RESEARCH ARTICLE

The Role of Data Warehousing Concept for Improved Organizations Performance and Decision Making

Nwakanma Ifeanyi[1], Egbivwie Oghenevwoke[2], Azike Uchenna[3], Nwaobilo Amarachi[4], Ibeji Chinedum[5], Ohia Nwabueze[6]

[1-6]Department of Information Management Technology, Federal University of Technology Owerri, Imo State, Nigeria

[1] fraircos@yahoo.com, [2] oghene_harly@yahoo.com, [3] azikenicho@yahoo.com,

[4] just2009great@yahoo.com, [5] chinaedum.ibeji@yahoo.com, [6] danny_coms@yahoo.co.uk

Abstract - In challenging times good decision-making becomes critical. The best decisions are made when all the relevant data available is taken into consideration. The best possible source for that data is a well-designed data warehouse. The concept of data warehousing is not hard to understand. The concept is to create a permanent storage space for the data needed to support analysis, reporting, and other organizational activities. The goal of this paper is to elicit the crucial role of data warehousing in an organization performance and decision making.

Keywords: Databases, OLAP, Meta Data, Data Warehouse, Data Mining, Data Mart, Flat Files

I. INTRODUCTION

A data warehouse is a storehouse of an organization's historical data; it is a consolidation of organizational data. It is a database of databases. More generally, data warehouse is a collection of decision support technologies, aimed at enabling the knowledge worker, such as executive, manager, and analyst, to arrive at better and faster decisions [11].

A data warehouse is defined as a subject-oriented, integrated, nonvolatile, time-variant collection of data in support of management's decisions [4]. Data warehouses provide access to data for complex analysis, knowledge discovery, and decision-making.

Data warehouses are widely used within the largest and most complex businesses in the world. Use within moderately large organizations, even those with more than 1,000 employees remains surprisingly low at the moment. The concept of data warehousing is deceptively simple.

Data is collected periodically from the applications that support business processes and copied onto special dedicated computers. There it can be validated, reformatted, reorganized, summarized, restructured, and supplemented with data from other sources. The resulting data warehouse becomes the main source of information for report generation, analysis, and presentation through ad hoc reports, portals, and dashboards.

Data warehousing systems are designed to support online analytical processing (OLAP).

II. FUNDAMENTAL CHARACTERISTICS OF DATA WAREHOUSE

Understanding the characteristics of a data warehouse expound what a data warehouse really is, and its importance. There are four fundamental characteristics of data warehouse which are listed below [4]:

2.1 Subject Oriented: In data warehouse, data are categorized by subjects such as products and sales which provide the decision makers with an all-encompassing picture of an organization and distinguish it from operational database which is product oriented and primarily deals with transactions that modify the database.

2.2 Integration: Data warehouse is expected to be completely integrated as it is a place where data from various places are stored. Thus, all data must be in a consistent format.

2.3 Time Variant: Time dimension is very important since data warehouse contains historical data that help with the forecast and decision making.

2.4 Nonvolatile: End users cannot update or change data, once they are keyed to the warehouse. Data in the data warehouse are loaded and refreshed from operational systems. Thus, data warehouse mostly deals with data access. With these characteristics in mind, it is apparent that data warehouse is the decision support tool that organizations can make use of.
The next section is going to address what role a data warehouse play in an organization.

III. OVERVIEW OF DATA WAREHOUSE ARCHITECTURE

Building an Effective Data Warehouse requires a good architecture that will include tools for mining or extracting data from multiple operational databases and external sources.

Data warehouses and their architectures vary with respect to an organizations situation. According to [6], a building is constructed using architectural diagrams (blueprints) that clearly depict the building's infrastructure (structural elements, walls, electrical wiring, plumbing, etc). The best data warehouses are built from architectural models of enterprise infrastructure (policies, goals, measures, critical success factors, etc.).

Designing and rolling out a data warehouse is a complex process, consisting of the following activities [7]

1. Define the architecture, do capacity planning, and select the storage servers, database and OLAP servers, and tools.
2. Integrate the servers, storage, and client tools.
3. Design the warehouse schema and views.
4. Define the physical warehouse organization, data placement, partitioning, and access methods.
5. Connect the sources using gateways, ODBC drivers, or other wrappers.
6. Design and implement scripts for data extraction, cleaning, transformation, load, and refresh.
7. Populate the repository with the schema and view definitions, scripts, and other metadata.
8. Design and implement end-user applications.
9. Roll out the warehouse and applications.

Three common architectures of data warehouse are [8]:
1. Data Warehouse Architecture (Basic)
2. Data Warehouse Architecture (with a Staging Area)
3. Data Warehouse Architecture (with a Staging Area and Data Marts)

3.1 DATA WAREHOUSE ARCHITECTURE (BASIC)

In the basic architecture the end users directly access data used for analysis, reporting, and mining in an organization from several data source systems (operational systems and flat files) through the data warehouse which consist of Meta data, Summary data, and raw data.

The operational data from the data source are not clean before they are putting it into the warehouse.

Flat files are data files that contain record with no structured relationships unlike relational database while Meta Data is information about the data.

3.2 DATA WAREHOUSE ARCHITECTURE (WITH A STAGING AREA)

This second architecture is just as the first one only that you need to clean and process your operational data before putting it into the warehouse. You can do this programmatically, although most data warehouses use a Staging area instead.
A staging area simplifies building summaries and general warehouse management.
Staging area is a place where data is processed (cleansed) before entering the warehouse.

3.3 DATA WAREHOUSE ARCHITECTURE (WITH A STAGING AREA AND DATA MARTS)

In the architecture beside a staging area that processes data before entering the warehouse, Data Marts are added to customize the warehouse's architecture for different groups within an organization.

A data mart is a focused subset of a data warehouse that deals with a singles area of data and is organized for quick analysis.

In organizations where there are many departments is this will be helpful because with the provision of data marts every department could analyze historical data quickly say for example an organization with purchasing, sales and inventories department. Each of these departments accessing the data warehouse can be separated with data Mata i.e. they have to access a subset of the data warehouse, this brings about quicker access.

IV. THE ROLE OF DATA WAREHOUSES IN ORGANIZATION

Data warehouses exist to facilitate complex, data-intensive and frequent adhoc queries [11]. The purpose of the Data Warehouse mostly is to integrate corporate data in an organisation.

It contains the "single version of truth" for the organization that has been carefully constructed from data stored in disparate internal and external operational databases. The amount of data in the Data Warehouse is massive.

Data is stored at a very granular level of detail. For example, coca cola want to introduce a new quantity of coke, say 50cl to a market in Africa but want to see which country in Africa has the highest demand for coke before deciding whether to introduce it or not. This decision could be reach using different approaches. The manual approach would be retrieving sale data directly from database in each country in Africa, and then compare the sets of data against each other to figure out which country has the highest sale. This approach would leave all the work to you. But a suitable approach would be that of common data storage method, also known as data warehousing.

Using this method, all the data from the various databases of countries you intend to integrate are extracted, transformed and loaded. That means that the data warehouse first pulls all the data from the various data

sources. Then, the data warehouse converts all the data into a common format so that one set of data is compatible with another. Then it loads this new data into its own database.

When you submit your query, the data warehouse locates the data, retrieves it and presents it to you in an integrated view. In the case of our example, the data warehouse would locate the latest information about the country with the highest sales send the view back to you.

Data warehouses must provide far greater and more efficient query support than is demanded of transactional databases. The data warehouse access component supports enhanced spreadsheet functionality, efficient query processing, structured queries, adhoc queries, data mining and materialized views [11]. Particularly enhanced spreadsheet functionality includes support for state-of-the art spreadsheet applications as well as for OLAP applications programs [11].

V. THE IMPORTANCE OF DATA WAREHOUSES IN ORGANIZATIONS

The importance of data warehouse in an organization tends to explain why data warehouse in needed in an organization. Since a data warehouse reflects the business model of an enterprise that make is an important aspect of an organization. The following are some importance of data warehouse:

1. Repository for historical information for comparative and competitive analysis
2. Ability to enhanced data quality and completeness.
3. Real-time consolidation of financial data becomes practical.
4. The IT costs and staff dedicated to reporting are greatly reduced.
5. Allow business process redesign and align with business strategy.
6. Give end users freedom to carry out wide-ranging analysis in various manners
7. Simplify the process of data access
8. Identify market trends
9. Reduce operation costs
10. Allow business process redesign and align with business strategy

The ability of a data warehouse to analyze and execute business decisions based on data from multiple sources is of utmost importance. For example, an organization has collected valuable data and stored it in 10 databases.

A data warehouse is not only a convenient way to analyze and compare data in all the databases, but it can also give historical data and perspective. Using data warehouse, one can look at past trends, whether they be product sales or customers or whatever and may be do some predictions of what is going to happen in the future.

Also data retrieved from multiple databases is not constrained by the tables in each of those databases. A data warehouse by itself does not create value, but value comes from the use of the data in the warehouse. In support of a low cost strategy, the data warehouse can provide savings in billing processes, reduce fraud losses, and reduce the cost of reporting. The data warehouses can provide analysts with pre-calculated reports and graphs. This increases the productivity of business analysts.

Most companies can benefit from a data warehouse when the proper tools are in place and users are trained in analysis of results.

5.1 KEYS TO A SUCCESSFUL WAREHOUSING PROJECT

1. Identified and involved warehouse users
2. Strong and committed leadership
3. Diversified project team
4. Established partnerships with all key source data holders
5. Incremental project plan the produces fast results
6. Correct design philosophy.

VI. CONCLUSION AND FURTHER STUDY

Data warehousing provides leverage for management in an organization. Effective decision making is the major function of every management in an organization; data warehouses facilitate meaningful research which facilitates effective management processes. With data warehouse in place, each department in an organization can share data and though the costs of operations will be reduced, this also allows users or management to perform extensive analysis across all departments in the organization.

As a future work we plan to conduct a field study to identify the critical success factors that affects Data warehouse implementation in an organization.

REFERENCES

[1] Priya Chaplot, *An Introduction to Data Warehousing; Educational Research Assessment Analyst Research and Institutional Effectiveness*, Mt. San Antonio College, 2007

[2] Alex Berson, Stephen J. Smith, *Data Warehousing, Data Mining and OLTP*, Pg: 4, 14-16, 1997

[3] Joseph Guerra, *Why You Need a Data Warehouse*, SVP, CTO & Chief Architect David Andrews, 2003

[4] Inmon, William H, *Building the Data Warehouse*, 4th Edition, Wiley Publishing, Pg: 32, 2005

[5] *Data Warehousing and OLAP* " (PDF). Retrieved 2031-11-11.

[6] Alan Perkins, Critical Success Factors for Data Warehouse Engineering, Pg: 1- 4, 2003

[7] Jiawei Han & Micheline Kamber' Data Mining – Concepts & Techniques, Jain Pei, Morgan Kaufmann Publication, 2003

[8] Theju Paul, Data Warehouse Architectures; Management Consultant at BTS ViiN, 2008

[9] Data Warehousing Techniques and Standards, http://www.ecs.umass.edu/masssafe/PDFS for Site/Traffic Records/Data Warehousing Techniques and Standards.pdf, Accessed date: 15th September 2014.

[10] Data warehouse architectures, http://en.wikipedia.org/wiki/Data_warehouse_architectures,   Accessed date: 15th September 2014.

[11] http://www.peterindia.net/DataWarehousingView.html.