using Oracle Data Warehouse Builder. The system was able to calculate the staff-students ratio and make the budget projection based on the number of staff that are promoted. And for the promotion prediction the evaluation is shown in table 4. The accuracy of the proposed system is better than the one obtained by a relevant work which is 0.853 [16]. This shows that there is an improvement in our work.

Table 4. *Prediction Evaluation.*

Parameters	Result
ACCURACY	0.875
SENSITIVITY	0.864
SPECIFICITY	0.887

7. Conclusion

From the evaluation of the system it can be seen the Accuracy of the prototype system is 0.875, Sensitivity is 0.864 and Specificity is 0.887 which are all more 80%. It can

be concluded that the system offers high level of accuracy which can be used to assess staff promotion case in a Nigerian University. The proposed system may be applied by any Higher Education Institution desiring to integrate business intelligence into their tactical decision-support processes, especially in a University, in order to discover hidden treasures in the Data Warehouse. In future we hope to cover the aspect of Non-teaching staff promotion and other university's administrative related activities such as procurement.

References

[1] Akintola, K. G., Adetunmbi, A. O., & Adeola, O. S. (2011). Building Warehouse and Data Mining from Course Management System; A Case Study of FUTA Course Management Information System. International Journal of Database Theory and Application, 13-20.

[2] Sherenaz, A. B., Alesso, M., & Mauro, M. (2005). Anormaly Detection in Computing Network. A State of the Art Review. Journal of Wireless Mobile Network Ubiquitors Computing and Dependable Application, 29-64.

[3] Nafeez, A. F., & Rikita, M. (2018). Design of Data Warehouse for Medical Information System Using Data Mining Techniques. 5th IEEE International Conference on Parallel Distributed and Grid Computing (PDGC) (pp. 20-22). Solan, India: Solan Computing Press.

[4] Majid, A., Dmitriy, D., Brihat, S., Xiaoyuan, C., Jason, B., Steven, B., Ron, P. (2019). Development and Application of a High throughput Natural Language Processing Architecture to convert all Clinical Data Warehouse into Standardized Medical Vocabularies. Journal of the American Informatics Association, 1364-1369.

[5] Jie, L., Wei, Y., Wan, Z., Xinyu, Y., Hanlin, Z., & Wei, Z. (2017). A Survey on Internet of things: Architecture, Enabling Technologies Security and Privacy and Applications. Internet of Things Journal IEEE, 1-17.

[6] Pravin, C., & Manoj, K. G. (2018). Comprehensive Survey on Data Warehousing Research. International Journal of Technology, 217-224.

[7] Senda, B., Ahlem, N., & Faiez, G. (2019). Design a Data Warehouse Scheme from Document Oriented Database. 23rd International Conference on Knowledge Based and Intelligent Information and Engineering Systems. Procedia Computer Science. Elsevier, Science Direct.

[8] Azman, T., Mohamad, S. A., Suwannit, C. C., & Mohd, H. M. (2017). Data Warehouse System for Blended Learning in Institutions of Higher Education. e-academia Journal, 144-155.

[9] Antti, L., Juha-pekka, J., Mikko, R., Tommi, M., & Timo, L. (2017). Migrating from a centralized Data Warehouse to a Decentralized Data Platform Architecture. Retrieved from http://www.solita.com

[10] Youssef, B. (2012). A Data Warehouse Design for a Typical University Information System. Retrieved from Lebanese Association for computational Sciences: http://www.lacsc.org

[11] Quafafou, M., Naouali, G., & Nachouki, G. (2005). Knowledge Data Warehouse: Web Usage OLAP Application. International Conference on Web Intelligence, (pp. 19-22).

[12] Pant, S., & Hsu, C. (1995). Information Resource Management Association. International Conference, (pp. 21-24). Georgia, Atlanta.

[13] Xingquan, Z., & Ian, D. (2007). Knowledge Discovery and Data Mining. Global Research Collection.

[14] Singh, Y., & Chauchan, A. S. (2013). Application of Data Mining Using Artificial Neural NetworkSurvey. International Journal of Database Theory and Application.

[15] Earlvon, F. L., John, K. J., Mark, J. K., & Pagatpat, D. D. (2018). Development of a University Financial Data Warehouse and its Visualization Tool. 3rd International Conference on Computer Science and Computational Intelligence (pp. 587-595). Elsevier, Science Direct.

[16] Kefi, H. and Koppel, N. (2011) 'Measuring data warehousing success: an empirical investigation applying the DeLone and McLean model', Int. J. Data Analysis Techniques and Strategies, Vol. 3, No. 2, pp. 178–201.