Modeling Architecture for Multimedia Data Warehouse to Support Decision-Making Processes in Industries

Elusoji, A.A (PhD)
Department of Computer Technology
Yaba College of Technology, Yaba – Lagos, Nigeria
elusoji872@yahoo.com

Aiyegbusi A.E
Department of Computer Science
Micheal Otedola College of Primary Education, Epe-Lagos, Nigeria
aiyegbusied@yahoo.com

Akeredolu, G.R
Department of Computer Science
Federal Polytechnic, Idah – Nigeria
akmasgbenga3@gmail.com

ABSTRACT

The data warehouses are considered modern ancient techniques, since the early days for the relational databases, the idea of the keeping a historical data for reference when it needed has been originated, and the idea was primitive to create archives for the historical data to save these data, despite of the usage of a special techniques for the recovery of these data from the different storage modes. Data Warehouse and Data mining are technologies that deliver optimally valuable information to ease effective decision making. This paper describes multimedia data ware house to extract, transform and load data, and the ways to improve these techniques to have maximum benefit of data warehouse in a business industrial environment. Further this paper uses the approach of indexing technique and partitioning technique to obtain an efficient storage and processing of multimedia data warehouse and also analyze the multimedia data better

Keywords: Data Warehouses, OLAP Operation, DSS, Data Quality.

I. INTRODUCTION

A data warehouse is a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management's decisions. The data warehouse contains granular corporate data. Data in the data warehouse is able to be used for many different purposes, including sitting and waiting for future requirements which are unknown today [1]. Data warehouse provides the primary support for Decision Support Systems (DSS) and Business Intelligence (BI) systems. Data warehouse, combined with On-Line Analytical Processing (OLAP) operations, has become and more popular in Decision Support Systems and Business Intelligence systems. The most popular data model of Data warehouse is multidimensional model, which consists of a group of dimension tables and one fact table according to the functional requirements [2]. The purpose of a data warehouse is to ensure the appropriate data is available to the appropriate end user at the appropriate time [3]. Data warehouses are based on multidimensional modeling. Using On-Line Analytical Processing tools, decision makers navigate through and analyze multidimensional data [4]. Data warehouse uses a data model that is based on multidimensional data model.

This model is also known as a data cube which allows data to be modeled and viewed in multiple dimensions [5]. And the schema of a data warehouse lies on two kinds of elements: facts and dimensions. Facts are used to memorize measures about situations or events [6]. Dimensions are used to analyze these measures, particularly through aggregations operations (counting, summation, average, etc.) [7]. Data Quality (DQ) is the crucial factor in data warehouse creation and data integration. The data warehouse must fail and cause a great economic loss and decision fault without insight analysis of data problems [8]. The quality of data is often evaluated to determine usability and to establish the processes necessary for improving data quality. Data quality may be measured objectively or subjectively. Data quality is a state of completeness, validity, consistency, timeliness and accuracy that make data appropriate for a specific use [9].

2. RELATED WORKS

In this section we will review related work in Data Warehouse Design and Implementation. The paper introduced by [15], covers the full lifecycle of the data warehouse, and allows capturing the interrelationships between different quality factors and helps the interested user to organize them in order to fulfill specific quality goals. Furthermore, they prove how the quality management of the data warehouse can guide the process of data warehouse evolution, by tracking the interrelationships between the components of the data warehouse.

Finally, they presented a case study, as a proof of concept for the proposed methodology [26]. The paper introduced by [16] describes a study which explores modeling of the dynamic parts of the data warehouse. This meta model enables data warehouse management, design and evolution based on a high level conceptual perspective, which can be linked to the actual structural and physical aspects of the data warehouse architecture. Moreover, this meta model is capable of modeling complex activities, their interrelationships, the relationship of activities with data sources and execution details [16]. The paper introduced by [17] is to discover the main critical success factors (CSF) that leaded to an efficient implementation of DW in different organizations, by comparing two organizations namely: First American Corporation (FAC) and Whirlpool to come up with a more general (CSF) to guide other organizations in implementing DW efficiently. The result from this study showed that FAC Corporation had greater returns from data warehousing than Whirlpool. After that and based on them extensive study of these organizations and other related resource according to CSFs, they categorized these (CSF) into five main categories to help other organization in implementing DW efficiently and avoiding data warehouse killers, based on these factors [17].

The paper introduced by [9], evaluates the data quality of decision databases and evaluates the model at different dimensions like accuracy derivation integrity, consistency, timeliness, completeness, validity, precision and interpretability, on various data sets after migration. The proposed data quality assessment model evaluates the data at different dimensions to give confidence for the end users to rely on their businesses. Author extended to classify various data sets which are suitable for decision making. The results reveal the proposed model is performing an average of 12.8 percent of improvement in evaluation criteria dimensions with respect to the selected case study [18].

3 THE DATA WAREHOUSE ARCHITECTURE

The architecture of Data warehouse is divided into two (2) portions / parts:

(a). The Back Room is composed of:

a) Operational Source System: At this stage data is gathered from different sources, and passed onto data staging area [4].
b) Data Staging Area: Here few operations are performed on data collected from various sources, operations are:
    i. Extract: Extract data from different sources.
    ii. Transform: Transform data into useful information; perform data cleaning / cleansing and data aggregation.
    iii. Load: Load the information on Data Warehouse / Data Marts [4].
c) Data Representation Area: It consists of "Data Warehouse and Data Marts". These are used to store historical data in a bulk, which are later use for decision making [4]. Data from here is passed to OLAP server, which apply techniques like MOLAP and ROLAP to increase performance [4].

(b). The Front Room consists of:

I). Data Access Tools: It consists of tools to access data like: Analysis / OLAP Query / Reporting Data Mining [4]

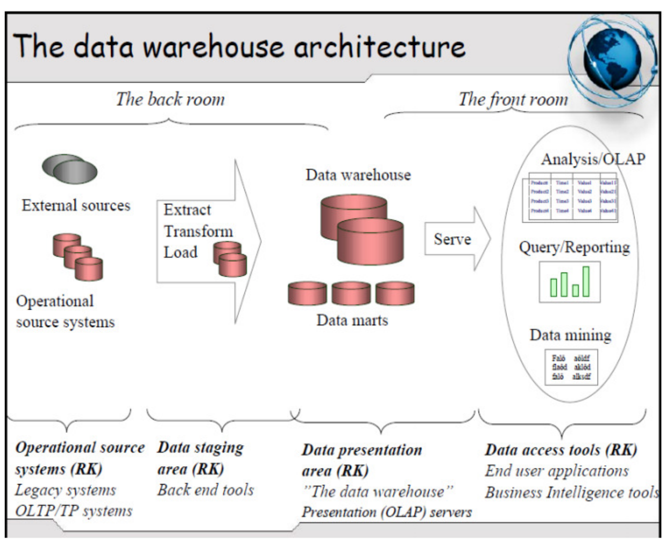

Figure 1. The Data Warehouse Architecture

4- DATA MINING TECHNIQUE

Data mining (DM) is one of the most important techniques that are used to discover required knowledge for intended enterprise. Data mining derives its name from the similarities between searching for valuable information in a large database and mining rocks for a vein of valuable ore. Since mining for gold in rocks is usually called "gold mining" and not "rock mining", thus by analogy, data mining should have been called "knowledge mining" instead [14]. Data mining is the knowledge discovery process by analyzing the large volumes of data from various perspectives and summarizing it into useful information [15]. Data mining is the process of discovering interesting knowledge, such as patterns, associations, changes, anomalies, and significant structures from large amount of data stored in databases, data warehouse, or other information repositories [16].

Data mining refers to discover useful, previously unknown knowledge by analyzing large and complex data sets. Data mining is defined as the extraction of patterns or models from observed data [12]. Data Mining refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. The goal of data mining is to allow a corporation to improve its marketing, sales, and customer support operations through a better understanding of its customers. Data mining, transforms data into actionable results [17].Other similar terms referring to data mining are: data dredging, knowledge extraction and pattern discovery [14].

5. DECISION MAKING USING A DATA WAREHOUSE

A data warehouse is an analytical database that is used as the foundation of a decision support system. It is designed for large volumes of read-only data, providing intuitive access to information that will be used in making decisions. A data warehouse is created as ongoing commitment by the organization to ensure the appropriate data is available to the appropriate end user at the appropriate time. A Decision Support System (DSS) improve the process of decision making in complex systems. It can range from a system that answer simple queries and allows a subsequent decision to be made, to a system that employ artificial intelligence and provides detailed querying across a spectrum of related datasets. Amongst the most important application areas of DSS are those complicated systems that directly "answer" questions, in particular high level "what-if" scenario modeling. Over the last decade there was a transition to decision support using data warehouses [18]. The data warehouse environment is more controlled and therefore more reliable for decision support than the previous methods.

The data warehouse environment supports the entire decision support requirements by providing high-quality information, made available by accurate and effective cleaning routines and using consistent and valid data transformation rules and documented pre summarization of data values. It contains one single source of accurate, reliable information that can be used for analysis. Data Warehouses (DW) integrate data from multiple heterogeneous information sources and transform them into a multidimensional representation for decision support applications [11]. Apart from a complex architecture, involving data sources, the data staging area, operational data stores, the global data warehouse, the client data marts, etc., a data warehouse is also characterized by a complex lifecycle. In a permanent design phase, the designer has to produce and maintain a conceptual model and a usually voluminous logical schema, accompanied by a detailed physical design for efficiency reasons [25]. The designer must also deal with data warehouse administrative processes, which are complex in structure, large in number and hard to code; deadlines must be met for the population of the data warehouse and contingency actions taken in the case of errors[13].

Finally, the evolution phase involves a combination of design and administration tasks: as time passes, the business rules of an organization change, new data are requested by the end users, new sources of information become available, and the data warehouse architecture must evolve to efficiently support the decision-making process within the organization that owns the data warehouse. All the data warehouse components, processes and data should be tracked and administered via a metadata repository. In [24], we presented a metadata modeling approach which enables the capturing of the static parts of the architecture of a data warehouse. The linkage of the architecture model to quality parameters (in the form of a quality model) and its implementation in the metadata repository Concept Base have been formally described in [24]. Presents a methodology for the exploitation of the information found in the metadata repository and the quality-oriented evolution of a data warehouse based on the architecture and quality model. In this paper, we complement these results with meta models and support tools for the dynamic part of the data warehouse environment: the operational data warehouse processes. The combination of all the data warehouse viewpoints is depicted in Fig. 2.

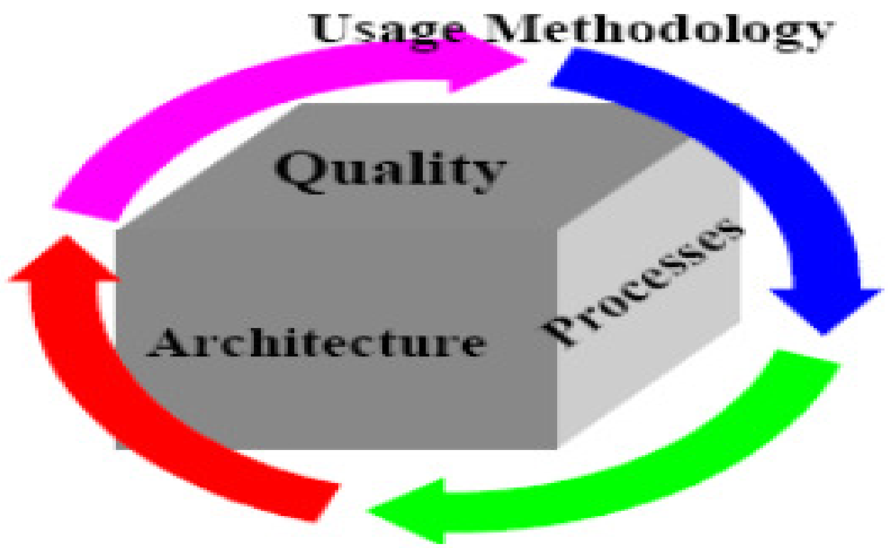

Fig. 2. The different viewpoints for the metadata repository of a data warehouse.

In [13] a basic meta model for data warehouse architecture and quality has been presented as in Fig. 2. The framework describes a data warehouse in three perspectives: a conceptual, a logical and a physical perspective. Each perspective is partitioned into the three traditional data warehouse levels: source, data warehouse and client level [22]. On the meta model layer, the framework gives a notation for data warehouse architectures by specifying meta-classes for the usual data warehouse objects like data store, relation, view, etc. On the metadata layer, the meta model is instantiated with the concrete architecture of a data warehouse, involving its schema definition, indexes, table spaces, etc. The lowest layer in Fig. 3 represents the actual processes and data.

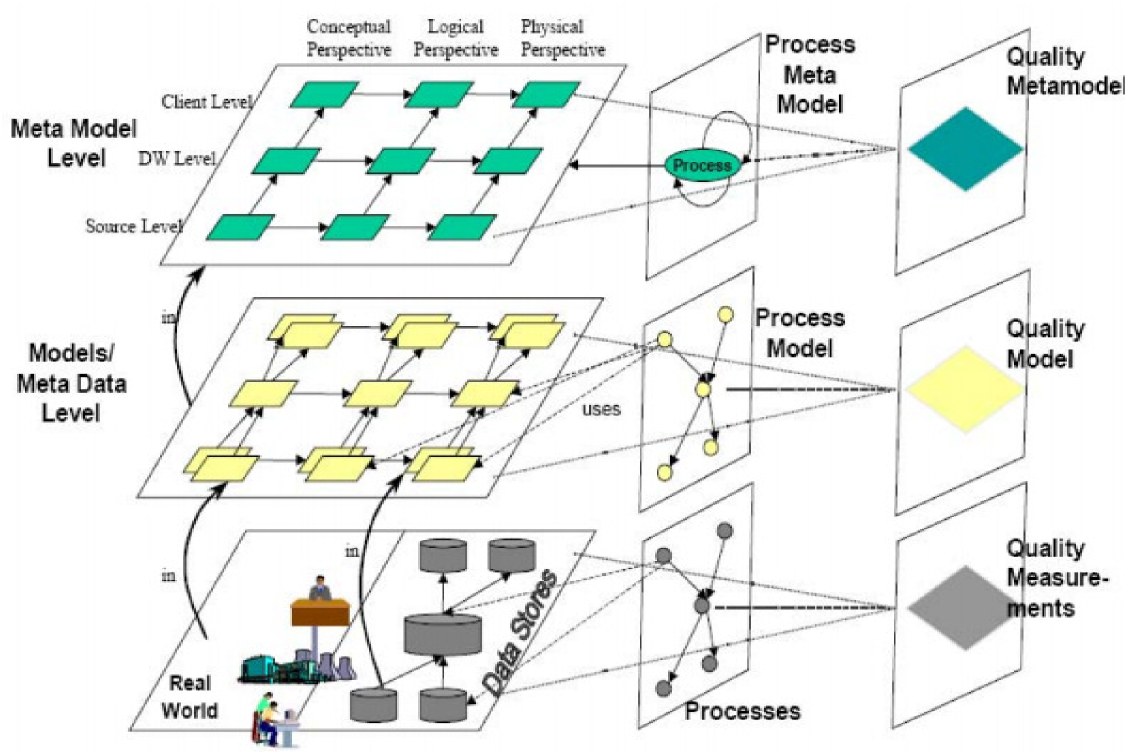

Figure 3: Frame work for Data Warehousing Architecture.

6. DATA WAREHOUSE DESIGN

A DW can be built using a top-down approach, a bottom-up approach, or a combination of both. The top-down approach starts with the overall design and planning. It is useful in cases where the technology is mature and well known, and where the business problems that must be solved are clear and well understood. The bottom-up approach starts with experiments and prototypes. This is useful in the early stage of business modeling and technology development [23]. It allows an organization to move forward at considerably less expense and to evaluate the benefits of the technology before making significant commitments. In the combined approach, an organization can exploit the planned and strategic nature of the top-down approach while retaining the rapid implementation and opportunistic application of the bottom- up approach.

In general, the warehouse design process consists of the following steps:

[1] Select the "business Process" to model, if business processes are multiple so "data warehouse model" must be followed, but if "business process" involve a single process so then "data mart" must be made [3].

[2] Select the "grain" of "business process", where "grain" is the level at which a fact in a fact table is represented [3].

[3] Select the "dimensions" that are applied to each fact, dimension is usually categorized as time, date, product and geography [3].

[4] Select the "measures" that will populate each fact, they are generally in a form of numbers [3]. Data warehouses normally adopt three-tier architecture:
a) Bottom tier: It is a "warehouse database server", which is almost "relational database system". Data from different sources are mined through gateways. A gateway is maintained by underlying DBMS [3].
b) Middle tier: It is an "OLAP server", which is applied by using "relational OLAP model" [3].
c) Top tier: It is a "Client", which consist of tools for querying, reporting, analysis and data mining [3].

7. Multimedia data extraction, Integration and modeling

Multimedia data is extracted from the operational sources. The relevant characteristic of multimedia data should be extracted according to the analysis goal. Multimedia data is usually described by different levels of feature of abstraction. Low level features (color, texture, shape, etc) are widely used to describe multimedia data as these data can be extracted automatically using program. These features seldom represent the semantic content of the multimedia object. With regard to the data access and analysis, the high level semantic content is important. For this reason, this work has included low level features descriptor, high level semantic feature descriptor and calculated feature descriptors of multimedia content.
Calculated features are features which can be extracted from multimedia data processing and can also be derived from the low level and high level features for example, face nodal points and distance between major nodal points can be extracted from face image. These extracted features are known as calculated feature [20]. After extracting multimedia data, low level feature extraction process takes place which is done automatically by using program and high level feature extraction process carried out manually. Some basic characteristics – file size, filename, author name, format, compression rate, duration for videos or sounds and resolution for images is also extracted automatically using program. After extracting these data, multimedia data is compressed using existing lossy technique [19]. After preparing the data in extraction part data is ready to get loaded in data warehouse. Following figure shows the low level and high level feature extraction process.

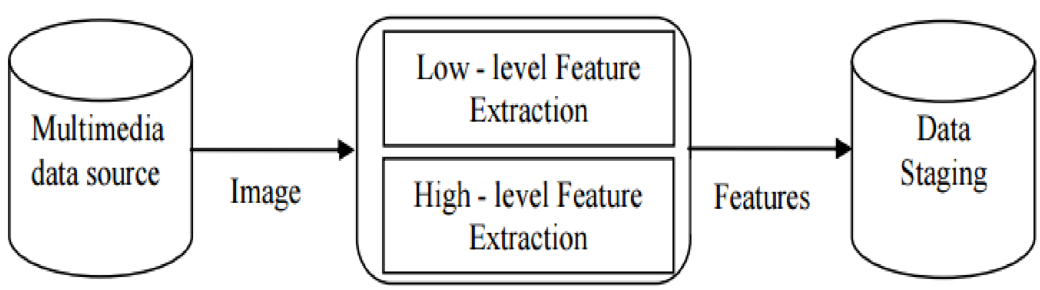

Fig. 4 Feature Extraction process

Data warehouse allows the data to be modeled in multidimensional way and to be observed from different perspective. Dimensional model of warehouse allows the creation of appropriate analysis contexts and the preparation of data for analysis [22]. This requires the building of multimedia data cubes on which OLAP operation are performed.

8. EXPERIMENTAL RESULTS

To validate our proposed architectural framework for multimedia data warehouse, experiment has been performed on small scale OLAP environment. Data compression is applied only on multimedia data. To shows the query duration for simple, middle and complex level queries, two sample queries for each level is evaluated. Following figure 5 shows the query performance for compressed image data without using indexing and partitioning approach.

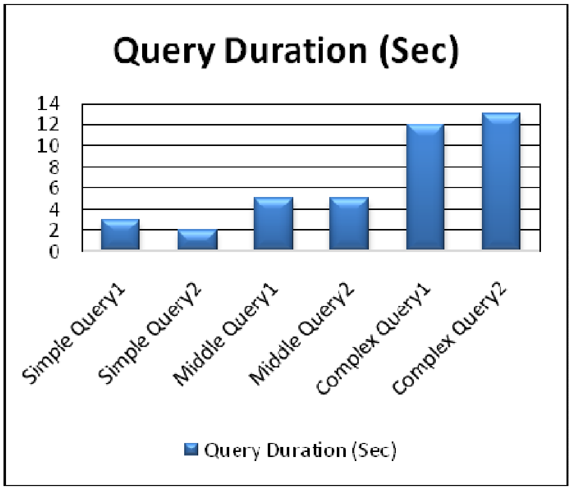

Fig. 5 Query duration for set of queries

To shows the query duration for simple, middle and complex level queries, the same two sample queries for each level is evaluated on small scale OLAP environment. Following figure 6 shows the query performance for compressed image data with using indexing approach.

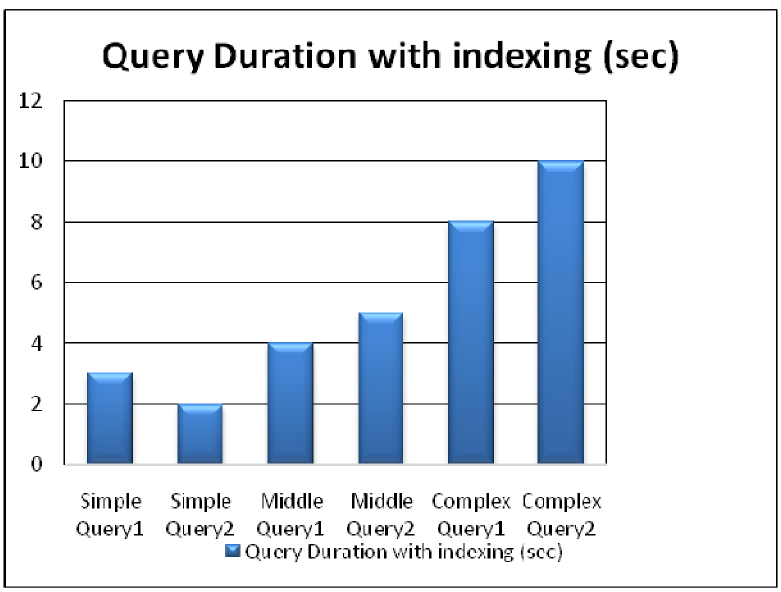

Fig. 6: Query duration for set of queries by using indexing

To shows the query duration for simple, middle and complex level queries, the same two sample queries for each level is evaluated on small scale OLAP environment. Following figure 7 shows the query performance for compressed image data with using indexing and partitioning approach together.

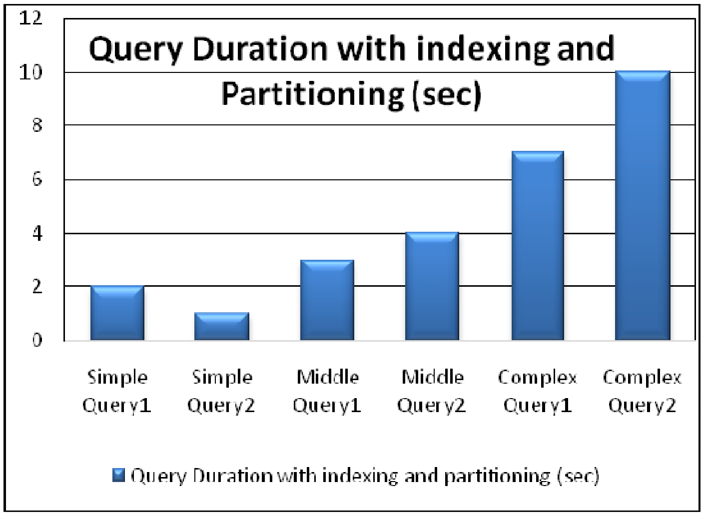

Fig. 7 Query duration for set of queries by using indexing and partitioning

Above figure shows the improved efficiency in data retrieval by using indexing and partitioning as compared to the data retrieval without using indexing and partitioning approach.

9. CONCLUSION

This paper has presented a systematic approach on building architectural framework for multimedia data warehouse in a generic way, so that the techniques can be applied to a wide range of multimedia data warehouse models and implementations. The implementation of these technique helps to build better multimedia model. Storage efficiency is improved by using provided compression and partitioning technique. Access and analysis efficiency is improved by representing multimedia data by multilevel features, by applying indexing technique and by using partitioning technique. By using this proposed approach, we not only obtain an efficient storage and processing of multimedia data warehouse but we also analyze multimedia data better.

REFERENCES

[1] [1] H. William Inmon, "Building the Data Warehouse". Fourth Edition Published by Wiley Publishing, Inc., Indianapolis, Indiana.2005.
[2] [2] R. Kimball, "The data warehouse lifecycle toolkit". 1st Edition ed.: New York, John Wiley and Sons, 1998.
[3] [3] K.W. Chau, Y. Cao, M. Anson and J. Zhang, "Application of Data Warehouse and Decision Support System in Construction Management". Automation in Construction 12 (2) (2002) 213-224.
[4] [4] Nicolas Prat, Isabelle Comyn-Wattiau and Jacky Akoka , "Combining objects with rules to represent aggregation knowledge in data warehouse and OLAP systems". Data & Knowledge Engineering 70 (2011) 732–752.
[5] [5] Singhal, Anoop, "Data Warehousing and Data Mining Techniques for Cyber Security". Springer. United states of America.2007.
[6] [6] Bhanali, Neera,"Strategic Data Warehousing: Achieving Alignment with Business". CRC Prees . United State of America. 2010.
[7] [7] Wang, John. "Encyclopedia of Data Warehousing and Mining ".Second Edition. Published by Information Science Reference. United States of America. 2009.
[8] [8] Yu. Huang, Xiao-yi. Zhang, Yuan Zhen , Guo-quan. Jiang, "A Universal Data Cleaning Framework Based on User Model ", IEEE,ISECS International Computing, Communication, Control , and Management, Sanya, China, Aug 2009.
[9] [9] T.N. Manjunath, S. Hegadi Ravindra, G.K. Ravikumar, "Analysis of Data Quality Aspects in Data Warehouse Systems". International Journal of Computer Science and Information Technologies, Vol. 2 (1), 2011, 477-485.
[10] [10] Paul Lane, "Oracle Database Data Warehousing Guide". 11g Release 2 (11.2), September 2011.
[11] [11] Stephen T C Wong, Kent Soo Hoo Jr, Robert C Knowlton, Kenneth D Laxer, Xinhau Cao, Randall A Hawkins, William P Dillon, Ronald L Arenson,"Design and Applications of a Multimodality Image Data Warehouse Framework," Journal of the American Medical Informatics Association Volume 9 No. 3, pp. 239-254, May / Jun 2002.
[12] [12] Tania Cerquitelli, Genoveva Vargas Solar, Jose Luis Zechinelli Martini, "Building Multimedia Data Warehouse from Distributed Data", eGnosis Vol-2, 2004.

[13] [13] Wookey Lee, Yongkyu Kim, Yunsun Lee, Jinho Kim, "Developing multimedia data warehouse of education on-demand systems", pp. 942- 945, IEEE 1999.
[14] [14] Xin Jin, Jiawei Han, Liangliang Cao, Jiebo Luo, Bolin Ding, Cindy Xide Lin, "Visual Cube and On-Line Analytical Processing of Images", ACM 2010.
[15] [15] Panos vassiladis, Mokrane Bouzegeghoub and Christoph Quix. "Towards quality-oriented data warehouse usage and evolution". Information Systems Vol. 25, No. 2, pp. 89-l 15, 2000.
[16] [16] Leo Willyanto Santoso and Kartika Gunadi. "A proposal of data quality for data warehouses environment". Journal Information VOL. 7, NO. 2, November 2006: 143 – 148
[17] [17] Amer Nizar AbuAli and Haifa Yousef Abu-Addose. "Data Warehouse Critical Success Factors". European Journal of Scientific Research ISSN 1450-216X Vol.42 No.2 (2010), pp.326-335.
[18] [18] T.N. Manjunath, S. Hegadi Ravindra, "Data Quality Assessment Model for Data Migration Business Enterprise". International Journal of Engineering and Technology (IJET) ISSN: 0975-4024, Vol 5 No 1 FebMar 2013.
[19] [19] Michael Goebel, Le Gruenwald, "A Survey of Data Mining and Knowledge Discovery Software Tools", SIGKDD Explorations. Copyright © 1999 ACM SIGKDD, June 1999, Volume 1, Issue 1, pages 20-33, provider: citeseer 2009.
[20] [20] Max Bramer, the book "Principles of Data Mining", Printed on acid-free paper © Springer-Verlag London Limited 2007.
[21] [21] CMPUT690, the book "Principles of Knowledge Discovery in Databases"\Chapter I: Introduction to Data Mining, © Osmar R. Zaïane, 1999.
[22] [22] Rupali, Gaurav Gupta, "Data Mining: Techniques, Applications and Issues", International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), ISSN: 2277 – 9043, Volume 2, Issue 2, February 2013.
[23] [23] Slavco Velickov and Dimitri Solomatine, "Predictive Data Mining: Practical Examples", Artificial Intelligence in Civil Engineering. Proc. 2nd Joint, Workshop, Cottbus, Germany. ISBN 3-934934-00-5, March 2000.
[24] [24] Radhakrishnan B, Shineraj G, Anver Muhammed K.M,"Application of Data Mining In Marketing" , IJCSN International Journal of Computer Science and Network, ISSN (Online) : 2277-5420 http://www.ijcsn.org , Volume 2, Issue 5, October 2013.
[25] [25] M. Jarke, M.A.Jeusfeld, C. Quix, P. Vassiliadis: Architecture and quality in data warehouses: An extended repository approach. Information Systems, 24(3): 229-253 (1999). A previous version appeared in Proc. 10th Conf. of Advanced Information Systems Engineering (CAiSE '98), Pisa, Italy (1998).
[26] [26] M.A. Jeusfeld, C. Quix, M. Jarke: Design and Analysis of Quality Information for Data Warehouses. In Proc. of the 17th Intl. Conf. on Conceptual Modeling (ER'98), pp. 349-362, Singapore (1998).