Global Scientific JOURNALS

GSJ: Volume 8, Issue 2, February 2020, Online: ISSN 2320-9186
www.globalscientificjournal.com

DATA WAREHOUSE IMPLEMENTATION FRAMEWORK FOR RETAIL BUSINESSES

Abubakar Sadiq Bunu, Abdulwahab Shehu

Lead Author
Abubakar Sadiq Bunu, Information Techology Dept. MAUTECH, Yola,Nigeria, +2348032475012,
E-mail: bunu2y@yahoo.com
Co Author
Abdulwahab Shehu, Information Techology Dept. MAUTECH, Yola,Nigeria, +2348036989864

KeyWords:
Business Intelligence, Data Warehouse, Super Market, Modelling.

ABSTRACT

Data warehouse is central to Business Intelligence (BI) applications. It integrates different data sources usually into a unified data source. The ability to utilize existing information is critical for the success of decision making. Transaction Processing Systems address only the operational needs of an organization, they are not suitable for decision support and other business questions the managers' need to address. The objective of the paper was to propose a framework that enables retail businesses to implement data warehouse. Thus, this was achieved by using SQL Server enterprise edition. The output showed that the business can generate reports on product sales and supplier performance though the data comes from different data sources.

I. Introduction

Information is an asset that provides benefit and competitive advantage to any organization. Almost every organization has a relational Database Management Systems (DBMS) that is used for organization's daily operations. Companies need to improve the value of their collected data by turning it into actionable information.

The primary purpose for the existing databases and other business applications is to serve operational needs of the business. The users have both read and write access, also the system is updated at real time and is event-driven that is the process generates data. Many computer applications are required to support various retail business processes such as order processing, inventory maintenance, keeping accounting books, receiving payments etc. The effectiveness of these applications depends on what they are designed to do, they produce different reports that help to run and monitor the business.

A data warehouse (DW) refers to database that is different from the organization's Online Transaction Processing (OLTP) database and that is used for the analysis of consolidated historical data Jindal and Taneja [1]. DW provides information from historical perspective, every key structure in DW contains an element of time either implicitly or explicitly. A DW generally stores data that is 5-10 years old, which is to be used for comparisons, trends and forecasting. Data in the warehouse are not updated or changed, so it does not require transaction processing, recovery and concurrency control mechanisms [2].

Using accumulated historical data that resides within different business applications, the research work designed a platform that assists management to make better decision about their retail business. Business users realized the challenge of generating reports from computer applications that interact with different data sources. The need to harmonize data and have consistent reports prompts the development of DW in retail business.

A super market is a type of retail business that sales different products such as packed food stuffs, drinks, dairy products, decorative items and cosmetics. It also stores costly items such as electronic goods, wrist watches, jewelries and other household goods. Managing these products in terms of sales records and sufficient stocks requires an effective information system.

A well designed DW has decision support tools that can help retailers to answer questions that were previously

unanswerable. Super market managers can use their transaction database to plan sales promotions, streamline budgeting and improve vendor tracking. Successful implementation of DW in supermarket business has a potential to decrease inventories, increase sales and productivity.

For the super market DW, customer loyalty programs mean linking customer information to transaction data to collect knowledge about purchase history and preferences and utilizing that knowledge in the organization to make integrated customer centric business decision [3].This research work is required in order to assist the management of a super market in making informed decisions based on the available historical data within the organization.

Problem Defintion

The data in an organizations' database need to be structured so that all related data elements are linked together. Also the database should contain data from all or most of the business's operational applications and this data must to be consistent.Unfortunately, sales database of many organizations that are in retail sale business contains information that is very detailed while decision making requires a framework that consolidates data from different heterogeneous sources. Also different sources typically use inconsistent data representations, codes and formats which have to be reconciled [3]. As the amount of the organizational data increases, it becomes more difficult to access and get information out of it simply because it exists in different formats, different platforms and resides on different structures [4].

However, the study aims at providing an approach for integrating operational databases of retail business specifically super market into useful and reliable information to support the decision making process and also provides the basis for data analysis such as data mining and business intelligence.

The most common super market information system where different business applications (software and data sources) handle different business operations makes it impossible for management to generate reports that cut across the whole business applications. Hence the need to design an additional system to complements the existing one so that the management of the super market can access an integrated and consolidated data in one single location.

Moreover, the new framework will be based on dimensional modeling of DW which is a technique that presents the data model in an accessible standard framework, the approach keeps the data that is numeric such as counts, values, balance, occurrences, and weights in the context of data warehouse that rearrange and summarize the data and presents the data views to support business analysis and helps managers who have little knowledge of computer to be able to make complex queries and generate reports.

II. Related Literatures

The concept of data warehousing arose in mid 1980s with the aim to support huge information analysis and management reporting [5]. Inmon [6] said that data warehouse is a data collection oriented to a subject, integrated, changeable in time and not volatile, to provide support to the decision making. A data warehouse is storage of convenient, consistent complete and consolidated data, which is collected for the purpose of making quick analysis for the end users who take part in Decision Support System. Data warehouse is the center of the architecture for information systems for the 1990's [7]. Data warehouse supports information processing by providing a platform of integrated and historical data from which to do analysis [6].

According to Macura [2] data warehousing process involves extraction of data from heterogeneous data sources, filtering and transforming data sources, cleaning, filtering and transforming data into common structure and storing data in a structure that is easily accessed and used for reporting and analysis purpose.

Sheta and Eldeen [5] Argued that the transactional databases are designed to answer questions such as *who* and *what*, they are not very good in answering what if why and what next type questions. The reason is that transactional databases are not organized to support analytical processing.

A comparative study between Data warehouse and traditional database was conducted by Jindal and Taneja [1] where they conclude that “A databases is a collection of operational data, stored and used by application systems from a specific organization” while data warehouse contained data that is derived from operational data in order to support decision making, “these derived data are often called analytical, informational or managerial. The table below presents the different characteristics of a data warehouse and traditional database.

Concept of Data Integration

Macura (2014) Said that gathered data often comes from heterogeneous (different) sources; therefore, integration activities are needed. Data integration is a “central problem in the design of Data Warehouses and Decision Support Systems”. For a data warehouse to provide a reconciled and integrated outlook of organizations’ data, potential redundancies and inconsistencies need to be resolved as the data passes to data warehouse from the operational environment [8].

According to Farhan, et al. [9] Extract Transform Load (ETL) can be an ideal solution for the bulk movement of large volumes of data. Packaged ETL products also offer advanced transformation capabilities. As for data acquisition, ETL tasks are executed intermittently, to ensure that data sources don’t change during data acquisition and lead to inconsistencies across online transaction processing (OLTP) systems and the data warehouse. According to Gour, et al. [10] data is “extracted” from the data sources (line of business applications) using a data extraction tool via whatever data connectivity is available. Macura [2] Found data quality and integrity checking is performed as part of the transformation process, and corrective actions are built into the process. In addition, the output design must incorporate all facts and dimensions required to present both the aggregation levels required by the BI solution and any possible future requirements [11].

Data Analysis Technique

Data warehouse is developed to provide an easy way to data. “It is not an end itself but a means to an end”. That end is usually the need to undertake analysis and make decision by the use of data source (Ravat, Teste & Tournier, 2007). According to Han, et al. [12] there are different technique being used for data analysis and also there is direct relationship between data analysis technique and data model type.

According to Chen, et al. [13] query and reporting technique refers to answering a question, data retrieval form the data warehouse, transforming data into the suitable context, displaying the data in a user readable format. The most popular technique to extent the performance of query/reporting is multidimensional analysis. That is instead of sending multiple queries data is typically structured to allow fast and easy access to answers for the questions asked.

Multidimensional analysis allows end user to view a large number of inter depended factors that are involved in business problem and also to look at data that is in complex nature [14].

Ravat, et al. [15] found that a new technique for data analysis that differs from multidimensional analysis and query & reporting which uses a “discovery technique” is data mining. In data mining a user needs not ask specific questions but instead use special algorithms to analyze data and report the discovery that was made. The discoveries may include finding relationship between given data elements, data element clusters and other hidden patterns in a certain data elements [16].

The Concept of Data Warehouse Implementation

Payton and Handfield [17] Argue that Current technologies, including data warehouse, continue to be met with limited implementation success. When organizations seek to implement these technologies and deploy an outsourcing strategy, implementation success can be further complicated.

According [18] data warehouse implementation has two important steps. First is the DW design schema, which is built from the requirements of the business itself, BI, and Data Mining, The Second Step, Which Is the Decision-Making Support

Hayen, et al. [19] Conducted a study which shows that a data warehouse project is risky and expensive and two-third of the effort of developing a DW will eventually fail. In the same vein, the researchers cited three contributing factors for the data warehouse failure which are environment, technical and project management.Environment refers to changes within and outside the organization in terms of politics, business, mergers and lack of management support. These involve corporate failure and human error [19, 20].

Secondly, lack of technical competence that relates to data requirement, data definition and analysis, and quality of data from different units of the organization may lead to data warehouse failure. Also lack of data extraction and integration technique, and wrong selection of data warehouse model are another factors. Lastly, poor project management may lead to failure of data warehouse project so also accessibility and ownership are potential causes of failure [19, 20].

Phiriyayotha and Rotchanakitumnuai [21] Stress the importance of reviewing the success factors for Data ware-

house. The researchers studied the work of six researches in data warehousing and highlighted Commitment management support and sponsorship, Business user oriented change management, clear business vision, business driven methodology, business centric championship, strategic and extensible framework as implementation success factors.

Data Warehouse Architecture

The DWs have seen many milestones. Major contributions are credited to Ralph Kimball, and WH Inmon, both of which are responsible for the "dimensional" and "normalized" approach respectively. These two are the most popular approaches in DWs designs. Both of these approaches can be represented in Entity Relationship ER diagrams, and both of these approaches can be implemented as Bottom-up or Top down implementations [18].

Top-Down architecture/ Inmon Architecture

Inmon (1997) introduced the first architecture for data house as shown in fig.1 The first step is data extraction, data transformation and data loading from external sources for legacy system. In the process the data collected is stored in data staging area then the data and metadata are loaded into the data warehouse.

According to Chowdhury and Pal [22] integration between data marts and data warehouse is automatic. On the other hand, [Kimball and Ross [23]] criticized the top-down model on the basis of high implementation cost and takes time to reap the benefits (results).

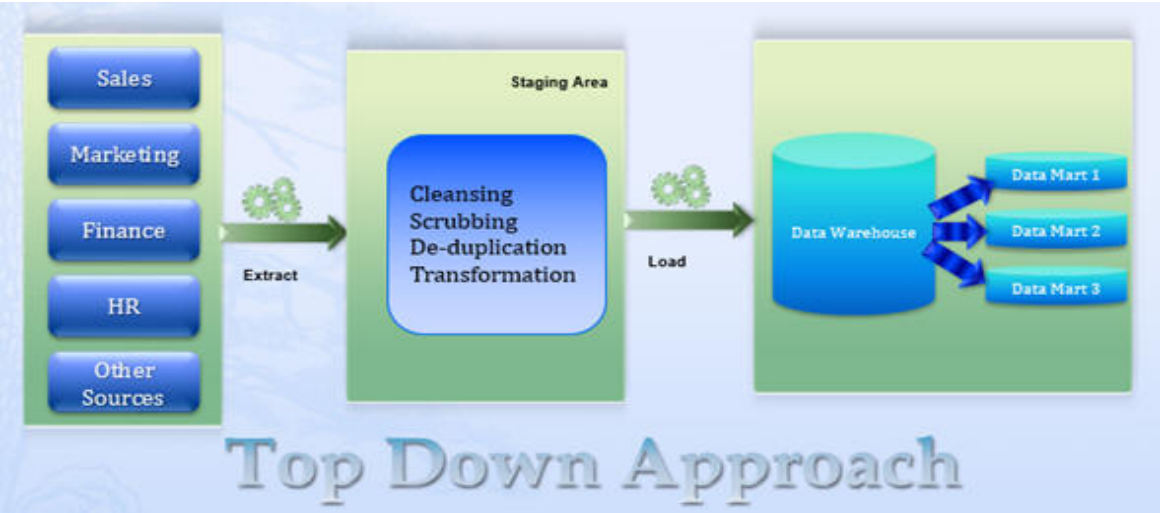

Figure 1: Top-Down Architecture of Data Warehouse
Source: Inmon (1997)

Bottom up Architecture/Kimbal Architecture

Another architecture for Data warehouse is bottom-up architecture as shown in figure 2, it offers incremental approach to Data warehouse design by Kimbal the process begins with designing departmental, then the staging area for data, independent data mart and the overall data warehouse.

Figure 2: Shows the Bottom up Architecture

Source: kimbal (1996)

III. The proposed framework

The paper adopt USE CASE as a common modeling language to bring together business analysts, software developers, architects, database designers, who are involved in software design and development so that they can understand the roles of sales manager, sales person, inventory/stock manager and customer as shown below.

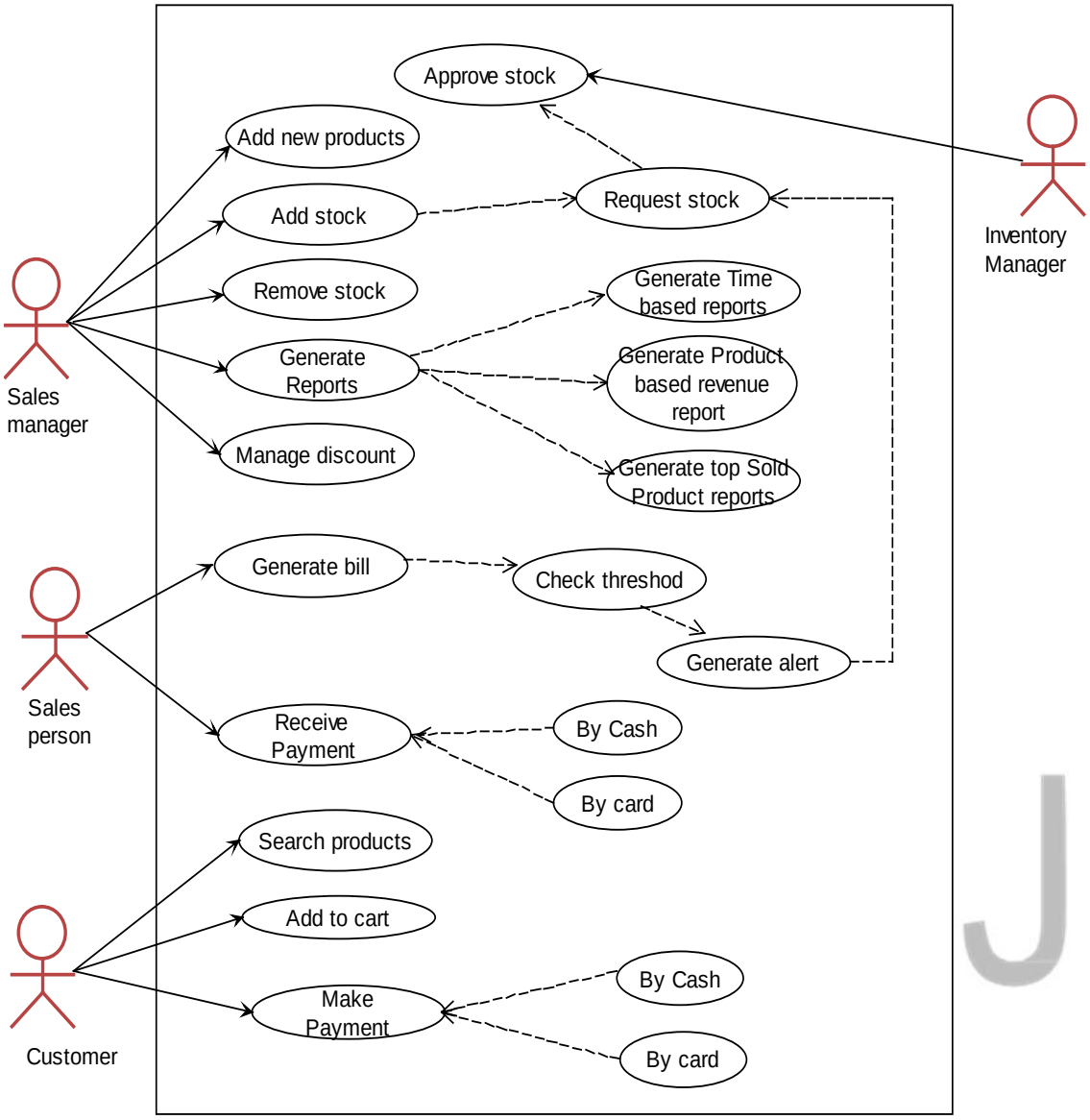

Figure 3 : USE CASE Model for a Super market

The Data Warehouse (DW) was developed using SQL server 2005 which is a Microsoft technology product that supports data warehousing and Business Intelligence (BI). The SQL server Express edition has a Relational Database Management Systems (RDBMS) that can store all the data required for the DW. It has a module that is capable of extracting data from different sources, integrating it and consolidating it in a single location for further analysis which is known as SQL Server Integration Services (SSIS). There is also functionality for developing a cube for the data mart, the front end application for analyzing the cube is known as SQL Server Analysis Services (SSAS).

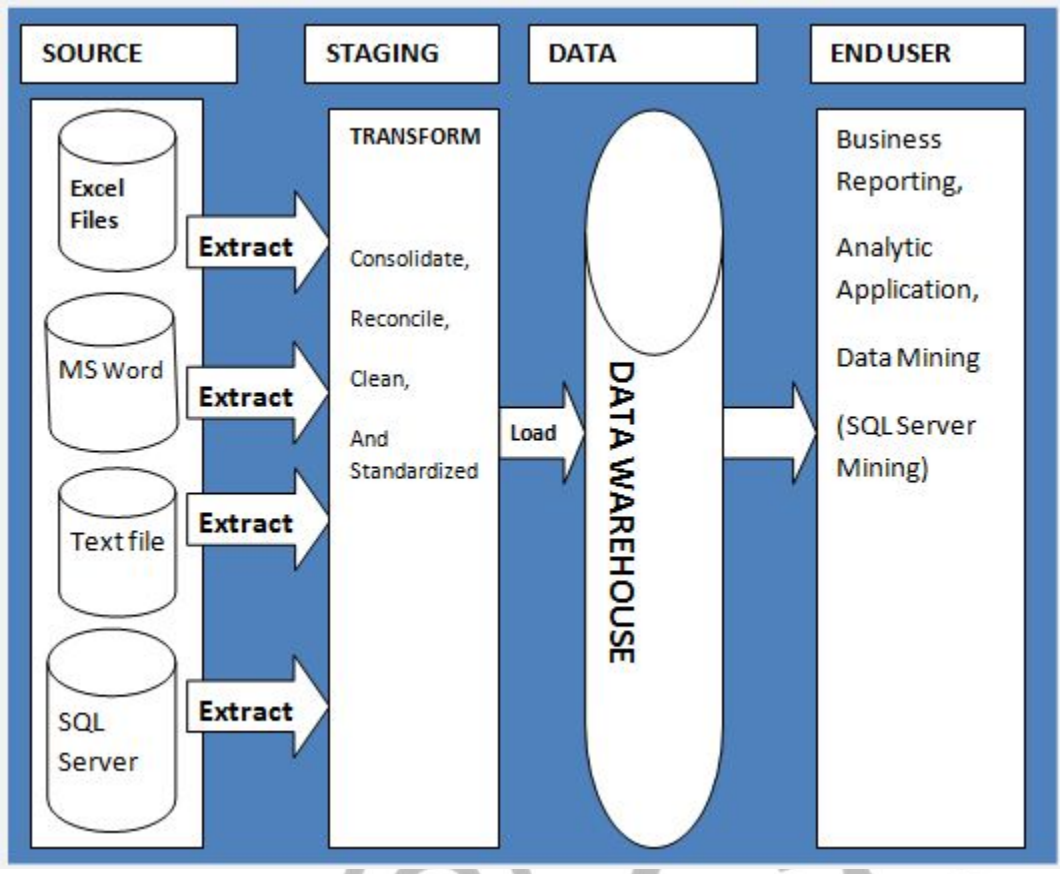

Figure 4: The Architectural Design for the Super Market DW

Source: The study, 2020

IV. The DW Implementation

The implementation covered both structured and unstructured data sources. The unstructured data source was MS Word files that keep store and supplier details which were transformed in to text file. While the structured data source was SQL Server that keeps Product sales data base and excel file that keeps customer details. After developing the DW with a well designed dimensional database with conformed dimension and surrogate keys. The conformed dimensions allow us to use dimensions in multiple data mart such as product dimension, customer dimension, date dimension, store dimension and supplier dimension. While the surrogate keys are artificial keys that link our dimensions and the two fact tables. The aim was to ensure uniqueness and improve performance. We use Microsoft SQL server Analysis Services which is business intelligence tool to perform the following analysis which are the identified user requirements.

Sample reports:

a) Product sales over time

The retrieved result from figure 20 showed that the profit margin recorded for the three stores. For the month of April 2015 furniture recorded highest margin of ₦ 89,500 while the grocery store recorded least i.e. ₦ 1,100. In the month of May 2015, it was furniture store that recorded the least and home and kitchen appliances recorded the highest margin. In 2016, it was furniture store that recorded highest margin for the two consecutive months totaled ₦ 496,500 while grocery is least for the two months aggregate. Based on the results from the two consecutive months for two different years, furniture store contributed most to the overall profit of the enterprise with a grand total of ₦ 604,600 while grocery store contributed only ₦ 104,150 for the period reviewed.

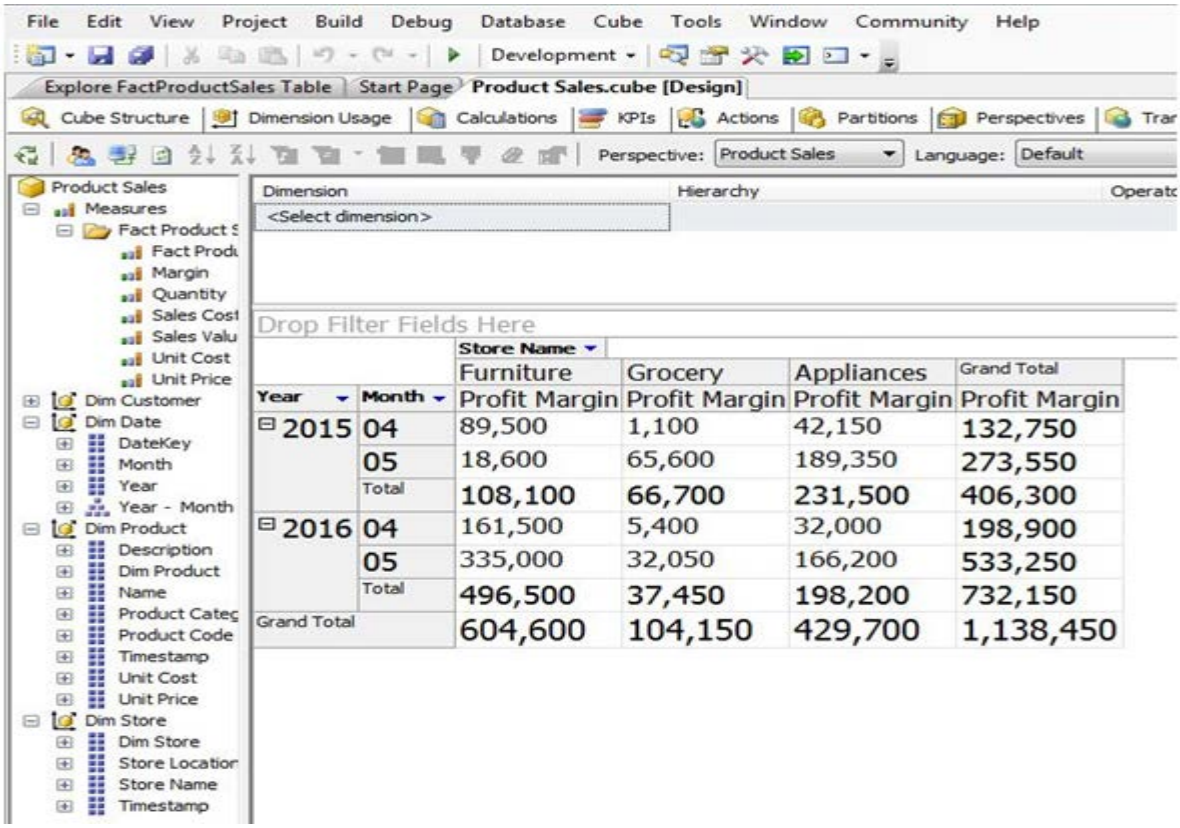

Figure 5: Retrieved Results for Product Sales Over Time based on store

b) Performance of a supplier based on his product

A super market usually deals with different vendors that supply goods to the business. Measuring their performance would make them to improve. Our DW evaluates suppliers based on two indicators (reject quantity and return quantity). Figure 21 showed the retrieved result of suppliers that supply Home and Kitchen appliances to the business. From the result in figure 21, we can discover the following instances. A given supplier has no

record of both rejected and returned products. (Refer to supplier with ID 2017003). Although, two suppliers supplied equal quantity of products but one has recorded higher number of returned products. (Refer to supplier with ID 2016001 and 2015004).Similarly, a supplier supplied higher quantity than others but they have the same number of rejected and returned products (refer to supplier with ID 2017002 and 2015005).

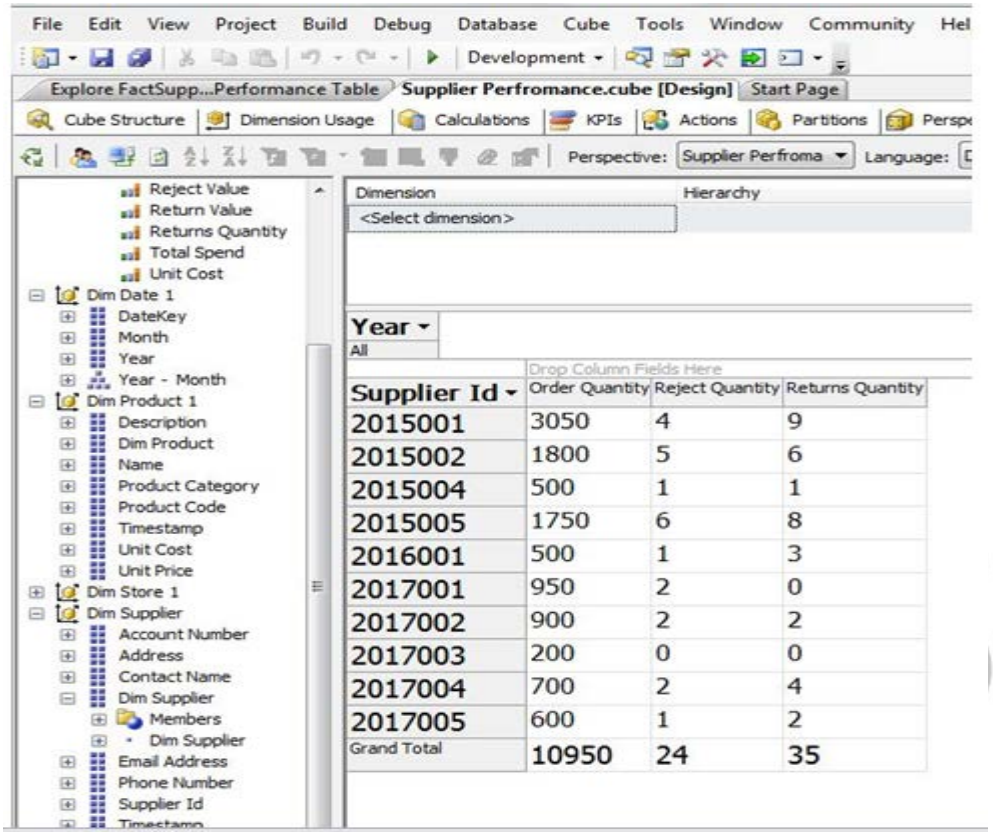

Figure 6: Result set for suppliers Performance

Conclusions

The desire for business organization to turn their data into actionable information that provide benefit and competitive advantage makes data warehousing to be an interesting area of study for both business managers and information technology specialists. In this study we have explored data ware house and its implementation in the context of retail business based on super market requirements.Data warehousing helps to standardize data across a particular organization thereby help in making smarter decision for companies. Moreover Sales retail businesses can identify products that are not doing well and drop them; thereby minimizing cost and also increase revenue (concentrate on high selling product).

The study would aid in identifying customer's common characteristics and behavior that purchases the same

products of a company and to predict customers who are likely to quit patronizing the company and move to a competitor company. DW serves as a foundation for data analysis. It basically supports business decision through encouraging users to examine data and make analysis in a better way. Since the data has been gathered in a particular repository, it easily facilitates measurement of the effect of different combination of factors such as demography, supply chain, customers' preference and can helps the analyst to work out the customer retention process and trend. Considering the importance of information to super markets the thesis intends to develop a DW for a super market.This research work will also served as a basis for other researchers in the field of Data mining and Business Intelligence which are emerging field of studies especially in our institutions

The study did not cover all the areas of Retail business; the focus was on product sales and supplier performance. The study focused only on consolidating data from two structured data sources and two unstructured data sources. Presentation and analysis was carried out in Business Intelligence studio.

References

[1] R. Jindal and S. Taneja, "Comparative Study of Data Warehouse Design Approaches: A Survey," *International Journal of Database Management Systems,* vol. 4, p. 33, 2012.
[2] M. Macura, "Integration Of Data From Heterogeneous Sources Using ETL Technology," *Computer Science,* vol. 15, pp. 109-132, 2014.
[3] I. Guratan, "The Design And Development Of A Data Warehouse Using Sales Database And Requirements Of A Retail Group," Master of science, Computer Engineering, Izmir Institute of Technology, 2005.
[4] E. M. Leonard, "Design and Implementation of an Enterprise Data Warehouse," Masters, Marquette University, 2011.
[5] O. E. Sheta and A. N. Eldeen, "The Technology of Using a Data Warehouse To Support Decision-Making In Health Care," *International Journal of Database Management Systems,* vol. 5, pp. 75-86, 2013.
[6] W. H. Inmon, *Building the data warehouse*: John wiley & sons, 2005.
[7] B. P. Başaran, "A Comparison Of Data Warehouse Design Models," Masters, Computer Engineering, ATILIM UNIVERSITY, 2005.
[8] R. Sherman, "Chapter 11 - Data Integration Design and Development," in *Business Intelligence Guidebook*, ed Boston: Morgan Kaufmann, 2015, pp. 275-299.
[9] M. S. Farhan, M. E. Marie, L. M. El-Fangary, and Y. K. Helmy, "Transforming Conceptual Model into Logical Model for Temporal Data Warehouse Security: A Case Study," *International Journal of Advanced Computer Science and Applications,*, vol. 3, pp. 115-122, 2012.
[10] V. Gour, S. S. Sarangdevo, G. S. Tanwar, and A. Sharma, "Improve Performance of Extract, Transform and Load(ETL) in Data Warehouse " *International Journal on Computer Science and Engineering* vol. 2, pp. 786-789 2010.
[11] R. Sherman, "Chapter 12 - Data Integration Processes," in *Business Intelligence Guidebook*, ed Boston: Morgan Kaufmann, 2015, pp. 301-333.
[12] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*: Elsevier, 2011.
[13] L.-D. Chen, T. Sakaguchi, and M. N. Frolick, "Data mining methods, applications, and tools," *Information systems management,* vol. 17, pp. 65-70, 2000.
[14] D. L. Moody and M. A. Kortink, "From enterprise models to dimensional models: a methodology for data warehouse and data mart design," in *DMDW*, 2000, p. 5.
[15] F. Ravat, O. Teste, and R. Tournier, "Olap aggregation function for textual data warehouse," in *ICEIS (1)*, 2007, pp. 151-156.
[16] M. J. Berry and G. Linoff, *Data mining techniques: for marketing, sales, and customer support*: John Wiley & Sons, Inc., 1997.
[17] F. C. Payton and R. Handfield, "Data Warehousing Implementation And Outsourcing Challenges: An Action Research Project With Solectron," *Communications of the Association for Information Systems,* vol. 1, pp. 633-648, 2003.
[18] A. R. Alazmi, "Data Warehousing Implementations: A Review," *International Journal on Data Mining and Intelligent Information Technology Applications (IJMIA),* vol. 4, pp. 9-25, 2014.
[19] R. L. Hayen, C. D. Rutashobya, and D. E. Vetter, "An investigation of the factors affecting data warehousing success," *Issues in Information Systems,* vol. 8, pp. 547-553, 2007.
[20] H. J. Watson and B. H. Wixom, "The current state of business intelligence," *Computer,* vol. 40, pp. 96-99, 2007.
[21] T. Phiriyayotha and S. Rotchanakitumnuai, "Data Warehouse Implementation Success Factors and the Impact of Leadership and Personality on the Relationship between Success Factors," 2013.
[22] R. Chowdhury and B. Pal, "Proposed hybrid data warehouse architecture based on data model," *International Journal of Computer Science and Communication,* vol. 1, pp. 211-213, 2010.
[23] R. Kimball and M. Ross, *The Data Warehouse Toolkit,* 3rd ed. New York: Robert Ipsen, 2002.