University of Ibadan
Journal of Science and Logics in ICT Research

An Integrated Multi-Dimensional Data Warehouse for University Payroll Management

Onyekwelu, A. O.,
University of Ibadan, Ibadan, Nigeria
talk2obi007@yahoo.com

Adeyemo A. B.
University of Ibadan, Ibadan, Nigeria
sesanadeyemo@gmail.com

Abstract

Payroll is traditionally the most data driven function relating to employee information. The challenge for many large organizations lies in the fragmented nature of payroll. Payroll data generally sits siloed or isolated in a multitude of different local systems. The implication of this is very frustrating: Given the time- consuming manual labour involve in aggregating the payroll data. The lack of integrated consolidated payroll data in a single location as well as the functional arrangement of these data makes it difficult to access timely information and carry out effective analysis to support management decision. This work presents an integrated multi-dimensional data model for a University Payroll System. The model integrates all disparate silos of payroll data sources into a single location for ease and speed of reporting and efficient analysis. The model was simulated and tested using Talend Open Source Data Integration tool while MySQL was used for data storage. Reports were developed from the integrated multi-dimensional data warehouse using a reporting tool, and performance was evaluated and compared relative to same reports generated directly from the various disparate payroll systems. The time it takes to obtain reports from the integrated data warehouse was a lot faster and lot easier than having to obtain same reports from each payroll system and manually aggregating the reports from each of the multiple disparate payroll systems. One obvious reason is that all data in are now integrated in a single location instead of wasting time traversing various payroll data sources. In conclusion, the integrated multi-dimensional data warehouse for payroll system will improve information access, reduce drastically the time to reconcile and obtain reports and provide for far more efficient information analysis for decision makers in the University to make effective strategic decisions based on facts

Keywords**:** *Data warehouse, University payroll management, Payroll system, Talend open source data integration*

1. Introduction

How to extract and retrieve timely and efficient information from very large data in different operational database sources and legacy systems has become an important issue especially with the rapid development of information technology [7]. Managers and policy makers of large enterprises are concerned with this issue as the need to always base their decision on solid, accurate, timely, well organized data during decision making becomes very essential. In this digital era, organizations are increasingly relying on database as the main component of their record keeping system. Examples include payroll system, stock management, student records system amongst others; all rely on internal databases to store and manage their data. However, at the same pace the amount and detail of information contained in such database systems grows. Also grows the concern that in few years most of it may be lost, when the current hardware, operating system, or actual application become obsolete and turn the data repositories unreadable thus posing a risk of losing vital historical information which are very important for making strategic business analysis and decisions. Historical and current data are both very essential to get strategic information.

It is often a challenging task for organizations to effectively manage, integrate and analyse huge volume of data from various sources fast in order to make timely and effective decision. In enterprises payroll systems, data may reside in different databases, computers and servers in various locations for different reasons ranging

Onyekwelu A. O. and Adeyemo, A. B. (2020). An Integrated Multi-Dimensional Data Warehouse for University Payroll Management, *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR)*, Vol. No. 1, pp. 112 – 119.

from policies of the company, location, or even security. The concept of data warehouse was developed to provide single access point to data from variety of operational sources. Data warehouse provides a central repository in which data from multiple sources are brought together, integrated, manage and easily accessible to use to obtain faster reports and make effective analysis [4].

Figure 1 presents a schematic diagram of a data warehouse. This architecture includes tools for extracting data from multiple operational databases and external sources; for cleaning, transforming and integrating this data; for loading data into the data warehouse; and for periodically refreshing the warehouse to reflect updates at the sources and to purge data from the warehouse, perhaps onto slower archival storage. In addition to the main warehouse, there may be several departmental data marts. Data in the warehouse and data marts are stored and managed by one or more warehouse servers, which present multidimensional views of data to a variety of front-end tools: query tools, report writers, analysis tools, and data mining tools. Finally, there is a repository for storing and managing metadata, and tools for monitoring and administering the warehousing system.

The warehouse may be distributed for load balancing, scalability, and higher availability. In such a distributed architecture, the metadata repository is usually replicated with each fragment of the warehouse, and the entire warehouse is administered centrally. An alternative architecture, implemented for expediency when it may be too expensive to construct a single logically integrated enterprise warehouse, is a federation of warehouses or data marts, each with its own repository and decentralized administration Warehouse [5].

Payroll traditionally is the most data driven function relating to employee information. Many companies have recognized the financial and strategic importance of their employees whose total cost (salary) represents at least 50-60% of the total cost base for most companies according to study by Deloitte [6]. Yet, in many organizations the ability to really understand their employee base through quantitative analysis remains under-developed. Most organizations make decisions based largely on instinct, experience and gut feeling instead of solid, accurate, timely, well organized data. More importantly, many organizations consider Payroll to hold the ultimate version of the truth and hence the most reliable source of data. Hence the quest for better Human Resource (HR) analytics by building on payroll data to enhance ease of access to timely and efficient information. However, the challenge for many large organizations lies in the fragmented nature of payroll.

Payroll data generally sits siloed or isolated in a multitude of different local systems. The implication of this is very frustrating: Given the time-consuming manual labour involve in aggregating the payroll data. Thus, there is a need for large organizations to have a single location for storage, analysis and sharing of payroll data where users and management can easily utilize to make effective and quality business decisions, rather than trying to traverse the multiple data sources that exist to generate reports [1]. An integrated university payroll data warehouse provides a single location where university management and users can gain easy access to integrated employee data gathered from various operational payroll systems and legacy systems in various locations. It can facilitate efficient storage, enhances timely analysis and increases real-time decision-making process.

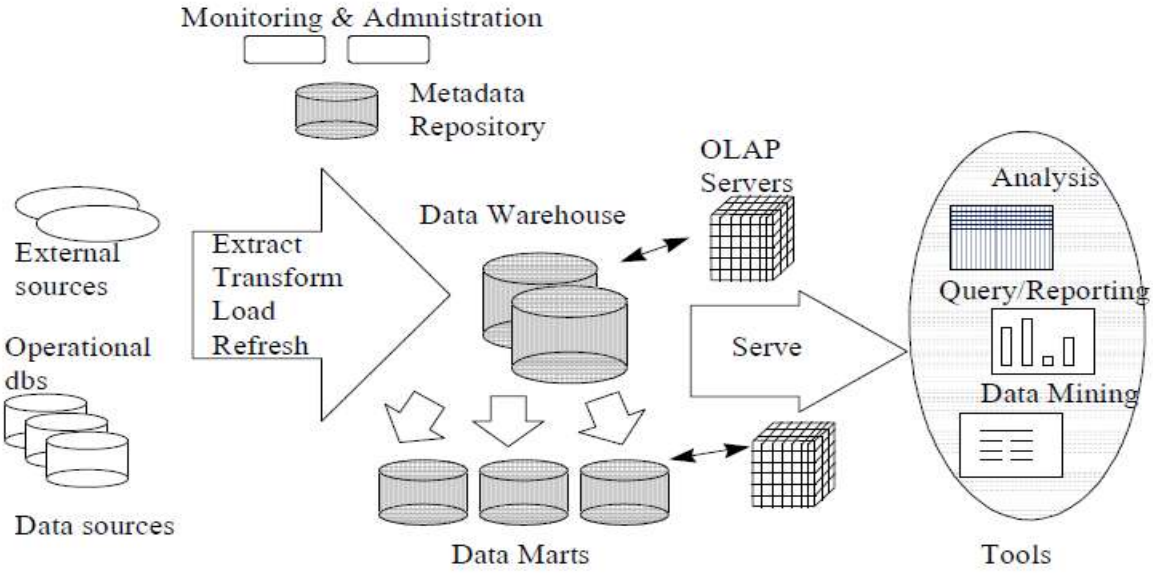

Figure 1. The Basic Architecture for Data Warehouse [5]

2. Materials and Methods

The requirements gathering phase of the study involved the use of the following requirements gathering methods: Observation of the operations of the existing payroll systems, Interview of the users and study of the existing business reports and requirements of the institution. In the Observation stage a detailed notation of behaviours, events and the context surrounding the payroll system processes in the university was carried out. Physical observation of what tools are used to collect, analyse and disseminate information was conducted.

In this particular case study university for instance, if there is a management staff (Vice Chancellor, Bursar etc.) request for information from the main payroll unit in the University (for example to get the list of staff recruited from 2010 to 2017 into the University system), the request is reviewed by the system analyst in the payroll unit to ascertain if a similar report has been generated in the past. If it is a new request, the system analyst in the main payroll unit will have to reach out to other analysts managing the other sub payroll system in other locations in the university campus to intimate them with the new request. Each of the system analysts in the three (3) payroll units will then have to search through payroll database tables to provide this information and send this to the main analyst in the major payroll unit. This process can take a day or two depending on the efficiency of the analysts.

The main analysts on receiving these reports on staff recruitment from other sub payroll units which often comes in different formats and style, must then clean, standardize, integrated and consolidated them into a single report for onward passage to management to assist them in decision making. This process is time consuming and could span for days and is also prone to error due to the manual process involved. Interviews were held with the Head of the Payroll unit, staff, users and few stakeholders. The purpose of this interview was to identify the existing payroll processes and determine the essential entities that are usually required to meet their reporting and analysis needs. It was also, to learn about the operation and processes of the payroll systems, its rules and to get useful data. The interviews showed that there were three payroll points, each of them ran different customized application based on Microsoft SQL database (there is a plan to make them all use a uniform payroll application software).

There were also legacy systems that hold vital historical information which is often referenced to carry out analysis. Some of this information

had been extracted to Microsoft Excel spreadsheet formats from which the data was integrated to carry out analysis and consolidated reports, from these disparate payroll system and legacy system. Discussions with the various stakeholders of the payroll systems highlighted the following business areas for critical analysis: Employee distribution pattern in terms of state and local government area, Employee earnings pattern, Employee deduction or spending pattern, Compliance to laws and regulations, Employee payroll status and other bio-data patterns. These business areas are analysed to understand the business process and events that led to generation of data and how the data model to be developed will meet the business requirements and solve the inherent problem of data inaccessibility, lack of integration and speed of reporting.

2. 1 Data Warehouse design

This research work developed a prototype integrated multi-dimensional data warehouse model that will meet the business requirements and also provide easy accessibility to data for effective analysis and fast reporting. This will enable the institution gain deeper insight on their data and provide easy access to data to meet the various business requirements for analysis and decision making.

The data warehouse design proposed for the case study university will integrate data from the four different sources within the university. Payroll history from the University main campus, College of Medicine, Grant payroll system and archive of historical payroll records. These payroll data from various disparate sources will all be integrated to give users the ability to generate vital reports that can be used for decision making. The schematic diagram of the design is given in Figure 2.

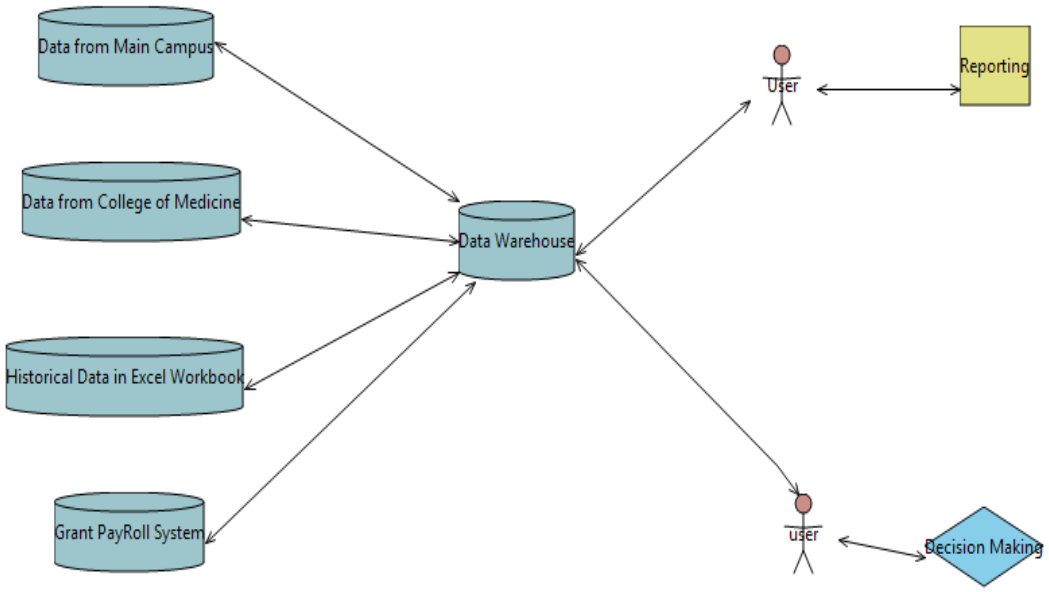

Figure 2: Schematic diagram of the university payroll system

The Entity Relation (ER) Modelling is a logical design technique that seeks to eliminate data redundancy. ER models show relationship between data. It is based on two basic concepts: entities and the relationship between those entities. The ER diagram showing the entities and their relationships in the existing payroll system is shown in Figure 3.

In data warehousing, dimensional modelling is simpler, more expressive, and easier to understand than ER modelling. Unlike ER modelling, dimensional is powerful in representing the requirements of the business as it is structured basically for optimized querying and analysis. (Kimball, 1996). In the design of a new payroll system a data warehouse using dimensional modelling is proposed using the star schema. The star schema is the focal point of data warehouse development, as it provides a de-normalized database structure for reporting the data that is vital to the business needs.

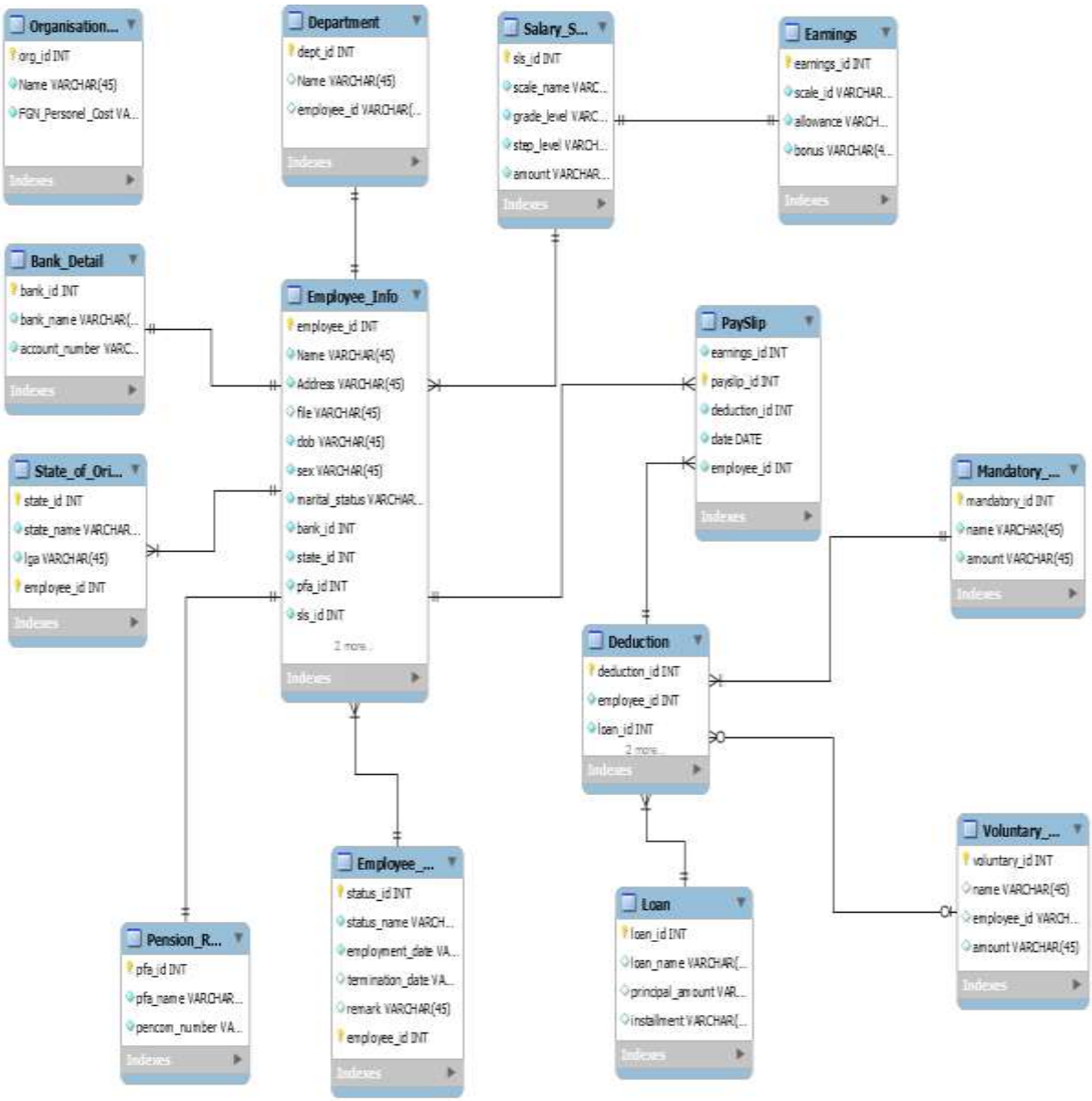

Figure 3: ER model of the existing payroll system,

3. System Implementation

End users' requirements collected and documented during the requirements gathering process which included observation, interview and also studying of the existing documents and reports were used to design a new payroll model. A summary of the information requirements derived from the payroll units are:

1. Keep track of new staff recruitment of the University on a yearly/monthly basis
2. Keep track of retirees, dead staff of the University on a yearly/monthly basis
3. Keep track of total payroll amount of staff on a yearly/monthly basis.
4. Keep track of the compliance to statutory deductions such tax, pension on a monthly/yearly basis
5. Keep track of the percentage of staff on various loans.
6. Keep track of the payroll status of each staff on a yearly/monthly basis

The ETL tool in Talend Open Studio was used to develop a prototype Employee Star Schema and Payroll Star Schema (figure 3). The Payroll database, which is a transactional database was one of the base databases for the development of the Employee Star Schema. The "Payroll" database is used for keeping track of employees' job assignments, salary and personal information. Records include employee identification number, e-mail address, job identification code, salary, and department. The Talend Open Studio tool was used to build the Employee Fact and Dimension tables (figure 4). The Payroll star schema was developed from the HR (Human Resources) database which is a Microsoft SQL database. The Payroll Star schema was linked to the Employee Star Schema previously created.

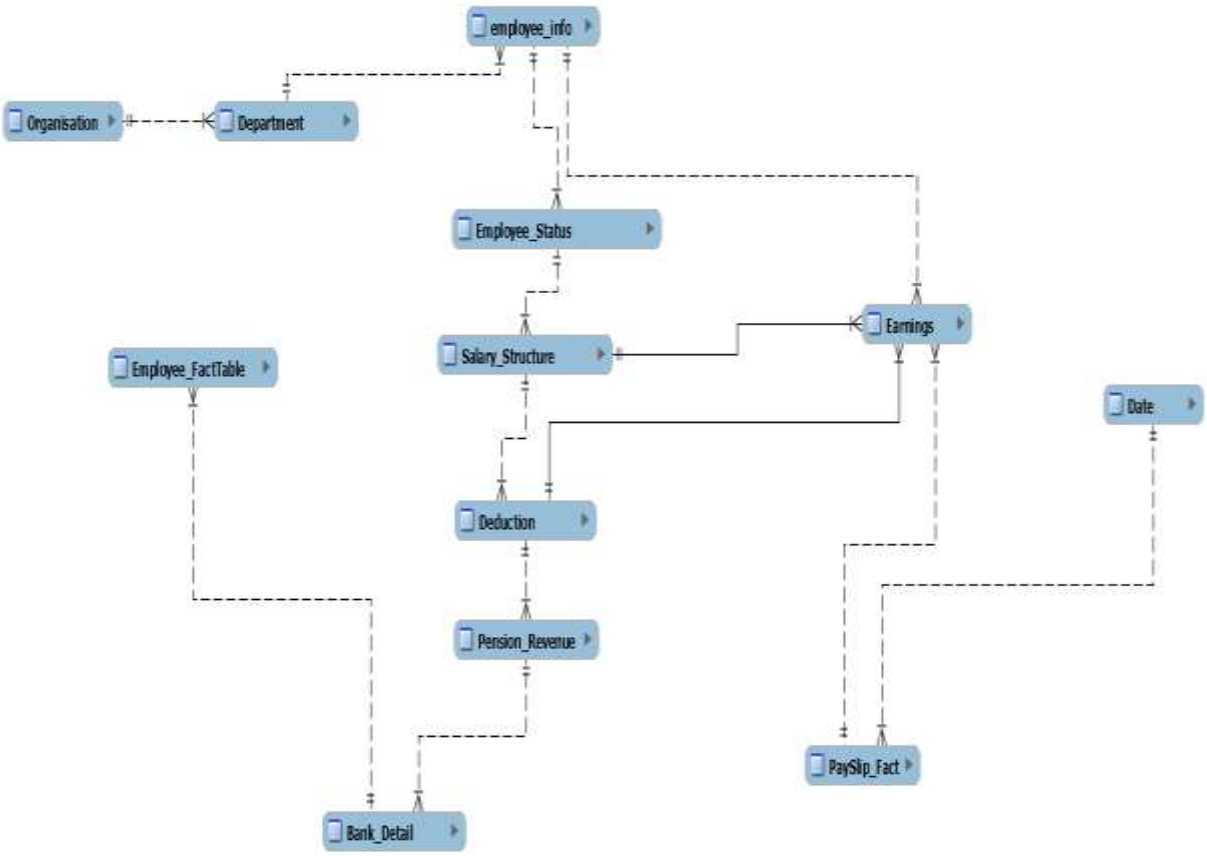

Figure 3: Combined Employee and Payroll Star Schema

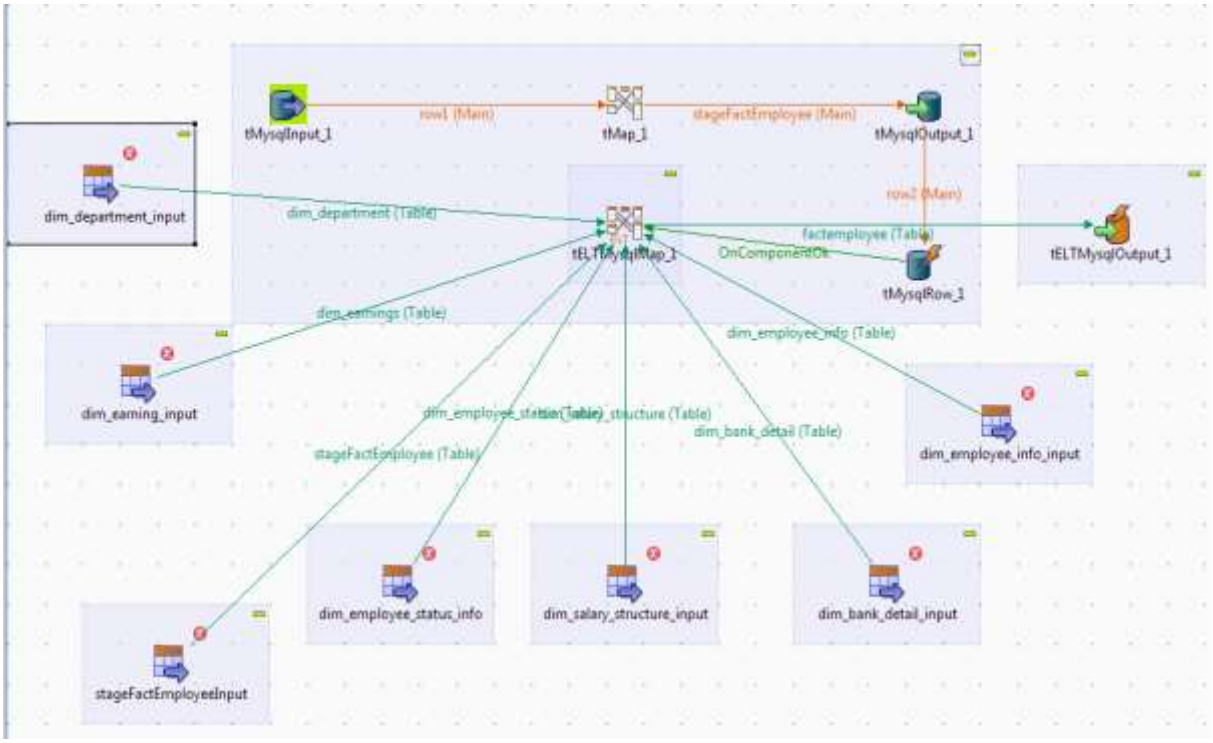

Figure 4: Building the Employee Fact Table.

Figure 5 presents a sample report generated from the data warehouse prototype developed. Reports can be easily generated from the prototype data warehouse when compared with the transactional payroll system because of the fewer number of tables involved in the report generation process.

PF NO	SURNAME	deduction
20001	STAFF 1	21,444.80
20002	STAFF 2	68,910.73
20003	STAFF 3	1,669.40
20004	STAFF 4	7,232.98
20005	STAFF 5	52,428.80
20006	STAFF 6	16,142.48
20007	STAFF 7	29,037.98
20008	STAFF 8	60,624.67
20009	STAFF 9	24,792.60
20010	STAFF 10	24,849.19
20011	STAFF 11	43,399.19
20012	STAFF 12	20,449.19
20013	STAFF 13	6,680.94
20014	STAFF 14	128,168.03

Figure 5: Report of staff deductions

4. Conclusion

In this study a prototype integrated multi-dimensional data warehouse for payroll system within the context of a university environment was developed to integrate disparate silos of payroll systems data for simpler and improved access to information, reporting and analysis. The prototype will also help end users and management to provide answers to complex queries and carry out more effective payroll analysis. Top management and heads of departments are generally concerned with aggregated reports. This data warehouse model has the ability to store and organize data needed for informative analytical processing over a long historical time perspective and which can be used for analysis that supports business decisions at many levels, from strategic planning to tactical planning as well as prediction. In order to implement a data warehouse, it is necessary to have a strong business commitment spearheaded by an executive sponsor who believes in the project and is responsible for keeping the project team motivated.

References

[1] Awogbamila Y. and Adeyemo A. B. (2012). Design of An Integrated Agrarian Data Dimensional Data Warehouse 3(4), 21–30. A *Journal of Computing, Information Systems & Development Informatics* Vol. 3 No. 5

[2] Bassil, Y. (2012). A Data Warehouse Design for A Typical University Information System. *Journal of Computer Science & Research*, 1(6), 12–17. Retrieved from http://arxiv.org/abs/1212.2071

[3] Daniel L.M, Mark A.R.K. (2003). From ER Models to Dimensional Models: Bridging the Gap Between OLTP and OLAP design, Part I. *Business Intelligence Journal*.

[4] Kimball, Ralph (1996). The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses. John Wiley & Sons.

[5] Kimball, R. and Ross, M. (2002). The Data Warehouse Toolkit. Second Edition. Wiley Computer Publishing. New York.

[6] https://www2.deloitte.com/content/dam/Deloitte/us/Documents/human-capital/us-cons-workforce-costs-are-hiding-with-laborwise.pdf.pdf

[7] Bo Xiao, Chaolin Zhang, Yici Mao, and Guohui Qian (2015). Review and exploration of metadata management in data warehouse