A Superficial Exposé of Data Warehousing: An Intrinsic Component of Modern Day Business Intelligence

Oyerinde, O.D[1], Adekunle, A. Y[2], Ebiesuwa, O.O[3]

[1]CIS Department, ICT Directorate, University of Jos. PMB 2084, Bauchi Road, Jos, Nigeria
oyerindeo@unijos.edu.ng

[2]School of Computer Science and Engineering, Babcock University, Ilishan-Remo, Ogun State, Nigeria
phdlectures@gmail.com

[3]School of Post Graduate Studies, Babcock University, Ilishan-Remo, Ogun State, Nigeria
seunebiesuwa@yahoo.com

Abstract: ***This paper aims to give a superficial exposé of Data Warehousing technology as a possible effective tool for organizations Business Intelligence. The key components of a Data Warehouse will be discussed as they offer a part of the core requirements for successful Business Intelligence deployment in an organization. Universally accepted Data Warehousing and Business Intelligence Models will be discussed and highlighted in order to ascertain the effectiveness of Data Warehousing as a tool for efficient Business Intelligence deployment. Traditionally, data warehouses are designed to collect and organize historical business data so it can be properly analyzed to enable management make optimal business decisions. Effective Business Intelligence can help companies gain a comprehensive understanding of the factors affecting their business, enabling them to make informed decisions for the competitive edge (Gutierrez, 2007)***

Keywords: Data Warehousing, Business Intelligence, Operational Data Store, Multi Dimensional Data Model**.**

1. Introduction

Different people have different definitions for a data warehouse. The most popular definition however is "A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process"[2]. Another concise definition is " A data warehouse is a copy of transaction data specifically structured for query and analysis"[3]. This is a more functional view of a data warehouse. Ralph Kimball did not address how the data warehouse is built like Bill Inmon did; rather he focused on the functionality of a data warehouse.

During the mid to late 1990s, data warehousing became one of the most important developments in the information systems field. It is estimated that 95% of the Fortune 1000 companies either have a data warehouse in place or are planning to develop one [4].

A data warehouse (or smaller-scale data mart) is a specially prepared repository of data created to support decision making [5]. Data are extracted from source systems, cleaned, transformed, and placed in data stores [6]. A data warehouse has data suppliers who are responsible for delivering data to the ultimate end users of the warehouse, such as analysts, operational personnel, and managers.

Data warehousing came about as a product of business need and technological advances. The business environment has become more complex, competitive, global, and volatile. Customer relationship management and e-commerce initiatives are currently creating requirements for large, integrated data repositories with advanced analytical capabilities. More data are captured by organizational systems (e.g., barcode scanning, clickstream) or can be purchased from companies like Dun &Bradstreet. Through hardware advances such as symmetric multi-processing, massive parallel processing, and parallel database technology, it is now possible to load, maintain, and access databases of terabyte size [7]. All of these changes are affecting how organizations conduct business, especially in sales and marketing, allowing companies to analyze the behavior of individual customers rather than demographic groups or product classes.

2. Review of Literature

Different data warehousing systems have different structures. Some may have an Operational Data Store (ODS), while some may just have multiple data marts. Some may have a small number of data sources, while others may have dozens. In view of this, it has become far more reasonable to present the different layers of data warehouse architecture rather than to discuss the specifics of any one system.

In general, all data warehouse systems have the following layers:

- Data Source Layer

- Data Extraction Layer
- Staging Area
- Extraction Transformation and Loading (ETL) Layer
- Data Storage Layer
- Data Logic Layer
- Data Presentation Layer
- Metadata Layer
- System Operations Layer

Figure 1 shows the relationships among the different components of the data warehouse architecture:

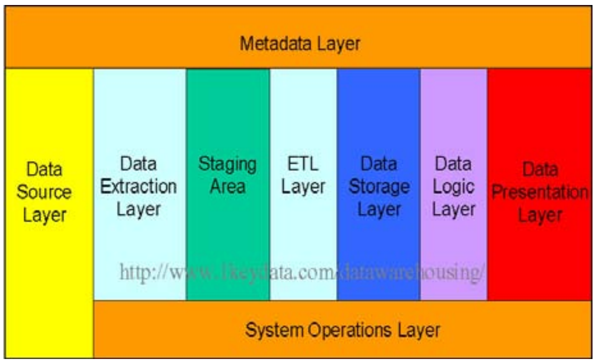

Figure 1: Components of Data Warehouse Architecture [8]

Data Warehouses are meant to support managers with answers to important business questions which require complex analytics such as roll-ups, drill-downs, pivoting, aggregations and data slicing and dicing [9]. This becomes essential because all levels of management decision-making processes are supported by incorporating Data Warehousing technology through the collection, transformation, integration, and interpretation of both the internal and the external data.

A Data Warehouse is the repository of summarized data [10], whereas Data Warehousing revolves around the development, management, methods, and practices that define how these summarized data are acquired, integrated, interpreted, managed, and used within business organizations [11].

There is, however, a significant differentiation between an Operational Data Store (ODS) and a Data Warehouse. Although an Operational Data Store uses Data Warehouse technology (i.e. star schema) to provide an integrated view of data, it is intended to assist day-to-day operations and not decision making [12]. This is so because an Operational Data Store is a subject oriented, integrated, and volatile (updatable) data store that contains only business organization detailed data for operational usage [13].

Initially, Data warehousing was viewed as a means by which business organization could solve the problems usually associated to their independently legacy systems which more often than not contained inaccurate, duplicate, and dissimilar data about the same entity. Data Warehouse technology, also known as star schema, can help managers make more effective decisions [14] by providing them with suitable information which is fundamentally different from the type of information that businesses use in their day-to-day operations [15].

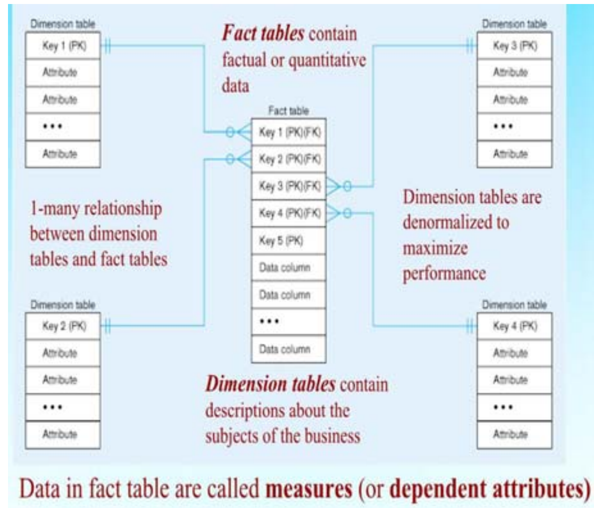

Figure 2: Example of Star Schema [16]

A Data Warehouse allows a business organization to manipulate a great deal of data in ways that are useful to it. This may be by: describing, organizing, cleansing, summarizing and storing large volumes of data to be transformed analyzed and reported [17].

Summer and Ali [18] argued that A Data Warehouse is the way in which a business organization converts its data into information which can be represented into different ways (textual, graphically, etc.) based on its reporting capabilities [19]. They also postulated that an ad hoc system which is provided to managers based on Data Warehousing allows them to generate speculative information such as projections, as well as allowing them to explore "what-if" analysis [20]. Thus, the desire to improve decision-making and business performance has been the fundamental business driver behind data warehousing [21].

Furthermore, Data Warehousing can be seen as a form of analytical processing and its key role is to offer correct and compulsive business intelligence to business organizations' decision-makers, through enriching their abilities in understanding business problems, exploiting opportunities and improving business performance [22]. Hence, by leveraging on Data Warehousing technology for business intelligence initiatives, business organizations will gain strategic competitive advantage [23].

Data warehouses integrate data from various transactional, legacy, or external systems, applications, and sources [24]. The data warehouse provides an environment separate from the operational systems and is completely designed for decision-support, analytical-reporting, ad-hoc queries, and data mining. This isolation and optimization enables queries to be performed without any impact on the systems that support the business' primary transactions (i.e transactional and operational systems).

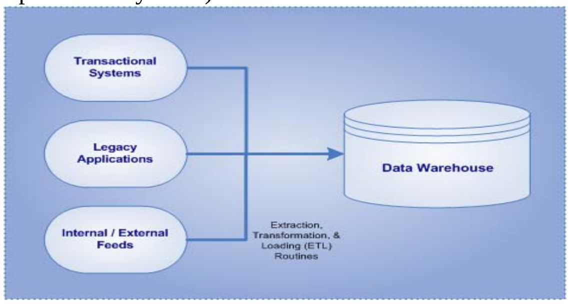

Figure 3: Data Warehouse [25]

3. Methodology

We realize that the scope of Business Intelligence more often than not should include making the best use of information for strategic, tactical, and operational needs. The basic purpose in building Business Intelligence strategy is to help businesses with long-term planning, help middle management with tactical reporting, and help operations with day-to-day decision making to run the business efficiently [26]. Business Intelligence is all about providing people with the information they need to do their jobs more effectively.

A wide range of Business Intelligence services need to be provided to meet a wide range of requirements. The Scope of a Business Intelligence Strategy should be determined by the business drivers and business goals. Scope should always account for the changing business requirements to keep the Business Intelligence strategy aligned with business.

Below is a functional model of what a Business Intelligence strategy entails and what it offers an organisation.

Figure 4: Business Intelligence [27]

For effective Data Architecture, an effective Data Warehousing Design and Implementation can adequately suffice as an efficient means of Data Provision for enhancement of the organization's Business Intelligence requirements.

Arnott and Pervan [28] argue that data warehousing provides the large scale IT infrastructure for contemporary decision support and business intelligence. They argue that the main reasons behind that is the use of Multi-Dimensional Data Model "MDDM" or cubes [29], which organizes large data sets in ways that are meaningful to managers besides being relatively easy to query and analyze. It has been proven that MDDM is the most suitable for On-Line Analytical Processing "OLAP" applications, data mining and advanced reporting functions [30].

The conceptualized MDDM can be physically realized in two ways: first, by using trusted relational databases "star schema/snowflake schema" or, by using specialized multidimensional databases. Many computer assisted analytical processes such as data mining and OLAP [31], are used to analyze data from different angles and distill it into actionable information are run over Data Warehouses [32].

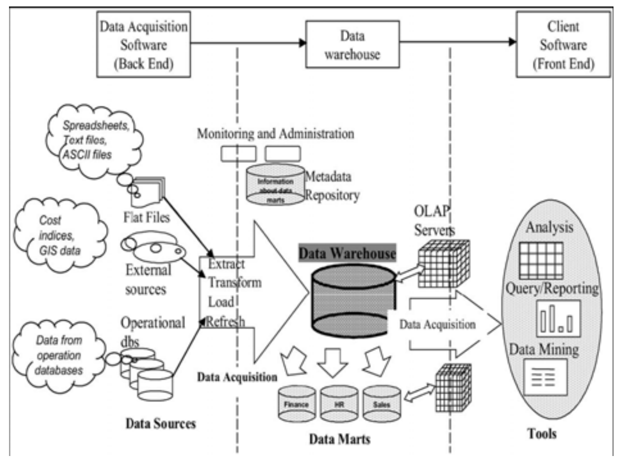

Figure 5: Data Warehouse Components Model [33]

Business Intelligence architecture varies in each enterprise but there are some common components of Business Intelligence architecture, which are found in some shape or form in all Business Intelligence solutions. What is included in the Business Intelligence architecture will be driven by the objectives, goals and requirements of the enterprise [34].

The following diagram depicts a comprehensive representation of the typical Business Intelligence architecture. Multiple disparate Data sources, data integration services, data management services, reporting, analytical services, information delivery, and consumption services form the broad spectrum of the Business Intelligence architecture.

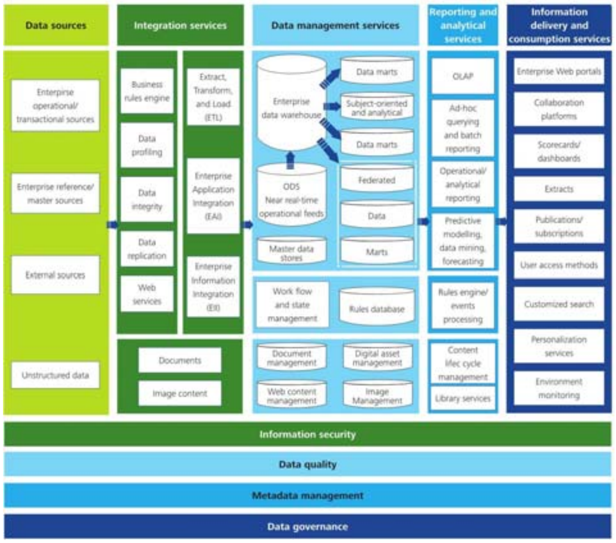

Figure 6: Business Intelligence Components Model [35]

A superficial comparison of both models reveals that the components of effective Data Warehousing are implemented in Business Intelligence. We can therefore theoretically infer that an effective utilization of Data Warehousing Technology can be used as a tool for data architecture and requirements for successful implementation of Business Intelligence.

4. Discussions and Implications

There has long been a keen interest in identifying the factors that contribute to the success or failure of data warehouses. In an early review of the literature, Vatanasombut and Gray [36] identified 51 success factors that may be classified into 12 categories. Most of these 51 factors, however, apply not only to data warehousing, but also to large systems development projects in general; only nine factors are specific to data warehousing. Other researchers provided their own lists of critical success factors. For example, Watson and Haley [37] identified eight critical success factors, whereas Sammon and Finnegan [38] discussed their “ten commandments of data warehousing.”

Even though there are many success stories [39], a data warehousing project is an expensive, risky undertaking. The typical project costs over $1 million in the first year alone [40]. While hard figures are not available, it is estimated that one-half to two-thirds of all initial data warehousing efforts fail [41]. The most common reasons for failure include weak sponsorship and management support, insufficient funding, inadequate user involvement, and organizational politics [42].

Practitioners and researchers need to better understand data warehousing to ensure the success of these promising, yet risky and costly, IT undertakings. The IT literature contains many studies that investigate the factors that affect the implementation of decision-support applications. While these studies are helpful, a data warehouse is arguably different in that it is an IT infrastructure project, which can be defined as a set of shared, tangible IT resources that provide a foundation to enable present and future business applications [43]. The capability of such an infrastructure is thought to impact business value by supporting (or failing to support) important business processes [44].

The table below summarizes the five published survey studies, which differ widely in the variables measured. Some studies measured critical success factors while others measured data warehousing success; however, only one Wixom and Watson [45] measured both critical success factors and data warehousing success. Without including both in the same study, the effect of any success factor on data warehousing success cannot be substantiated.

Table 1: Data Warehouse Success Study Survey [45]

Study	DW Success Factors Measured	DW Success Measured	Results Reported	N
Watson & Haley (1997)	Upper management support; User involvement; Having a business need: User support, Using a methodology, modeling; Defined, understandable goals; Good, clean data; Managing expectations	Not Available (N/A)	Ordered list of success factors	121
Chen et al. (2000)	N/A	Support for end users; Accuracy, format, and Preciseness; Fulfillment of end users needs; User satisfaction	Support for end users affects user satisfaction	42
Wixom & Watson (2001)	Management support; Champion; Resources; User participation; Team skills; Source systems; Development technology; Organizational implementation success; Project implementation success; Technical implementation success	System quality; Data quality; Net benefits	Some success factors affect DW success	111
Watson et al. (2001)	N/A	Reduced effort by developers to produce info; Improved user ability to produce info; More and better info; Better decisions; Improvement for business process; Support for the accomplishment of strategic business objectives	Ordered list of success measures	106
Hwang & Cappel (2002)	N/A	N/A	Development/management Practices	27
Shin (2003)	N/A	System quality; Information quality; Service quality; User satisfaction	System quality affects user satisfaction	64

Researchers have also defined and measured different success factors and data warehousing success variables. For example, user satisfaction was used as a measure for success in two studies (Chen et al.. 2000; Shin. 2003), but not in the others (Watson et al.. 2001; Wixom and Watson, 2001). The two studies conducted by Watson and colleagues used different success measures too. It appears that even the fundamental question of what constitutes data warehousing success has not been resolved. Few studies have examined the implementation success of infrastructure projects instead, infrastructure research focuses on the innovation and diffusion of such phenomenon.

An implementation is not successful unless the system it produces is accepted into the organization and integrated into work processes. However, an information system implementation can cause considerable organizational change that people tend to resist [46]. The likelihood of this resistance increases with the scope and magnitude of the changes that the system creates [47].

Data warehousing, in particular, has profound effects on organizations because it can shift data ownership, use, and access patterns; change how jobs are performed; and modify business processes [48]. It moves data ownership from the functional areas to a centralized group, shifts the responsibilities for data access from information systems personnel to end users, changes how users perform their jobs as a result of having access to warehouse data, and allows businesses to operate differently. These changes potentially lead to resistance from managers, data suppliers, and end users.

5. Conclusions

Insights will be gained through improved information access. Managers and executives will be freed from making their decisions based on limited data and their own initiative. Decisions that affect the strategy and operations of organizations will be based upon credible facts and will be backed up with evidence and actual organizational data.

Moreover, decision makers will be better informed as they will be able to query actual data and will retrieve information based upon their unique personal needs. In addition, data warehouses and its related business intelligence can be applied directly to intrinsic business processes such as marketing segmentation, financial management, inventory management, and sales.

For many organizations, enterprise information systems are comprised of a number of multiple subsystems which are physically separated and built on different platforms. Furthermore, merging of data from multiple independent and disparate data sources is a common need when conducting business intelligence. In order to solve this problem, the data warehouse performs an integration of existing disparate data sources and makes them readily accessible in one place.

Therefore, business users will spend little time in the data retrieval process. Scheduled data integration routines, known as Extraction Transformation and Loading, are leveraged within a data warehouse environment. These routines help consolidate data from multiple source systems and transform the data into a useful, readily accessible and understandable format. Subsequently, business users can then easily access data from one customized interface.

Businesses and profit oriented organisations expect high returns on investment. Return on Investment (ROI), refers to the amount of increased revenue or decreased expenses a business will be able to realize from any project or investment of capital. Subsequently, successful implementations of data warehouses and other complementary business intelligence systems have enabled businesses to generate higher amounts of revenue and provide substantial savings in capital expenditure. According to a 2002 International Data Corporation (IDC) study “The Financial Impact of Business Analytics”, analytics projects have been achieving a substantial impact on a business’ financial status. Importantly also, the study found that business analytics implementations have generated a median five-year return on investment of 112% with a mean payback of 1.6 years. Of the businesses included in the study, 54% have had a return on investment of 101% or more [48].

A Data Warehouse is highly recognized as an infrastructure and allows many applications run over it such as Decision Support Systems. Many techniques, such as data mining, OLAP and dashboards have been rising to prominence to extract business intelligence from Data Warehouses [49]. Therefore, Data Warehouses are meant to be used by managers since they support decision-making process.

6. Suggestions for Further Research

Data warehousing success is an important issue for both researchers and practitioners however, not many studies have empirically assessed data warehousing practices in general and critical success factors in particular [50]. Although plenty of guidelines for implementation exist, few have been subjected to rigorous empirical testing. Another problem is that researchers have used different variables in individual studies, thus making comparison and integration of the results from different studies difficult.

Data Warehouses have experienced relatively high failure rates and its spread and/or use has been to some extent limited [51]. This may be due to the fact that designing and developing a Data Warehouse is a risky, costly and complex process. It requires huge amounts of capital investments, it is time consuming and spans years, and it needs a wide variety of technical and managerial skills. Generally speaking, social aspects are shaping the technology. Hence, the interaction of technology and social context is a key determinant of a Data Warehouse success [52]. Nevertheless, despite the technical complexity of a Data Warehouse design and implementation, social/cultural and organizational factors are the most cited reasons behind Data Warehouse failures [53].

In another data warehousing research, Doherty and Doig [54] concluded in a case study that the success of a data warehouse implementation depends on how well the resulting culture changes are measured and managed. The potential impact of cultural changes, as well as other political, social, and economic factors [55] should be further researched to allow a fuller understanding of data warehousing success.

References

[1] Gutierrez, N.A. "Effective Visualization Technique for Business Intelligence". *INFOSYS: Building Tomorrow's Enterprise.* (2007).

[2] Inmon, W.H., "What is a Data Warehouse?" *Prism, Volume 1, Number 1*, 1995.

[3],[29] Kimball, R., "The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses." 1996.

[4] META Group. "Industry Overview: New Insights in Data Warehousing Solutions," *Information Week*, 1996, pp. 1-27HP.

[5],[7],[45],[48] Wixom, B. H. and Watson, H. J. (2001) ‘An Empirical Investigation of the Factors affecting Data Warehousing Success’. *MIS Quarterly*, 25 (1), pp. 17-41.

[6],[21] Gray, P., and Watson, H. J. *Decision Support in the Data Warehouse*, Prentice Hall, Upper
Saddle River, 1998.

[8] http://www.1keydata.com/datawarehousing.html

[9],[23] Ramamurthy, K. R., Sen, A. and Sinha, A. P. (2008) ‘An empirical investigation of the key determinants of data warehouse adoption’. *Decision Support Systems*, 44 (4), pp. 817-841.

[10] Ang, J. and Teo, T. S. H. (2000) ‘Management issues in data warehousing: insights from the
Housing and Development Board’. Decision Support Systems, 29 (1), pp. 11-20.

[11],[22] March, S., Hevner, A. (2007) 'Integrated Decision Support Systems: A data warehousing perspective'. *Decision Support Systems*, 43 (3), pp. 1031-1043.
[12] Shin, B. (2002) 'A Case of Data Warehousing Project Management'. *Information &Management*, 39 (7), pp. 581-592.
[13] Boar, Bernard (Bernie)." Understanding Data Warehousing Strategically." *White paper commissioned by NCR's Communication Industry Line of Business*. The Data Warehousing Institute, Gaithersburg, MD, 1996.
[14],[17] Griffin, R. K. (1998) 'Data Warehousing'. *Cornell Hospitality Quarterly*, 39 (4), pp. 28-35.
[15],[18],[19] Summer, E. and Ali, D. L. (1996) 'A Practical Guide for Implementing Data Warehousing'. *Computers and Industrial Engineering*, 31 (1), pp. 307-310.
[16] Hoffer, J.A., Prescott, M.B., McFadden, F.R. "Chapter 11: Data Warehousing, Modern Database Management, 7th Edition," *Prentice Hall*, 2005.
[20],[31],[33],[49],[52],[53] Al-Debei, M. M. "Data Warehouse as a Backbone for Business Intelligence: Issues and Challenges" *European Journal of Economics, Finance and Administrative Sciences* 33 (2011): pp 153-166.
[24],[25],[48] http://bi-insider.com/portfolio/benefits-of-a-data-warehouse/
[26],[27],[34],[35] Pant, P. "Business Intelligence (BI) How to Build Successful BI Strategy" *Deloitte Consulting LLP* (2009).
[28] Arnott, D. and Pervan, G. (2005) 'A Critical Analysis of Decision Support Systems Research'.
Journal of Information Technology, 20 (2), pp. 67-87.
[30] Bellatreche, L., Karlapalem, K., Mohania, M. "Some issues in design of data warehousing systems", *Data warehousing and web engineering*, 2002, pp. 22-76
[32] Gunnarsson, C. L., Walker, M. M., Walatka, V. and Swann, K. (2007) 'Papers Lessons Learned: A case study using data mining in the newspaper industry'. *Database Marketing & Customer Strategy Management*, 14 (4), pp. 271-280.
[36] Vatanasombut, B., and Gray, P. "Factors for Success in Data Warehousing: What the Literature Tells Us," *Journal of Data Warehousing* (4:3), 1999, pp. 25-33.
[37],[40] Watson, H. J., Haley, B. (1997) 'Data warehousing: A framework and survey of current practices'. *Journal of Data Warehousing*, 2(1), pp. 10-17.
[38] Sammon, D. and Finnegan, P. (2000) 'The Ten Commandments of Data Warehousing'.*Database*, 31 (4):82-91.
[39] Beitler, S. S., and Leary, R. "Sears' EPIC Transformation: Converting from Mainframe Legacy Systems to On-Line Analytical Processing (OLAP),"*Journal of Data Warehousing* (2:2), 1997, pp. 5-16.
[41] Kelly, S. *Data Warehousing in Action*, John Wiley & Sons, Chichester, 1997.
[42] Watson, H. J., Gerard, J. G., Gonzalez, L. E., Haywood, M. E., and Fenton, D. "Data Warehousing Failures: Case Studies and Findings," *Journal of Data Warehousing* (4:1), 1999, pp. 44-55.
[43],[45],[50] Xu, Hongjiang and Hwang, Mark I., "The Effect of Implementation Factors on Data Warehousing Success: An Exploratory Study" (2007). *Scholarship and Professional Work - Business.* Paper 77. http://digitalcommons.butler.edu/cob_papers/77
[44],Xu, Hongjiang and Hwang, Mark I., "A Structural Model of Data Warehousing Success" (2008). *Scholarship and Professional Work Business.* Paper 78. http://digitalcommons.butler.edu/cob_papers/78
[46] Markus, M.L. "Power, Politics and MIS Implementation," *Communications of the ACM*, 26, 1983, pp. 430-444.
[47] Tait, P and Vessey, I. (1988) 'The Effect of User Involvement on System Success: A Contingency Approach'. *MIS Quarterly*, 12 (1), pp. 91-108
[51] Quaddus, M. and Intrapairot, A. (2001). "Management policies and the diffusion of data warehouse: a case study using system dynamics-based decision support system." *Decision Support Systems.* 31 (2): pp. 223-240. http://espace.library.curtin.edu.au/R?func=dbin-jump-full&local_base=gen01-era02&object_id=116573
[54] Doherty, N.F.; Doig, G.,(2003) "An analysis of the anticipated cultural impacts of the implementation of data warehouses," *Engineering Management, IEEE Transactions on* , 50 (1), pp.78,88
[55] Bussen, W., Myers, M., (1997) "Executive Information Systems Failure: A New Zealand Case Study", *Journal of Information Technology*, 12 (2), 145-153.

Author Profile

Oyerinde, Oyeyinka Dantala received the B.Sc. degree in Computer Science (Information Systems) from Babcock University, Ilishan-Remo, Ogun State, Nigeria in 2005. He is currently pursuing his M.Sc. degree in Computer Science (Management Information Systems) from the same Babcock University. Mr. Oyerinde is a staff of the Corporate Information Systems Department of the ICT Directorate of University of Jos, Jos Nigeria. Having been employed in 2006, he has been on various committees and development teams and was the Acting Head of CIS from December 2010 till February 2013.