# An Improved Data Warehouse Architecture for SPGS, MAUTECH, Yola, Nigeria

**Ibrahim Inuwa[1] and E.J. Garba[2]**

[1]Department of Information Technology, Modibbo Adama University of Technology, Yola. Adamawa State Nigeria.
ibrahiminuwa1985@gmail.com. +2348065273810.
[2] Department of Computer Science, Modibbo Adama University of Technology, Yola. Adamawa State Nigeria.
e.j.garba@mautech.edu.ng +2348036943881.

.

## Abstract

*Data warehousing is a very important contemporary technology that is useful in decision making, relating it to software development, the data warehousing technology is indeed a very new discipline and does not until now offer well established approaches and procedures for the development process in the educational sector. In this paper, we have considered using SPGS MAUTECH, Yola as our case study and present the design of a proposed data warehousing architecture within the context of the University atmosphere, to better incorporate systems for simpler and improved data analysis, reporting and querying activities. We have considered the ideologies of data warehousing in the course of this study and demonstrated how data can be incorporated from diverse heterogeneous source systems into a sole historical repository that is capable of supporting Decision Support System (DSS) for University administrators and other end users.. University data management can step up towards adopting and implementing this proposed architecture.*

**Keywords -** Data Warehouse, Architecture, DSS, Decision Making.

## I. Introduction

Data warehouse (DW) has become an important knowledge in the realm of Information Technology (IT) today, for organizations to make quick and precise decisions, data warehousing is one of the very best technology to be consider. Educational institutes, industries or businesses, need to advance their information record system so as to continue in the modest setting. The establishments have to increase their efficiency and effectiveness in maintaining the sequence of activities, in their planning, decision making processes, and analytical needs.

In the era of big data, organizations today rely of huge quantity of data from diverse sources and need to integrate this data in a speedy manner to gain any strategic advantage out of the data. DW is becoming increasingly popular in organizations due to the need for enterprises to gather all of their data in a single place for in-depth analysis and also to segregate such analytical work form on-line transaction processing systems [2]. DW is an essential current issue for many establishments and is relatively a new field in the realm of IT. As data warehousing is a new field, a small number of research has been done regarding the characteristics of academic data and the complexity of analyzing such data. Educational institutions measure success very differently from business oriented organizations and the analyses that are meaningful in such environments pose unique problems in DW. The DW database is used to store information that will satisfy decision making requests. A very common problem with enterprises is the difficulty to access corporate data, complete and integrated information of the enterprise that can satisfy decision-making requests. A paradox occurs: data exists but information cannot be obtained. In general, a DW is constructed with the goal of storing and providing all the relevant information that is generated along the different databases of an organization [1]. Nowadays, almost every enterprise uses a database to store its vital data and information. For instance, dynamic websites, accounting information systems, payroll systems, stock management systems all rely on internal databases as a container to store and manage their data [14]. In addition, online applications had no considerable amount of past data because they discarded their historical data as quickly as possible in the name of high performance. Hence, organizations had heaps of data and very little information [8].

Architecture is a fixed guideline to stick to when constructing or building a DW. Data warehousing is quite a large and complex thing to do, architecture is necessary for success. DW architecture is a way of representing the overall structure of data, communication, processing and presentation that exists for end-user computing within the enterprise [10]. The Architecture process specifies elements of the technical foundation and architectural design of the DW. Throughout the process, the focus is on the integration of many different products and various DW components to provide an extendible and scalable architecture. The Technical Architecture, DW Architecture and Infrastructure Roadmap are defined to outline the design and implementation of the architecture. The DW Architecture provides an integrated data warehouse environment while delivering incremental solutions. The architectural design focuses on the application of a centralized data warehouse, data marts, individual marts, metadata repositories, and incremental solution architectures. As the process continues, the development and execution of the integration plans are completed and the compliance of incremental solutions with the strategic architecture is validated [6].

Making a good choice for data warehouse development methodologies requires thorough understanding of two main data warehousing methodologies, namely bottom-up and top-down approach. Understanding of similarities and differences provides solid foundation knowledge for an organization before applying it [9]. Inmon's top-down architectural approach includes information systems and their databases from all departments of an organization. He named this monstrous size of database as Corporate Information Factory (CIF). This approach insures that complete information is consistent because all departmental information originates from a single Atomic DW [7]. On the other hand, Kimball's bottom-up approach builds the data marts independently at different times while the business requirements become available from each department. These data marts are later combined and merged into a corporate Data Warehouse. [11].

DW development is a very motivating problem, because relating it to software engineering; the DW is indeed a very new discipline and does not until now offer well established approaches and procedures for the development process most especially in the educational sector. Hardly have you seen a DW architecture that is designed specifically for a University DSS. The architectural methodologies of Inmon and Kimball joint together can form DW architecture that can be suitable for a University DSS. It is in this regard we have proposed for DW architecture for a University DSS in the SPGS MAUTECH. This paper provides an improved Data Warehousing architecture that could assist Universities to discover knowledge and improve services. In this paper, section 2 has reviewed the available literature which relates to the study; we discussed the different architectures of DW and analyze their structures and features. In section 3, we have compared the Inmon's and Kimbal's Architectural Methodologies. In section 4, we have presented our proposed DW Architecture for a University DSS. Section 5 provides a brief discussion and conclusion based on the study.

## II. Literature Review

### *The Data Warehousing Architectures*

We start by reviewing the Traditional Data Warehousing architecture; it encompasses of the following components [17]:

i. Data sources as external systems and tools for extracting data from these sources.

ii. Tools for transforming, which is cleaning and integrating the data.

iii. Tools for loading the data into the DW.

iv. The DW as central, integrated data store.

v. Data Marts as extracted data subsets from the DW oriented to specific business lines, departments or analytical applications.

vi. A metadata repository for storing and managing metadata

vii. Tools to monitor and administer the DW and the extraction, transformation and loading process.

viii. An OLAP (online analytical processing) engine on top of the DW and Data Marts to present and serve multi-dimensional views of the data to analytical tools

ix. .

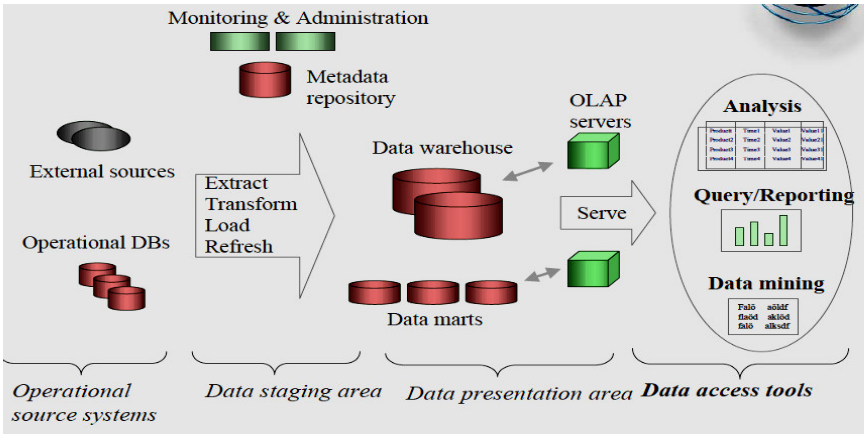


**Figure 1: The Traditional Data Warehousing Architecture [17].**

**x. Tools that use data from the DW for analytical applications and for presenting it to end-users.**

This architecture exemplifies the basic idea of physically extracting and integrating mostly transactional data from different sources, storing it in a central repository while providing access to the data in a multi-dimensional structure optimized for analytical applications [18]. However, the architecture is rather old and, while this basic idea is still intact, it is rather unclear and inaccurate about several facts:

Firstly, most modern data warehousing architectures use a staging or acquisition area between the data sources and the actual DW. This staging area is part of the extract, transform and load process (ETL process). It temporarily stores extracted data and allows transformations to be done within the staging area, so source systems are directly decoupled and no longer strained [18]. Secondly, the interplay between DW and Data Marts in the storage area are not completely clear. Actually, in practice this is one of the biggest discourses about data warehousing architecture with two architectural approaches proposed by Bill Inmon and Ralph Kimball [12]. Inmon places his data warehousing architecture in a holistic modeling approach of all operational and analytical databases and information in an organization, the Corporate Information Factory (CIF). What he calls the atomic DW is a centralized repository with a normalized, still transactional and fine-granular data model containing cleaned and integrated data from several operational sources [7].

Inmon's approach, also called enterprise DW architecture by [18] is often considered a top-down approach, as it starts with building the centralized, integrated, enterprise-wide repository and then deriving Data Marts from it to deliver for departmental analysis requirements. It is however possible, to build the integrated repository and the derived Data Marts incrementally and in an iterative fashion. Kimball on the other hand proposes a bottom-up approach which starts with process and application requirements [11].With this approach, first the Data Marts are designed based on the organization's business processes, where each Data Mart represents data concerning a specific process. The Data Marts are constructed and filled directly from the staging area while the transformation takes places between staging area and Data Marts. The Data Marts are analysis-oriented and multi-dimensional as described above. The DW is then just the combination of all Data Marts, where the single Data Marts are connected and integrated with each other via the data bus and so-called conformed dimensions that are Data Marts use, standardized or 'conformed' dimension tables. If two Data Marts use the same dimension, they are connected and can be queried together via that identical dimension table. The data bus is then a net of Data Marts, which are connected via conformed dimensions. This architecture (also called Data Mart bus architecture with linked dimensional Data Marts by [18] therefore forgoes a normalized, enterprise-wide data model and repository.

In figure 2 there are a number of options for architecting a Data Mart. For example:

i. Data can come directly from one or more of the databases in the operational systems, with few or no changes to the data in format or structure. This limits the types and scope of analysis that can be performed. For example, you can see that in this option, there may be no interaction with the DW Meta Data. This can result in data consistency issues.

ii. Data can be extracted from the operational systems and transformed to provide a cleansed and enhanced set of data to be loaded into the

v. .

Data Mart by passing through an ETL process. Although the data is enhanced, it is not consistent with, or in sync with, data from the DW.

iii. Bypassing the DW leads to the creation of an independent Data Mart. It is not consistent, at any level, with the data in the DW. This is another issue impacting the credibility of reporting.

iv. Cleansed and transformed operational data flows into the DW. From there, dependent Data Marts can be created, or updated. It is a key that updates to the Data Marts are made during the update cycle of the DW to maintain consistency between them. This is also a major consideration and design point, as you move to a real-time environment. At that time, it is good to revisit the requirements for the Data Mart, to see if they are still valid

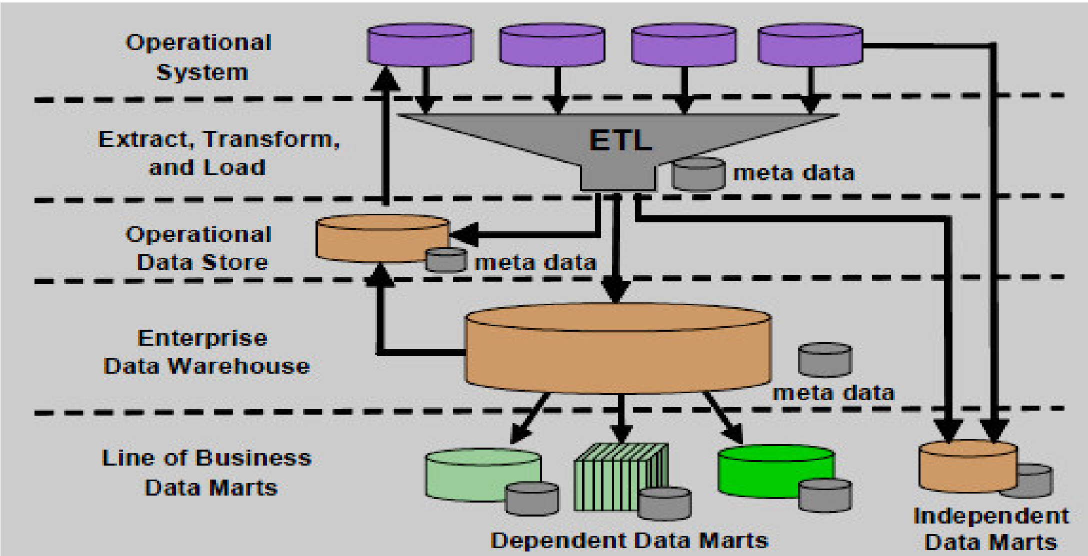


**Figure 2: DW Architecture [4]**

However, there are also many other data structures that can be part of the data warehousing environment and used for data analysis, and they use differing implementation techniques. Although Data Marts can be of great value, there are also issues of currency and consistency. This has resulted in recent initiatives designed to minimize the number of Data Marts in a company. This is referred to as Data Mart consolidation (DMC). Data Mart consolidation may sound simple at first, but there are many things to consider. A critical requirement, as with almost any project, is executive sponsorship, because you will be changing many existing systems on which people have come to rely, even though the systems may be inadequate or outmoded. To do this requires serious support from senior management. They will be able to focus on the bigger picture and bottom-line benefits, and exercise the authority that will enable making changes [4].Some of the existing literatures on DW have recognized five other architectures: Independent Data Marts, Bus Architecture, Hub and Spoke, Centralized and Federated [13]. These architectures are collected in Figure 3

.

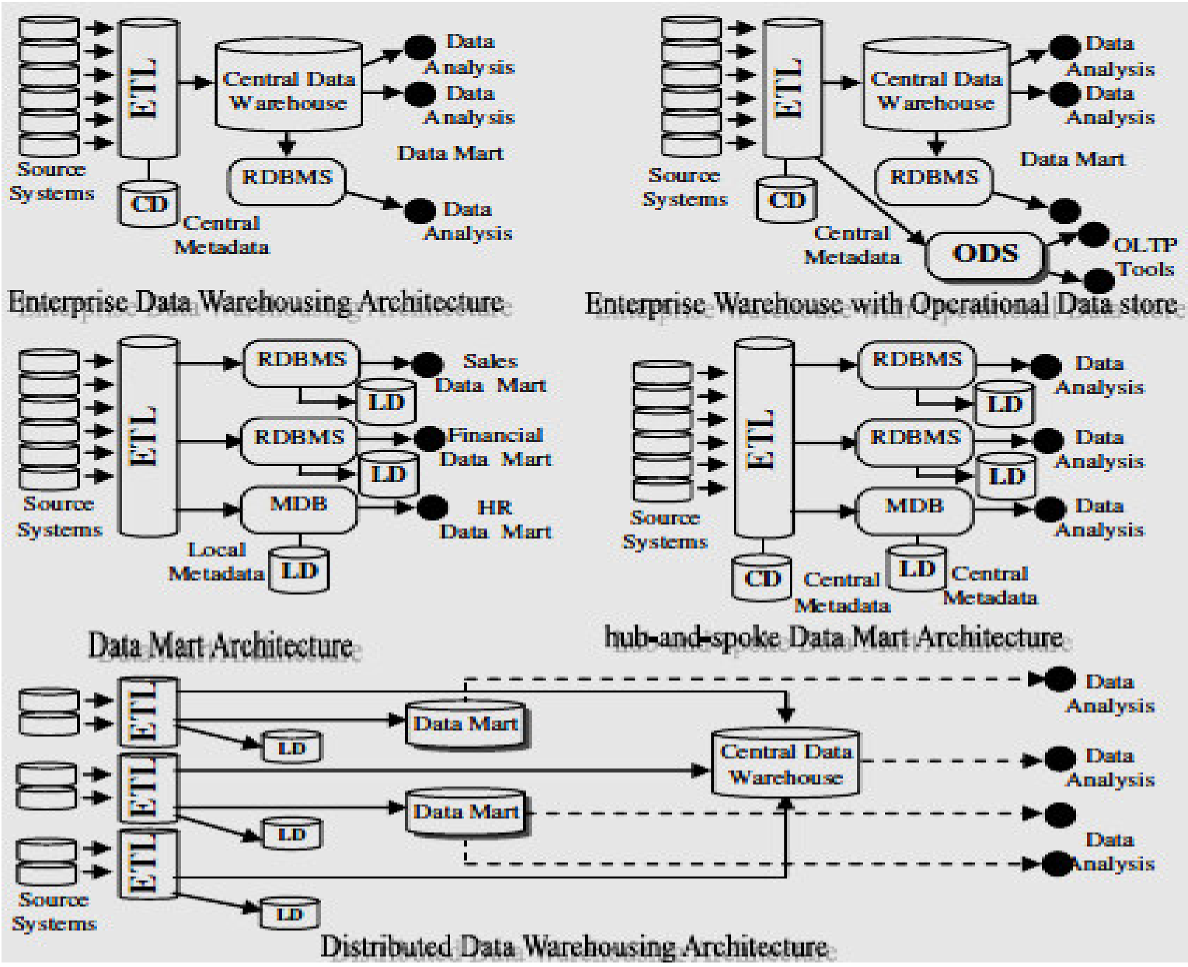


**Figure 3: The common architectures of data warehouse [15]**

Figure 3 illustrates the most important architectures of the DW, Independent Data Marts (IDM), which are the first accomplishments to provide a repository of decision support data, are usually independent of other data stores and serve specific and localized needs such as providing data for a particular application or business unit, do not provide "a single version of the truth". Data mart bus architecture with linked dimensional data marts (DBA) has data marts that support various business processes, the first mart is built for a single business process using dimensions and measures that are used with other marts (i.e., conformed dimensions), additional marts are developed using these conformed dimensions, which results in logically integrated marts and an enterprise view of the data [3]. Hub and Spoke Architecture is developed in an iterative manner, subject area by subject area. In this architecture, atomic level data is maintained in the warehouse in 3rd normal form. Dependent data marts are created that source data from the warehouse, thus maintaining a "single version of the truth" The dependent data marts may be developed for departmental, functional area, or specialized purposes (e.g., data mining) and may have normalized, summarized dimensional data structures depending on user needs. Federated is recommended when there is a fragmented decision support data environment and there is a need to integrate at least some of the data. These data are either logically or physically integrated using shared keys, global metadata, distributed queries, or other methods [13].

## III. Comparing Inmon's And Kimbal's Architectural Methodologies

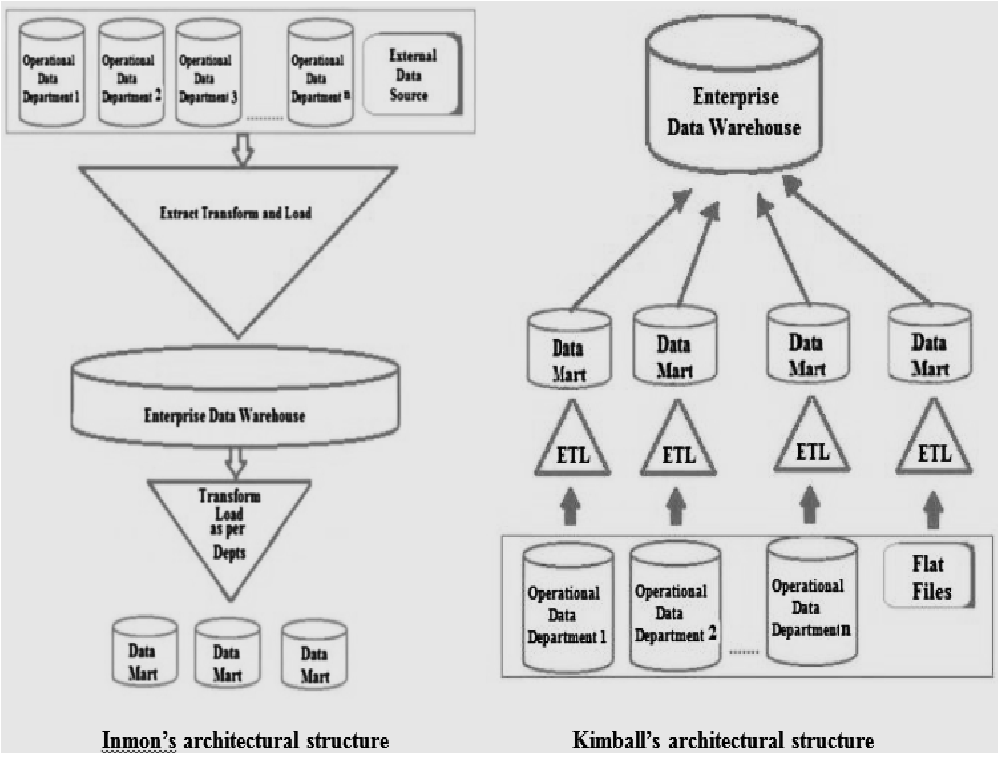


**Figure 4: Inmon's and Kimball's Architectural Methodologies [9]**

The first level contains daily transactional processing data, and the last three levels become part of a DW, which provides a logical framework for DSS and business management capabilities. Inmon believes that the initial efforts to construct atomic data warehouse later helps in the creation of any number of departmental DWs without risking data incompatibility between them. Kimball's architecture starts from individual department's data and builds data marts. Then, it uses these individual data marts to build the enterprise data warehouse. In this architecture all data marts are modeled within reliable data standards called adapted dimensions [12]. Table 1 summarizes the most essential characteristic differences between the two ideologies. According to Watson [19], not long ago, BI managers and professionals were struggling with architecture decisions. "Is the Inmon hub-and-spoke or the Kimball data mart bus architecture best? (Both can be successful) Should we build logical or physical data marts? (You will end up with at least a few physical ones)." Watson argues that these decisions do not seem challenging today; they were for the people developing their first DW at a time when data warehousing knowledge was less codified. Conradie 2005 [5] investigated and compared the views of Inmon and Kimball and found that the concept of the CIF is appealing [13]. Although he fully supported the concept of a CIF, the Kimball approach to the design of the data warehouse (simple data mart by data mart, driven by specific business needs and glued together by the Bus architecture of conformed dimensions), led him to lean towards the Kimball approach when developing the Bigger Picture BI context Model in his thesis. He claimed that the idea to accommodate the detailed transactional data requirements in a detailed data mart as part of the data warehouse (instead of a separate ODS), is a further plus point for the Kimball's approach [13].

**Table 1: Comparison of Essential Features of Inmon's and Kimbal's Ideology [16]**

| Kimball's Approach | | Inmon's Approach |
|---|---|---|
| 1. | Everyone is allowed to fabricate their database according to their requirements and department structure. All these independent repositories can be integrated as and when required. This methodology is known as bottom up approach. | Supports a top down approach. Here no one is allowed to develop any database independently. The database for an organization should be planned and designed centrally. Every department within the organization will follow the centrally designed schema to fabricate their database. |
| 2. | This structure is easier to build. | The structure proposed is very typical one to craft. |
| 3. | It is a nimble approach. | Rigorous analysis and designing is required. |
| 4. | Problematic to maintain as an enterprise resource. | Easier to maintain as an enterprise resource. |
| 5. | Data is often redundant. | Redundancy is regulated to a great extent. |
| 6. | Very difficult to integrate independent data marts with varying structure. | Integration of data marts is comparatively easier. |
| 7. | This approach is flexible. | This approach is comparatively rigid. |

## IV. Developing An Improved Data Warehouse Architecture For A University Dss

Kimball's idea of the DW starts from individual department's data and builds data marts (meaning departmental data marts are fully independent of the organization's DW), so that the integration of all the data marts forms the DW of the organization. Inmon's idea of the DW is to construct an atomic DW which will later help in the creation of any number of departmental data marts (meaning departmental data marts are fully dependent on the organization's DW) without risking data mismatch between them.

To sum it all, in Kimball's approach departmental data marts are independent of the DW while in the Inmon's approach departmental data marts are dependent of the DW. The two approaches combined together forms the basis of our argument that leads to the development improved DW architecture for a University DSS. We developed our proposed architecture by considering these approaches. We have combined their ideas to form the independent and dependent DW architecture for the SPGS MAUTECH.

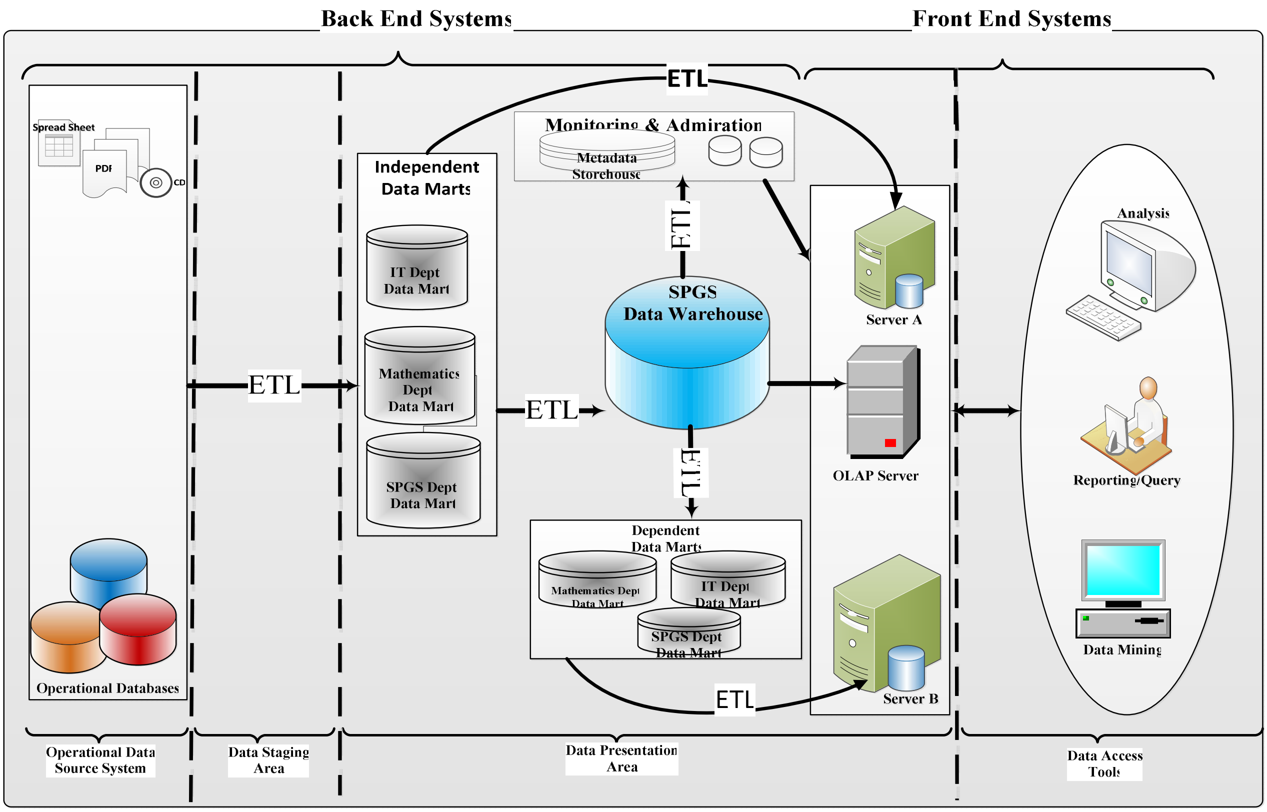


**Figure 5: The Proposed DW Architecture Design for a University DSS**

Figure 5 shows the improved architectural design for the DW Architecture and DSS. The improved architecture has the followings abilities:

i. Integration of the independent and dependent Data Marts within the architecture.
ii. It has a multiple tier ETL which are simpler and in multiple stages.
iii. It has three access points (Using a three tier server).

The proposed DW architecture was developed based on three departments (IT, MA and PG) within the SPGS MAUTECH; Data Marts were designed based on these departments and the integration of these data marts forms the Proposed SPGS DW (Kimbal's approach). Out of the proposed SPGS DW a subject based dependent departmental (IT, MA and PG) data marts were generated (Inmon's approach). The ideologies we had followed to develop this architecture makes it scalable and as such, it can be adopted for any institution of learning in Nigeria.

The proposed architecture has a back and frond end system in which so many activities is carried out. The back end systems comprises of the operational data source system, data staging area and the data presentation area. Data are first extracted from different operational data source systems and then stored briefly into an ETL server at the data staging area where it is being processed as soon as it is captured. The activities of the ETL server at the data staging includes data scouring and cleansing, data integration, data fixing, data entry errors, transforming and refreshing data into a new normalized standard. As soon as data is cleansed, the transformed data are loaded and indexed into the data presentation area where the data marts (independent and dependent) and DW are located. In this process, tables are released, new tables are created, columns are castoff, and new ones are created based on the user requirements.

The front end systems in the other hand comprise of the main servers (OLAP, server A and B) and data access tools. The OLAP server hold the copy of the data contained in DW while the servers A and B holds data about the independent and the dependent Data Marts respectively. The data access tool is the interface where applications are stored, which allows for data Analysis, Reporting/Querying and Data mining activities.

## V. Conclusion

Data warehousing and functional data sources share certain similarities and there are significant synergies in development and such opportunities should be exploited [20]. Research in these area points that there are methodological reuses potential for project justification and systems development and any expertise in this area should be exploited by organizations for optimization of resources [20]. The greatest potential benefits of data warehousing are when the DW is used in the redesign of business processes and to support strategic business objectives [19]. Improved decision making usually results from the better information available from a data warehouse.

In this paper we have focused on developing an improved data warehousing architecture for a University decision making using SPGS MAUTECH, Yola as our case study. We have considered the ideologies of data warehousing in the course of this study and demonstrated how data can be incorporated from diverse heterogeneous source systems into a sole historical repository (DW) that is capable of conveying a DSS to the University's administrators and other end users. The case study used in this study stresses that there is value in addressing the data needs of the institution in a holistic way. The proposed DW architecture would combined the departmental data marts to reduce time for reconcilements and report development for end users; reduce risk from manual exposure calculations due to service availability and expand user access to data. It will help for gradually moving towards more analytical processing at a faster pace. On this context, University data management should step up towards implementing this project.

---

# References


[1] Alejandro, G. and Adriana, M. (2000). An Overview of Data Warehouse Design Approaches and Techniques. Instituto de Computación, Facultad de Ingenieria, Universidad de la República, Montevideo, Uruguay.

[2] Amit P. and Sushma M. (2014) Moving from traditional data warehouse to enterprise Data management: a case study *Journal of issues in information systems Volume 15, issue ii, pp. 133-140, 2014.*

[3] Ariyachandra, T. and Watson, H. J. (2010). Key organizational factors in data warehouse architecture selection. Decision Support Systems 49 200–212.

[4] Chuck, B. Daniel, M. F. Amit, G. Carlos, M. & Stanislav, V. (2006). Dimensional Modeling: In a DSS Environment. IBM Corporation, North Castle Drive Armonk, NY 10504-1785 U.S.A.

[5] Conradie, P. J. (2005). An industrial engineering perspective of business intelligence. Doctorate thesis, Faculty of Engineering, Built Environment and Information Technology, University of Pretoria.

[6] GRT Corporation. Data Warehousing and Business Intelligence Development Approach. Retrieved: February 14, 2015 from http://www.grtcorp.com/solutions/data_warehouse_business_intelligence.

[7] Inmon, W.H. [Ed.] (2002). *Building the Data Warehouse.* New York. John Wiley & Sons, Inc.

[8] Inmon, W. H. (2006). Data Warehouse 2.0 Architecture for the Next Generation of Data Warehousing. DM Review, Vol. 16 Issue 4, p.8-25.

[9] Intikhab, K. (2012). Modern Methodology and Tools for Data Warehouse Development. Master's thesis, Department of Information Systems, Athabasca University, Alberta.

[10] Jerry, Y. D. and Mohammad, S. (2012). Three Tier level Data Warehouse Architecture for Ghanaian Petroleum Industry. *International Journal of Database Management Systems (IJDMS)* Vol.4, No.5.

[11] Kimball, R. & Ross, M. (2002). *The DW Toolkit: The Complete Guide to Dimensional Modeling*.New York. John Wiley & Sons, Inc.

[12] Mary, B. (2004). *Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models*. *Journal* of *Business Intelligence*.9, 1, 6–20.

[13] Moh'd, A., Kamal M., and Mieczysław L. O. (2012). A survey of Data Warehouse architectures: preliminary results. Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1121–1126.

[14] Pant, S. and Hsu, C. (1995). *Strategic Information Systems Planning*: A Review Information Resources Management Association International Conference, Atlanta.

[15] Sen, A. and Sinha, A. P. (2005). A Comparison of Data Warehousing Methodologies Using a common set of attributes to determine which methodology to use in a particular data warehousing project. Communications of the ACM. Vol. 48, No. 3 pp. 79-84.

[16] Singh J., (2010). Promulgation of Contemporary Testing Techniques For Effective Engineering of Data Warehouses, Doctorate thesis, faculty of engineering and technology Punjabi University, India.

[17] Surajit, C. and Umeshwar, D. (1997). *An Overview of Data Warehousing and OLAP Technology. Journal of SIGMOD Record*. 26, 65–74.

[18] Thilini, A. and Hugh, W. (2010). Key organizational factors in DW architecture selection. *Journal of Decision Support Systems*. 49, 2, 200–212.

[19] Watson, H. J. (2012). This Isn't Your Mother's BI Architecture". *Business Intelligence Journal* Vol. 17 Issue 1, p4-6.

[20] Winter, R. (2001). The Current and Future Role of Data Warehousing in Corporate Application Architecture. Proceedings of the 34th Hawaii International Conference on System Sciences, Hawaii, USA.