International Journal of Computational Intelligence and Informatics, Vol. 2: No. 4, January - March 2013

An Improved Enhancement of Decision- making Analysis to increase the sales promotion using Hyper ETL in Data Mart

A Prema
Department of Computer Science
Bharathiar University
Coimbatore, Tamilnadu, India
Email: latharaman2012jr@gmail.com

A Pethalakshmi
Department of Computer Science
MVM Government Arts College for Women
Dindigul, Tamilnadu, India
Email: pethalakshmi@yahoo.com

***Abstract*- The multiplication in the number of corporations looking for data mart solutions, with the aim of adding major business gains, has created the need for a decision-aid has come near in preferring the right data mart system. Due to the indistinct concept often represented in decision-making procedure, to facilitate decision matrix analysis, with consideration given to both technical and managerial criteria. This paper illustrates the item- wise and place- wise analysis of sales promotion in sales data mart using Hyper ETL (Extract, Transform and Load) tool. We have used Laplacian method for ranking, which enables to make efficient decision- making in order to increase the sales promotion. The main objective of this paper is to determine the alternative courses of action (the movement of sales quantity and also the movement particular item in number of places) from which the ultimate choice to be made. This approach supports the business goals and requirements of an organization and to identify the appropriate attributes or criteria for evaluation. This improvement communicates information quickly and enlarges the aggregation process and helps to take an efficient decision.**

Keywords- Hyper ETL, Data Mart, Decision making, Decision Matrix, Decision analysis, aggregation

I. INTRODUCTION

Data mart conquers different troubles that result from the require to connect large numbers of decision support systems to large numbers of operational Data source systems .Many managerial decisions, however, are made with some uncertainty. Managers, for example, authorize substantial finical investments with less than complete information about product demand. As the decision taken by a manager governs the fortunes of business, right decisions will have a salutary effect while the wrong ones may prove to be disastrous, it is extremely important to choose the appropriate decision. Decision theory provides a rational approach to the managers in dealing with problems confronted with partial, imperfect or uncertain future conditions.

Under conditions of uncertainty, the decision maker has knowledge about the states of nature that happen but lacks the knowledge about the probabilities the source of their occurrence. Situations like launching a new product fall under this category. The process with insufficient data, leads to a more complex decision model and, perhaps, a less satisfactory solution. However, one uses scientific methods to exploit the available data to the fullest extents. Under conditions of uncertainty, a few decision criteria which are available could be of help to the decision maker and a choice among them is determined by the company's policy and attitude of the decision maker. In Laplace based method, the weight of each criterion and rating of each alternative are described using linguistic terms.

This paper proposes the design of Hyper ETL, which eliminates the mismanagement of metadata structure in data mart. But a data mart collects data on a specific subject area such as sales or production or accounts or Human Resource management or customer information. It can be a subset of company data warehouse and it is proposed to meet the desires of a single department. An operational data store is an updatable set of integrated data used for enterprise-wide strategic decision making. It consists of live data, not snapshots, and has least history retained. The most significant profit of data mart is a single view of data, whose result is more accurate information and it enables better decision making. A data mart is a persistent physical store of operational and aggregated data statistically processed data that supports business people in making decisions based primarily on analyses of past activities and results. A data mart contains a predefined subset of enterprise data organized for rapid analysis and reporting. This ETL Tool is used to simplify the process of migrating data, standardize the method of data migration, store all data transformation logic as Meta data which enable the users, managers and architects to understand, review, and modify the various interfaces and to reduce the cost and effort associated

with building interfaces. Extraction is the process of reading data from a specified source database and extracting a desired subset of data. Transformation phase applies a chain of rules or functions to the extracted data to derive the data to be loaded. Three forms of transformations are utilized, that is subsets of tables, formatting data and primary keys and indexes. Subsets are created to remove personally individual information. All tables except the reference table are transferred to the Data warehouses using an ETL process.

The ETL process includes designing a target, mapping sources to target, extracting data from sources, transforming, scheduling and monitoring processes, and managing the overall BI environment. The purpose of using ETL Tools is to save time and make the whole process more consistent. The ETL tools are customized to provide the functionality to meet the enterprise requirements. Hence, many of them choose to build their own data warehouse themselves. In this paper, we have proposed hyper ETL in order to improve the decision making process in sales data mart.

Section 2 of this paper deals with related work done in Data Mart, Extract, Transformation and Loading. Section 3, explains an actual process of Extract, Transform and Load, Section 4 explains the proposed work, in section 5, Experimental analysis, results are given, and finally, section 6 presents a conclusion of this paper.

II. RELATED WORKS

Data Mart can hold information which addresses both strategic and tactical information needs and provides information which allows key operating functions to effectively manage performance. It unifies information from various databases into a single database. A data mart contains data from a particular business area and multiple data marts can form a data warehouse. Data marts are the cornerstones of the enterprise, and each unique knowledge data mart is maintained by the divisional or departmental group. The motives for building a data mart are specified below [14].

- Easy access to frequently needed data
- Potential users are more clearly defined than in a full Data warehouse.
- Improves end-user response time
- Creates collective view by a group of users
- Provides ease of creation
- Lower cost than implementing a full Data warehouse

Over the years, data warehouse technology has been used for analysis and decision making in enterprises [4]. Different varieties of approaches for the integration of ETL tool in data warehouses have been proposed. Shaker H. Ali El-Sappagh tried to navigate through the efforts done to conceptualize abbreviations for ETL, DW, DM, OLAP, on-line analytical processing, DS, ODS, and DSA [9]. A data warehouse gives a set of numeric values that are based on a set of input values in the form of dimensions [6]. In automated data warehouse testing, the emphasis is the validation of data integrity between all points of comparison to ensure the proper implementation of the ETL mapping and transformations across the architecture.

Li Jian, conquered the weak points of traditional Extract, Transform and Load tool's architecture and proposed a three layers architecture based on metadata. That built ETL process more flexible, multipurpose and efficient and finally they designed and implemented a new ETL tool for the drilling data warehouse. A systematic review method was proposed to identify, extract and analyze the main proposals on modeling conceptual ETL processes for Data Warehouses. The main proposals were identified and compared based on the features, activities and notation of ETL processes and concluded the study by reflecting on the approaches being studied and providing an updated skeleton for future study [7].

A concrete ETL service framework was proposed and talked about metadata management service, metadata definition services, ETL transformation rules service, process definition service etc [3]. Two heuristic algorithms with greedy characteristics were proposed to reduce the execution cost of an ETL workflow [10]. Lunan Li, recommended intensively manage ETL by metadata repository and makes metadata easier to understand; therefore metadata management becomes more direct, simple, and centered. Sabir Asadullaev talked about centralized Extract, Transform and Load with similar Data warehouse and Data Mart, applications of data mart, data warehouse with integration bus and recommended data warehouse architecture[8].Numeric values of a classical data warehouse can be difficult to understand for business users, or may be interpreted incorrectly. Therefore, for a more accurate interpretation of numeric values, business users require an interpretation in meaningful non-numeric terms. However, if the transition between terms is crisp, true values cannot be measured and smooth, transition between classes cannot take place [1]. At last, definition method and related algorithms of ETL rules are designed and analyzed. A data mart contains data from a particular business area and multiple data marts can form a data warehouse [5]. ETL is a authoritative meta data based process that extract the data from source system and load into the data warehouse and this process improve overall data quality and report ability.

Radhakrishna and Sreekanth, proposed a web based framework model for representing the extraction of data from one or more data sources and use transformation business logic and load the data within the data ware house. This is a good starting point for gathering information in the existing documentation for the system and also research for ETL phase in web based scenario modeling in distributed environment a provide the effective decision results for various organization [23]. The models of the entire ETL process using UML because these structural and dynamic properties of an information system at the conceptual level are more natural than the naive approaches. It is more flexible and is used to support trading corporation, banks, financial and human resource management system of an organization at various levels. The future direction of this paper includes analyzing multimedia information sources automating mechanisms for ETL process.

Owen Kaser et al., "The Lito project data ware houses with Literature "describes to apply the business intelligence techniques of data warehousing and OLAP to the domain of text processing. A literary data ware-house is the conventional corpus but its data stored and organized in multidimensional stages, in order to promote efficient end user queries. This work improves the query engine, ETC process and the user interfaces. The extract, transform, load stage retains the information which are build by the data ware house. We believe the overall idea of applying OLAP to literary data is promising. The initial custom engine is slow for production use but until more optimization is attempted, its promise is unclear [22].

James F. Brule's "Fuzzy Systems – tutorial" demonstrates that the fuzzy system is an alternative to traditional notions of set membership and logic that has its origin in ancient Greek philosophy and its applications are the leading edge of artificial intelligence and it presents the foundation of fuzzy systems with formal mathematics [18]. It is used in many applications such as information retrieval system, a navigation system for automatic cars, a predictive fuzzy logic controller for automatic operation of trains, laboratory water level controllers for ROBOT arc welders, feature definition controllers for ROBOT vision, graphics controller for automated police sketchers and more. Fuzzy systems including fuzzy logic and fuzzy set theory provide a rich and meaningful addition to standard logic. The mathematics generated by theories is consistent; a fuzzy logic may be a generalization of classic logic. Many systems may be modeled and event replicated with the help of fuzzy systems.

Lior sapir et al., This paper "A methodology for the design of a fuzzy data warehouse" a data ware house is a special database used for storing business oriented information for future analysis and decision making. In business scenario, where some of the data or the business attributes are fuzzy, it may be useful to construct a ware house that can support the analysis of fuzzy data and also outlined the Kimball's methodology for the design of a data ware house can be extended to the construction of a fuzzy data ware house. A case study demonstrates the visibility of the methodology most commonly used methodology today is Kimball's. It describes the process of translating business data and process into a dimensional model. It has several advantages, such as users can make more intuitive and easy to understand queries in a natural language. Defining fuzzy dimensions allows the user to describe the facts with abstract human concepts which are actually more realistic [21]. The fuzzy dimensions also allow more flexible and interesting to filtering of the facts. We have demonstrated that fuzzy measures used with fuzzy aggregation operators allow the user to understand his business and the data warehouse measures better.

Daniel Fasel demonstrates the uses a fuzzy data ware house approach to support the fuzzy analysis of the customer performance measurement. The potential of the fuzzy data warehouse approach is illustrated using a concrete example of customer performance measurement of a hearing instrument manufacturer. A few for combining fuzzy concepts with the hierarchies of data ware house have been proposed. A method of summary can be guaranteed using this approach and the data ware house concepts retained flexibility. Using a fuzzy approach in data ware house concepts improves information quality for the company. It provides broader possibilities to create indicators for customer performance measurement as in the example given of a hearing instrument manufacturer. The proposed approach does not include fuzzy linguistic concepts directly in to the hierarchical structure of dimension or into fact tables of the data ware house model and also explains how the fuzzy concepts can be aggregated over dimensions without having to redefined the fuzzy sets in every degree of granularity[17]. Visualization should provide easily understand the results for fuzzy queries in the fuzzy data ware house.

D. Ashok Kumar and M.C. Loraine explained modern electronic health records are designed to capture and render vast quantities of clinical data during the health care prone. Utilization of data analysis and data mining methods medicine and health care is sparse. Medical data is one of the heavily and categorical type data. A Dichotomous variable is type of categorical variable which is binary with categories zero and one. Binary data are the simplest form of data used for medical database in which close ended questions can be used. It is very efficient based on computational efficiency and memory capacity to represent categorical type data. Data mining technique called clustering is involved here for dichotomous medical data due to its high dimensional and data scarcity. Usually the binary data clustering is done by using 0 and 1 as numerical value. The clustering is performed after transforming the binary data into real by wiener transformation. The proposed algorithm in this paper can be usable for large medical and health binary data bases for determining the correction are the health disorders and symptoms observed [16].

Kari Richardon and Eric Rossland described the hands-on work shop will give users a basic tour through the functionality of SAS ETL studio health to build a small data mart. The participants in this workshop will use SAS ETL studio to define necessary library definitions also source and target table definitions. Participants will create a process flow diagram using a simple transformation and load the target table. In the last step, participants will create 2 reports using target table[20]. Finally, this hands-on workshop provides an overview of SAS ETL studio and how it can be used to create a data mart.

Christ Sophie et al., focus that in the field of human resources there is a growing trend towards moving from activity based functions to a more strategic, business oriented role. The data mart defined on the HR information needs is the best solution to meet the objectives [15]. The main purpose of this paper is to explain how the SAS system can be used in top of SAP R/3 HR, and obtained real business benefits on a very short time. It is also based on the practical experience at the Belgian Gas and electricity provider. The structure of this paper first explained the business functions that are covered shortcomings of the system are pointed out. Next explained the solution of the short comings and discussed the business objectives for the data mart. Finally this paper explains the project approach and focuses on the specific attention points when building a data mart. It provides end to end solution and the data management facilities possible to deliver quick results to the end-users.

Jeremy Andeas et al., described in this paper about building powerful data marts that requires minimal administration and are simple to change. This may seem like an impossible goal to anyone who has been involved in the usual complexity but there are a number of simple, practical concepts and methodologies that have been employed and tested over many years of successful data ware house implementation that are repeatable and are easy to understand [19]. For the purposes of data ware housing ETL is used to pull data from business systems into a database that is designed for analysis and reporting. Building data mart and ETL Processes involves large volumes of complex business data and the easiest outcome is complexity. Lack of results are expected the easiest outcome of the more resources. It is also used to achieve powerful results in a short amount of time that is useful to users and fulfills the core requirement of effective visibility in to their complex business data.

III. EXTRACT, TRANSFORM AND LOAD (ETL)

A detailed explanation of the ETL process is Extract, Transform and Load (ETL)" The three data base functions are united into one tool that automate the process to drag data out of one database into another database.ETL is the process to allow business to combine their data while moving it from source system to data warehouse. Data can be taken from any source. Extraction is referred as extracting the data from various heterogeneous systems. Transform means applying the business rules on data which are derived from different sources. The process of pumping the data into the data warehouse for end user access is referred as loading [2]. Figure 1 shows the outlook of data warehouse. The Testing of ETL mainly deals with how, when, from where and what we carry in our data base.

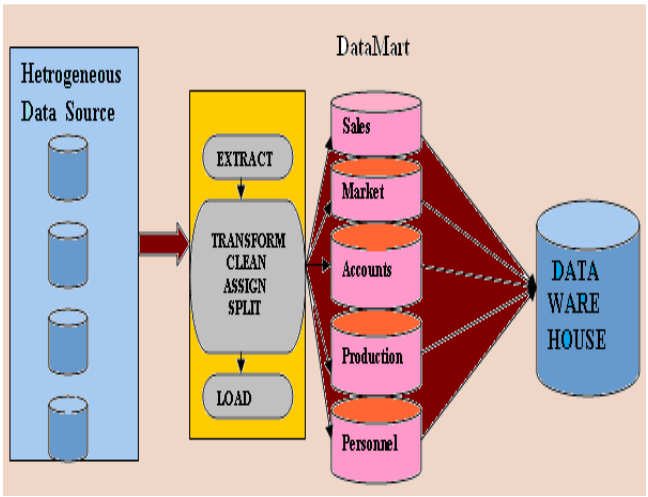

Figure 1: Schematic outlook of data warehouse

The star schema is the simplest data ware house schema looks like entity relationship model with points baking from a central table. The center of the star contains number of fact table. Snow-Flake Schema is a complicated data warehouse schema than a star schema. This structure resembles a snowflake. This schema normalizes dimensions data by grouping data into multiple tables rather than the one giant table. All tables except the reference table are transferred to the Data warehouse using an ETL process]. Many of the tables are split into smaller tables in order to expedite queries. The ETL process [11] includes designing a target, mapping sources to target, extracting data from sources, transforming data for the target, scheduling and monitoring processes, and managing the overall Business Intelligence environment.

Benefits of an ETL [12] Tools are given below:

- To Store all data transformation logic/rules as Meta data
- To Enable Users, Managers and architects to understand, review, and modify the various interfaces.
- To simplify the process of migrating data
- To reduce cost and effort associated with building interfaces
- To standardize the method of data migration

The ETL tools [13] were created to improve and facilitate data warehousing. ETL eliminates the step of loading the text files into intermediate storage, saving significant space and time. The ETL process consists of the following steps: 1. Initiation 2. Build reference data 3. Extract from sources 4. Validate 5. Transform 6. Load into stages tables 7. Audit reports 8. Publish 9. Archive10 Clean up. Transformation has been applied to achieve migrating from one database to another Data warehouse. End users can access the data via several methods i.e JDBC, ODBC, OLE, etc. The key factors underlying the main problems of ETL processes are vastness of the data volumes, quality problems, since data is not always clean and has to be cleansed, performance, since the whole process has to take place within a specific time window and evolution of the sources and the data warehouse can eventually lead, even to daily maintenance operations.

IV. PROPOSED WORK

This paper, presents Hyper ETL for data mart to get better performance. To design the Hyper ETL we exercised XML, Oracle and java. Here, ETL rules are designed and analyzed to remove the mismanagement of metadata in ETL processes and also improve the ETL process. This paper integrate the Hyper ETL and decision matrix analysis concept to reduces the storage space, time, cost and also this hyper ETL increase the throughput than existing one. Sales data are taken into the account for implementing this hyper ETL. Based on the Laplace criterion, equal probability values applied to each places and its corresponding items. The Process of Hyper Extract, Transform and Load tool in sales data mart is shown in figure 2.

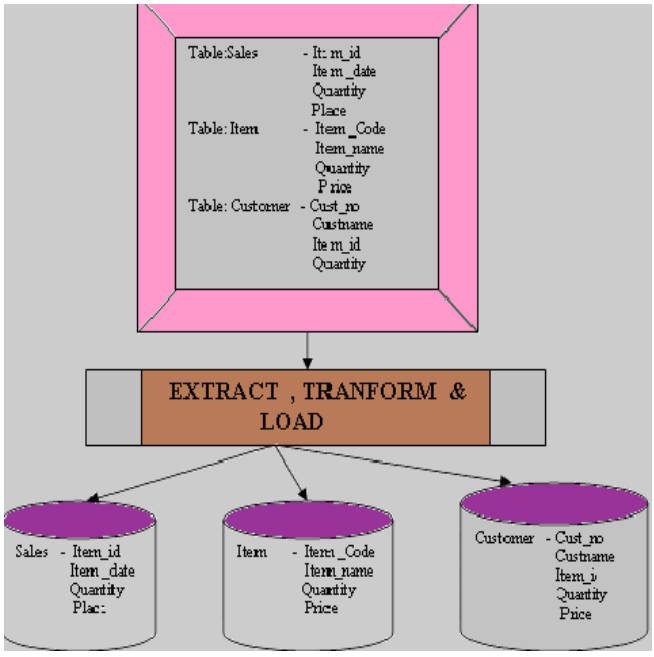

Figure 2: Hyper ETL Diagram

Table 1: Decision analysis matrix of sales

Places	Weight	I1		I2		I3		I4		I5		I6		Total
		Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	
P1	0.17	0.3	0.051	0.2	0.034	0.1	0.017	0.4	0.068	0.5	0.085	0.6	0.102	0.357
P2	0.17	0.1	0.017	0.4	0.068	0.2	0.034	0.3	0.051	0.6	0.102	0.5	0.085	0.357
P3	0.17	0.2	0.034	0.3	0.051	0.3	0.051	0.1	0.017	0.4	0.068	0.5	0.085	0.306
P4	0.17	0.4	0.068	0.1	0.017	0.4	0.068	0.2	0.034	0.3	0.051	0.1	0.017	0.255
P5	0.17	0.3	0.051	0.2	0.034	0.5	0.085	0.1	0.017	0.4	0.068	0.6	0.102	0.357
P6	0.17	0.5	0.085	0.5	0.085	0.6	0.102	0.3	0.051	0.2	0.034	0.1	0.017	0.374
Total		0.306		0.289		0.357		0.238		0.408		0.408		
RANK		4		5		3		6		1		1		

Algorithm for Decision Matrix analysis

Step 1: Take a sales table which is having places in Row and Items in column

Step 2: Assign equal probabilities to each place in a sales table.

Step 3: Multiply each Item quantity rating by the weight value. This gives score for that Item.

Step 4: Sum up the scores. This goes in the total, including both the column and the row total

Step 5: Repeat the above steps for every Item in the sales table.

Step 6: Put the rank by ordering the descending values of total score.

Step 7: Apply the Histogram analysis based on the total score.

Step 8: With the help of the Histogram analysis acquire the best effective decision for sales promotion.

V. EXPERIMENTAL ANALYSIS AND RESULT

We have proposed a Hyper ETL with decision matrix analysis for increasing an efficiency of ETL process in Local data warehouse (i.e., Data Mart) using decision matrix analysis. Experiments were conducted in XML, java and Oracle language. The worth of this Hyper ETL tool was checked through some sample sales records and we located that, it obtained 58 minutes for nearly one million records for which transformation time was fewer than the existing ETL tool. We used Intel® Xeon® E5-4600 Series Processor, 16GB DDR-III RAM to analyze this proposed Hyper ETL for sales data mart.

This test is put into operation by using Operation Research technique (Transportation problem). Here, the cost of an item is taken into the account. Cost of product may vary from place to place. The result of transportation problem shows that, the transportation cost is lower than transportation cost of sales data mart which used an existing ETL Tool.

The results of decision analysis are given in the form of histogram analysis.

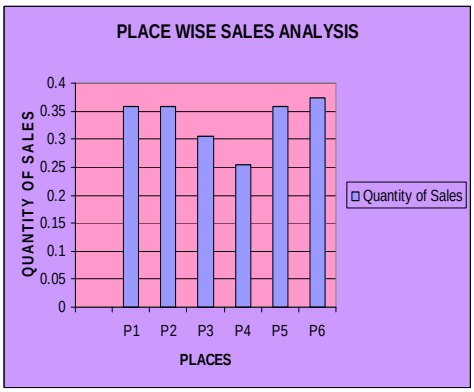

Figure 3: Placewise sales analysis

This place wise analysis shows that place4 has low sales rate than other places, place3 has a moderate sales rate and p6 and p1 has high sales rate.

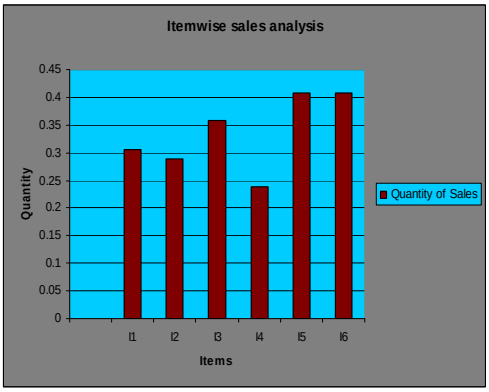

Figure 4: Itemwise sales analysis

The item wise analysis shows the demands of each item in all places. According to the analysis, item5 and item6 have high demand in all places, item1and item2 have moderate demand and item4 has low demand in all places.

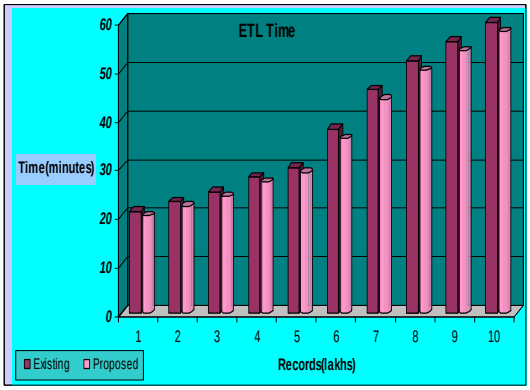

Figure 5: Transformation time of existing ETL and Hyper ETL

The result of transportation problem is given in the form of graph.

Figure 6: Transportation cost of Sales

The outcome of transportation problem shows that, the transportation cost which used an existing ETL tool is lesser than transportation cost of sales data mart using Hyper ETL Tool.

VI. CONCLUSION

We have suggested the sophisticated structural design of Hyper ETL and decision matrix analysis which achieves better performance of ETL Process through reduces the time, space and improves the decision-making process. This decision analysis matrix suggested the optimal solution by applying the Laplace's method and this holds great potential for dramatic business benefits and also provides decision makers access to consistent, reliable and timely data still much more to explore and enhance an efficient decision making to hike the sales promotion. The job obtainable in this paper is planned at discover an effective decision making to increase the sales promotion .The work presented in this paper is intended to get better performance of ETL process exploring an effective decision making.

REFERENCES

[1] D. Fasel and D. Zumstein., "A fuzzy data warehouse approach for web analytics", In MD. Lytras, E.Damiani, J. M. Carroll, R. D. Tennyson, D.Avison, A. Naeve, A. Dale,P. Lefrere,F. Tan, J. Sipior, and G. Vossen, editors, Visioning and Engineering the Knowledge Society – A Web Science Perspective, volume 5736 of Lecture Notes in Computer Science, pages 276–285. Springer, 2009.

[2] Master Data Management An Oracle White Paper September 2011.

[3] Munoz L, Mazon J, Trujillo J, "Systematic review and comparison of modeling ETL processes in data warehouse", Iberian Conference on information Systems and Technologies, June 2010.

[4] Sabir Asadullaev, "Data Warehouse Architectures", III SWG IBM EE/A 03.11.2009.

[5] Shaker H. Ali El-Sappagh a, Abdeltawab M. Ahmed Hendawi b, Ali Hamed El Bastawissy b, "A proposed model for data warehouse ETL processes", Journal of King Saud University – Computer and Information Sciences (2011), 91–104.

[6] Simitsis, A Vassiliadis, Sellis T, "State-space optimization of ETL workflows",IEEE Transactions on Knowledge and Data Engineering, Vol 17, Issue 10, Oct 2005.

[7] Inmon, William, "Data Mart Does Not Equal Data Warehouse", DMReview.com.

[8] Hariprasad T, "ETL testing Fundamentals", on March 29, 2012.

[9] Huamin Wang, "An ETL Services Framework Based on Metadata", 2nd International Workshop on Intelligent Systems and Applications, May 2010.

[10] W. H. Inmon., "Building the Data Warehouse", Wiley Publishing, Inc., 4 edition, 2005.

[11] Inmon, William (2000-07-18), "Data Mart Does Not Equal Data Warehouse", DMReview.com.

[12] Jeffrey R. Bocarsly, "Complex ETL Testing-A Strategic Approach".

[13] R. Kimball and M. Ross, "The Data Warehouse Toolkit", WileyPublishing, Inc., 2002.

[14] Li Jian, Xu Bihua, "ETL tool research and implementation based on drilling data warehouse", Seventh International Conference on Fuzzy Systems and Knowledge Discovery, Aug 2010.

[15] Lunan Li, "A framework study of ETL processes optimization based on metadata repository", International Conference on Computer Engineering and Technology, April 2010.

[16] D.Ashok Kumar and M.C. Loraine Charlet Annie,"Decision Making on Dichotomous Medical Data using Novel Clustering approach", National conference on Signal and Image processing(NCSIP) 2012.

[17] Daniel Fasel, "A fuzzy datawarehouse approach for the customer performance measurement for a hearing instrument manufacturing company", Sixth International conference on fuzzy systems and knowledge discovery,2009.

[18] J.F.Baldwin, "Fuzzy systems – logic and reasoning in fuzzy applications", London: Academic press,1981.

[19] Jeremy, Jean King and Andreas schindler, "Understanding of Business Intelligence: ETL and Data mart Practises".

[20] Kari Richardson and Eric Rossland, "A Quick Tour of SAS ETL to build a data mart",pp.131-30.

[21] Lior Sapir and Armin Shmilovice., "A methodology for the design of a fuzzy data warehouse", 2008.

[22] Owen kaser, Steven Keith and Daniel Lomire , "Dataware housing with literature", September 11,2006.

[23] Radha krishna and Sree kanth, "An Object Oriented modeling and Implementation of Web based ETL process", in IJCSNS, International Journal of Computer Science and Network Security, vol 10 no.2, February 2010.