RESEARCH ARTICLE OPEN ACCESS

Survey on Intelligent Data Repository Using Soft Computing

A. Prema [1], A.Pethalakshmi [2]
Research Scholar [1]
Research and Development Centre
Bharathiar University, Coimbatore
Associate professor [2]
MVM Government Arts College, Dindigul
Tamil Nadu – India

ABSTRACT
Data warehouse is one of the components of the overall business intelligence system. An enterprise has one data warehouse, and data marts source has their information from the data warehouse. The Data warehouse is a corporation of all data marts within the enterprise. Information is always accumulated in the dimensional model. In this paper, an intelligent data repository with soft computing is presented. It covers similarity metrics that are commonly used to improve the efficiency of data storages. It also covers multiple decision making methodologies to improve the efficiency of decision making. This chapter focuses on the review of the literature for Extraction, Transform and Load with Data Warehouse. Moreover the ETL hybridization with fuzzy optimization, Markov Decision model, Decision making criteria and Decision Matrix has also been reviewed. The Decision Matrix is a mathematical tool to deal with uncertainty and vagueness of decision systems. It has been applied successfully in all fields. This paper proposes Hyper ETL with an integration of decision making methodologies and fuzzy optimistic technique.

Keywords:- Hyper ETL, Data Mart, Data warehouse, Decision making Methodologies, Fuzzy optimization.

I. ETL DATA MART AND DATA WAREHOUSE

Data Warehousing has been around for twenty years and has become the part of the information technology infrastructure. Data warehouse originally grew in response to the corporate need for information not data and it supplies integrated, granular, and historical data to the corporation. The benefit of this is that people who are building or using a data warehouse can see what lies ahead, and can determine [21].In modern business, vast amount of data are accumulated, which complicates the decision making process. How to change the existing situation of "mass data, poor knowledge", support better business decision making and help enterprises increase profits and market share become the business and IT sector issues of mutual concern. Business intelligence technologies were emerged as the times require them. ETL plays an important role in BI project, which realizes the technical service and decision making support. An overview of ETL, the main module of ETL, the optimization scheme of ETL, as well as the specific implementation of the ETL process are included by Tang Jun[86]. PanosVassiliadis and Alkis Simits is highlighted Extraction, Transformation, and Loading (ETL) processes which are responsible for the operations taking place in the background of data warehouse architecture. In a high level description of an ETL process, first, the data are extracted from the source data stores that can be on-line transaction processing (OLTP) or legacy systems, files under any format, webpages, various kinds of documents (e.g., spreadsheets and text documents)or even data coming in a streaming fashion. Typically, only the data that are different from the previous execution of an ETL process(newly inserted, updated, and deleted information) should be extracted from the sources. Secondly, the extracted data are propagated to a special-purpose area of the warehouse, called the data staging area (DSA), where their transformation, homogenization, and cleansing take place the most frequently used being transformation[54].

Gregory S. Nelson et al. explained the methodology used to design the target database structure and transformations, create a mapping worksheet used to implement the ETL code, load the metadata, and create the process flows in Data Integration (DI) Studio. The paper further connects the dots for those interested in getting started with DI Studio not only as a tool, but also how practitioners think about the DI Studio process [15].Table 1 summarizes the different approaches with Data Mart and Data ware house.

Table 1:ETL process with Data Mart and Data warehouse.

Auhtor(s)	Purpose(s)	Description(s)
Inmon, William, 2000[22]	ETL, data ware house	The ETL procedure consists of designing a target, transforming data for the target, scheduling and monitoring processes. The reason for using ETL tools is to save time and make the whole process more consistent. The ETL tools were customized to provide the functionality to meet the enterprise
Simitsis, A Vassiliadis, P. Sellis, T. 2005.[76]	Data ware house	A data warehouse gives a set of numeric values that are based on set of input values in the form of dimensions
W. H. Inmon. 2005.[23]	ETL Process	Two heuristic algorithms with greedy characteristics were proposed to reduce the execution cost of an ETL workflow
Tec Ying Wah, Ng Hooi Peng,and Ching Sue Hok,2007[89]	ETL and Data warehouse	An attempt had been made to bring out a systematic process of crawl for only the data that the users need to insert into database instead of simply crawling all the data without planning and organizing the data structure for it. Building a data warehouse for library is an iterative process as the library data warehouse will be growing and evolving
Gregory S. Nelson et al, 2007, [15]	ETL	Explained the methodology used to design the target database structure and transformations, create a mapping worksheet used to implement the ETL code, load the metadata, and create the process flows in Data Integration (DI) Studio.
William H. Inmon, Derek Strauss and, Genia Neushloss,2008	Data Warehouse	Data Warehousing has been around for 20 years and has become part of the information technology infrastructure. Data warehousing originally grew in response to the corporate need for information.
Sabir asadullaev , 2009[71]	Centralized ETL With parallel Data warehouse	Discussed the advantages and limitations of the following architectures: centralized ETL with parallel DW and data marts, with intermediate application data marts, data warehouse with integration bus and recommended EDW architecture
Tang Jun, Feng Yu 2009[86]	ETL with Data warehouse	In modern business, vast amount of data are accumulated, which complicates the decision making process. How to change the existing situation of "mass data, poor knowledge", support better business decision making and help enterprises increase profits and market share become the business and IT sector issues of mutual concern. ETL plays an important role in BI project, which realizes the technical service and decision making support.
Panos Vassiliadis and AlkisSimitsis, 2009[54]	ETL	In a high level description of an ETL process, first, the data were extracted from the source data stores that can be on-line transaction processing (OLTP) or legacy systems, files under any format, web pages, various kinds of documents(e.g., spreadsheets and text documents) or even data coming in a streaming fashion.
D. Fasel and D. Zumstein, 2009[13]	ETL	Method and related algorithms of ETL rules were designed and analyzed.
Teori kontra praktik Ann Josefsson & IsabelZitoun, 2010[90]	ETL	Examined the theory behind the ETL process and subsequently investigate how it may be applied by comparing the theory and how the company knows it.
Huamin Wang, 2010[19]	ETL	Different kinds of approaches for the integration of ETL tool in data warehouses had been proposed.

Table 1: ETL process with Data Mart and Data warehouse (cont.)

Auhtor(s)	Purpose(s)	Description(s)
Radha Krishnan and Sree Kanth,2010[64]	ETL, data ware house	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 warehouse. This is the good starting point for gathering information in the exiting documentation for the system and also researching for ETL phase in web based scenario modeling in the distributed environment which provides an effective decision results
Master Data Management An 2011[44].	ETL and data warehouse	Extract, Transform and Load (ETL) is a process that involves extracting data from produce source. It has been transforming it through encoded business rules to fit business needs, and loading it into the data warehouse from where reports are

Shaker H. Ali El-Sappagh, Abdeltawab M. Ahmed Hendawi, Ali Hamed El Bastawissy,2011[75]	ETL	This problem represented a real need to find a standard conceptual model for representing the simplified way for the extraction, transformation, and loading (ETL) processes. Some approaches have been introduced to handle this problem.
Hariprasad T, 2012[18].	ETL , Data Mart	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
Stephen Overton, 2012[79]	ETL	Presented a flexible change data capture process to extract and load new data during any phase of loading a data warehouse. The process can run dynamically at any time and requires no set schedule. This paper demonstrates a data retention process using
Nitin Anand,2012[50]	ETL	Discussed an important part of BI systems which is a well performing implementation of the Extract, Transform, and Load (ETL) process and in typical BI projects, implementing the ETL process can be the task with the greatest effort.
Osama E.Sheta and Ahmed Nour Eldeen, 2013[51]	Data warehouse	Described the technology of data warehouse in healthcare decision-making and tools for support of these technologies, which are used for cancer diseases. The healthcare executive managers and doctors need information about and insight into the existing health data, so as to make decision more efficiently without interrupting the daily
S. Saagari, P. Devi Anusha, Ch. Lakshmi Priyanka, V. S. S. N. Sailaja, 2013[70]	Data warehouse	Presented an overview of Data warehousing, Data Mining, OLAP, OLTP technologies, exploring the features, applications and the architecture of Data Warehousing. The data warehouse supports on-line analytical processing (OLAP), the functional and performance requirements of which are quite different from those of the on-line transaction processing (OLTP) applications traditionally supported by the operational databases
K. Srikanth et al, 2013[78]	Data warehouse	Presented the information about a previous value of a dimension that is written into the database for SCD (Slowly Changing Dimensions) type 3. In this article, the authors discussed the step by step implementation of SCD Type 3 using Informatica Power Center. The number of records stored in SCD Type 3 does not increase exponentially as they do not insert a record for each and
A.Prema and A.Pethalakshmi 2013[60]	ETL	Discussed the Improved decision making using novel ETL by mapping the multiple sources into multiple Targets and eliminate the duplicate fields from the table.
A.Prema and A.Pethalakshmi, 2013[61]	Hyper ETL	Demonstrated the comparative analysis of ETL and Hyper ETL Hyper ETL tool broadens the aggregation method, conveys information intelligently and is useful for an effective decision making. ETL rules are designed to eliminate the
A.Prema and A.Pethalakshmi , 2013[59]	HyperETL and Data warehouse	Presented the refined design of Hyper ETL which accomplishes enhances show of ETL, through reducing the data transformation time and cost and improves the throughput and amalgamate the contribution of enhanced Hyper ETL Tool with decision analysis methodologies

Osama E.Sheta et al. described the technology of data warehouse in healthcare decision-making and tools for support of these technologies, which are used for cancer diseases. The healthcare executive managers and doctors need information about and insight into the existing health data, so as to make decision more efficiently without interrupting the daily work of an On-Line Transaction Processing(OLTP) system. This is a complex problem during the healthcare decision-making process. To solve this problem, building a healthcare data warehouse seems to be efficient. The authors explain the concepts of the data warehouse, On-Line Analysis Processing (OLAP). Changing the data in the data warehouse into a multidimensional data cube is then shown. Finally, an application example is given to illustrate the use of the healthcare data warehouse specific to cancer diseases developed in this study. The executive managers and doctors can view data from more than one perspective with reduced query time, thus making decisions faster and more comprehensive [51].

Tec Ying Wah, et al described steps in the development of library data warehouse especially extracting data, transforming data and loading data into database. Due to complexity of data, more time is spent in these tasks. In order to reduce the time consumed, an attempt has been made to bring out a systematic process of crawl for only the data that the users need to insert into database instead of simply crawling all the data without planning and organizing the data structure for it.
Building a data warehouse for library is an iterative process as the library data warehouse will be growing and evolving. Hence, flexibility and extendable issues are important as the author's framework will include this portable feature. The goal is to produce a framework that simplifies the process of building a library data warehouse and shares knowledge and problems that are being faced due to reducing the work. Through this iterative process, the user needs to enhance the crawling and cleansing process in order to achieve consistency and guarantee for an updated data warehouse [89].
S. Saagari et al. presented an overview of Data warehousing, Data Mining, OLAP, OLTP technologies, exploring the features, applications and the architecture of Data Warehousing. The data warehouse supports on-line analytical processing (OLAP), the functional and performance requirements of which are quite different
from those of the on-line transaction processing (OLTP) applications traditionally supported by the operational databases. Data warehouses provide on-line analytical processing (OLAP) tools for the interactive analysis of multidimensional data of varied granularities, which facilitates effective data mining. Data warehousing and on-line analytical processing (OLAP) are essential elements of decision support, which has increasingly become a focus of the database industry. OLTP is customer-oriented and is used for transaction and query processing by clerks, clients and information technology professionals. An OLAP system is market-oriented and is used for data analysis by knowledge workers, including managers, executives and analysts. Data warehousing and
OLAP have emerged as leading technologies that facilitate data storage, organization and then, significant retrieval. Decision support places some rather different requirements on database technology compared to traditional on-line transaction processing applications [70].
Nitin Anand presented an important part of BI systems which is a well performing implementation of the Extract, Transform, and Load (ETL) process and in typical BI projects, implementing the ETL process can be the task with the greatest effort. He proposed the templates of set of generic meta model with a palette of frequently used ETL activities. [50]. What a data warehouse is and how the ETL process is used for data storage in the data warehouse are included in "Uppsala Universitet ETL-processen". The purpose of this paper is to examine the theory behind the ETL process and subsequently investigate how it may be applied by comparing the theory and how the company knows it[90].
K. Srikanth et al. described the information about a previous value of a dimension that is written into the database for SCD (Slowly Changing Dimensions) type 3. In this article, the authors discussed step by step implementation of SCD Type 3 using Informatica Power Center. The number of records stored in SCD Type3 does not increase exponentially as they do not insert a record for each and every historical record. Hence they might not need the performance improvement techniques used in the SCD Type 2 Tutorial. It is better to know more about SCDs at Slowly Changing Dimensions Concepts. The new incoming record replaces (changes/modifies data set) the existing old record in target. Comprehensive ETL criteria are identified, testing procedures are developed and this work is applied to commercial ETL tools. The study covers all major aspects of ETL usage and can be used to effectively compare and evaluate various ETL tools[78].
Stephen Overton presented a flexible change data capture process to extract and load new data during any phase of loading a data warehouse. The process can run dynamically at any time and requires no set schedule. This paper demonstrates a data retention process using Base SAS ®. Both processes are centrally managed and operate independent of each other[79].
Sabir asadullaev discussed the advantages and limitations of the following architectures: Centralized ETL with parallel DW and Data Marts, with intermediate application data marts, data warehouse with Integration bus and recommended EDW architecture. The importance of various approaches, methods and recommendations make a mess of concepts, advantages and drawbacks, limitations and applicability of specific architecture solutions. Recommended corporate data warehouse architecture allows creating a short time and with minimal investment a workable prototype that is useful for business uses. The key to this architecture that provides an evolutionary development of EDW which is the introduction of meta data and master data management systems at the early stage of development [71].
Sabir Asadullaev proposed a methodology for data warehouse design, when sources of data are

XML schemas and conforming XML documents in "A Tool for Data Warehouse Design from Xml Sources". A prototype tool has been developed to verify and support the methodology. The tool automations have many parts in the conceptual and logical design process. Thus it helps the designer in designing faster and more accurately.

In this paper the main features of the tool for data warehouse design from xml source are presented [43]. For many years, data warehouse technology has been used for analysis and decision making in the enterprises [71].

Shaker H. Ali El-Sappagh et al investigated a very important problem in the current research of data warehouse. This problem represents a real need to find a standard conceptual model for representing the simplified way for the extraction, transformation, and loading (ETL) processes. Some approaches have been introduced to handle this problem. These approaches have been classified into three categories: first one is modeling based on mapping expressions and guidelines, second one based on conceptual constructs, and the last one based on UML environment. Building a data warehouse requires focusing closely to understand three main areas: the source area, the destination area and the mapping area (ETL processes). The framework of ETL processes consist of data source part, data warehouse schema part, and mapping part. Both data sources and data warehouse schemas should be defined clearly before starting to draw EMD scenario. And also it is an attempt to navigate through the efforts done to conceptualize the ETL processes [75].

Extract, Transform and Load is a process that involves extracting data from produce source. It has been transforming it through encoded business rules to fit business needs, and loading it into the data warehouse from where reports are generated. One can customize the ETL jobs to suit one's specific business requirements. The three database functions are combined into one tool that automates the process to pull data out of one database into another database [44]. The ETL procedure consists of designing a target, transforming data for the target, scheduling and monitoring processes. The reason for 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 necessity. Hence, many of them choose to construct their own datawarehouse themselves[22,28 ,34].

Li Jain conquered the weak points of traditional Extract, Transform and Load tool's architecture and proposed a three-layer architecture based on metadata. They built ETL process more flexible, multipurpose and efficient and finally they designed and implemented with a new ETL tool for drilling the ware house. A systematic review method was proposed to identify, extract and analyze the main proposals on modeling conceptual ETL processes for Data Warehouse. The main proposals are identified and compared based on the features, activities and notation of ETL processes and the study is concluded by reflecting on the approaches being studied and providing an update skeleton for future study [22].

Sabir Asadullaev stressed 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 [18].

Different kinds of approaches for the integration of ETL tool in data warehouses had been proposed. Shaker H. Ali El- Sappagh tried to navigate through the effort done to conceptualize abbreviations for ETL, DW, DM, OLAP, on- line analytical processing, DS, ODS, and DSA[19]. A data warehouse gives a set of numeric values that are based on set of input values in the form of dimensions [76].

A concrete ETL service framework was proposed and talked about metadata management service, metadata definition service, ETL transformation rules service, process definition service etc [47]. Two heuristic algorithms with greedy characteristics were proposed to reduce the execution cost of an ETL workflow [23].

Lunan Li recommended to Intensively manage ETL by metadata repository and makes metadata easier to understand; therefore metadata management becomes more direct, simple and centered. Numeric values of a classical data warehouse can be difficult to understand for business users, or may be interpreted incorrectly. Therefore, for more accurate interpretation of numeric values, business users require an interpretation in meaningful non-numeric terms. However, if the transition between the terms is crisp, true values cannot be measured and smooth transition between classes cannot take place [13].

At last, definition method and related algorithms of ETL rules are designed and analyzed.

Radhakrishnan and Sreekanth proposed a web based framework model for representing the extraction of data from one or more data sources using transformation business logic and loading the data within the data warehouse. This is the good starting point for gathering information in the existing documentation for the system and also researching for ETL phase in web based scenario modeling in the distributed environment which provides effective decision results for various organizations.

The models of the entire ETL process use 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 it is used to support trading corporation, banks, finance and Human Resource Management System at various levels. The future direction of this paper includes analyzing multimedia information sources automating mechanisms for ETL process [64].

A data mart contains data from a particular business area and multiple data marts can form a data warehouse. ETL is an authoritative meta data based on process that extracts the data from source system and loads into the data warehouse and this process improves overall data quality and report ability [75].

Jeremy, Andeas et al., had built powerful data marts that require minimal administration and are simple to change. This may seem like an impossible goal to anyone who is involved in the usual complexity but there are number of simple, practical concepts and methodologies that have been employed and tested over many years, of successful data warehouse implementation that are repeatable and easy to understand [29].

Data Mart can hold information which addresses both strategic and tactical information needs and provides information which allows key operating function to manage effectively. It unifies information from various databases into a single database. Data marts are the cornerstones of an 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 [36].

a) Improves end- user response time
b) Creates collective view by a group of users
c) Provides ease of creation
d) Easy access for frequently need data
e) Lower cost than implementing a full Data warehouse

Data mart conquers different troubles that result from the requirements to connect from a large numbers of decision support systems to a large number of operational Data source systems including many managerial decisions. However they are made with some uncertainty. Managers, for example, authorize substantial, financial investments with less than complete information for product demand. As the decision taken by a manager who governs the fortunes of business, right decisions will have a salutary effect while the wrong one may be proved to be disastrous, it is extremely important to choose the appropriate decision. Moreover, Decision theory provides a rational approach to managers in dealing with problems confronted with partial, imperfect or uncertain future conditions. Under the conditions of uncertainty, the decision maker has knowledge about the states of nature that happen but the lack of knowledge brings about the probabilities of the source of their occurrences. Situations like launching a new product falls under this category. The process with insufficient data, leads to a more complex decision model perhaps, a less satisfactory solution. However, one uses scientific methods to exploit the available data to the fullest extents. Under the conditions of uncertainty, a few decision criteria which are available could be helpful 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 alternative are described by using the linguistic terms [57].

Steven Scherma et al. described the use of data marts. Data Ware housing concepts are used to expedite retrieval and display of Complex attribute data from multi-million record database. Los Alamos National Laboratory has developed an Internet Application (SMART) using ArcIMS that relies on data marts to quickly retrieve attribute data, but has not contained within GIS layers. The volume of data and the complex relationships within the transactional database make data display within ArcIMS; impractical without the use of data marts. The technical issues and solutions involved in the development are also discussed. It has been demonstrated that this approach integrates well into a GIS framework and can be used successfully on the web [80].

Christ Sophie et al., focused 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 [42]. This paper explained how the SAS system can be used on the top of SAP R/3 HR, and obtains real business benefits in 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 explains the business functions that cover shortcomings of the system. The solution to short comings is explained and business objectives for the data mart are discussed. 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 data management facilities possible to deliver quick results to the end users.

For the purposes of data ware housing, ETL is used to pull data from business system into a database that is designed for analysis and reporting. Building data mart and ETL process involves

large volumes of complex business data and the outcome is complexity. It is also used to achieve powerful results in a short span of time that is useful to users and fulfills the core requirement of effective visibility in to the complex business data. Fuzzy union and intersection are used to take optimal solution [32].

A.Prema et al. proposed an integrated Quick ETL engine with Markov analysis algorithm. Which eliminated the mismanagement of meta data structure in data mart and improves the movement of sales item to the right place for increasing the sales rate. The movement of items in a particular place is studied and the work presented in this paper is aimed at exploring an effective decision making to increase the sales promotion by Quick ETL Engine with Markov analysis decision making process[62].

A.Prema et al. analyzed the troubles of existing ETL tools, and compare the parameter of Hyper ETL with existing ETL. This Hyper ETL tool broadens the aggregation method, conveys information intelligently and is useful for an effective decision making.ETL rules are designed to eliminate the negligence of metadata in ETL processes and improve an effectiveness of an ETL process. This Hyper ETL reduced the transformation time, maintenance cost and increase the throughput and reliability than an existing one. presented the comparative study of Existing ETL and proposed Hyper ETL. They took about 15 essential parameters and we have given the difference of existing and proposed Hyper ETL. Based on the study, Scalability, CPU utilization, throughput, reliability, execution speed are high and maintenance cost is low than Existing ETL[61].

II. DATA MART, DATA WAREHOUSE AND FUZZY CONCEPT

This section reviews the perception of data ware house with Fuzzy logic concepts. Fuzzy logic is a form of many-valued logic; it deals with reasoning that is approximate rather than fixed and exact. Compared to traditional binary sets, fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false. Furthermore, when linguistic variables are used, these degrees may be managed by specific functions.

Lior Sapir et al outlines how Kimball's methodology for the design of a data warehouse can be extended to the construction of a fuzzy data warehouse. A case study demonstrates the viability of the methodology. A data warehouse is a special database used for storing business oriented information for future analysis and decision-making. In business scenarios, where some of the data or the business attributes are fuzzy, it may be useful to construct a warehouse that can support the analysis of fuzzy. The users can make more intuitive and easy to understand queries in a natural language[44].

Rohit Ananthakrishnal et al. developed an algorithm for eliminating duplicates in dimensional tables in a data warehouse, which are usually associated with hierarchies to increase high quality, scalable duplicate elimination algorithm, and evaluate it on real database from an operational data warehouse. The duplicate elimination problem of detecting multiple tuples, which describe the same real world entity, is an important data cleaning problem. The users exploit dimensional hierarchies in data warehouses to increase high quality, scalable, and efficient algorithm for detecting fuzzy duplicates in dimensional tables [67].

Fasel, D. and Shahzad, K presented a fuzzy data warehouse model facilitates smooth transition between classes, have been proposed. By using the fuzzy data warehouse model, data can be classified both fuzzily and sharply. Because of this, the FDWH supports qualitative and quantitative analyses without affecting the core data warehouse schema. In addition, querying can be done based on natural language through direct use of the terminologies of the fuzzy classifications. A fuzzy data ware-House (FDWH)modeling approach, which allows a Integration of fuzzy concepts without affecting the core of A DWH is presented. The use of the proposed approach is demonstrated by a retail company. Finally, a comparison of fuzzy and classical data Warehousing approaches is presented [12].

Table 2 summarizes different approaches of fuzzy logic with data warehouse.

Table 2: Different approaches of fuzzy logic with data warehouse.

Author(s)	Purpose(s)	Description(s)
Kankana Chakrabarty, Ranjit Biswas and Sudarsan Nanda[32]	Fuzzy,data ware house	A justification, such attempt was made with examples on real life problems. The occurrence of union/intersection of two fuzzy sets in two different universe is very common in many real life problems.
R. E. Bellman and L. A. Zadeh1970[3]	Fuzzy optimization	The study on the theory and methodology of the fuzzy optimization had been active since the concept of fuzzy decision and the decision model under fuzzy environments were proposed
H-J. Zimmermann, 1976[98]	Fuzzy Mathematical Programming	Symmetric approach is an important approach for Fuzzy Mathematical Programming. The word 'Symmetric' used here comes originally from the symmetric model by Zimmermann
J. F. Baldwin,1981[2]	Fuzzy system	Demonstrated that the fuzzy system is an alternative to traditional notions of set membership and logic that has had 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
H-J. Zimmermann,1985 and M. K. Luhandjula,1980[100]	Symmetric and Asymmetric	Classified the fuzzy Mathematical Programming into symmetric and asymmetric models and categorized the fuzzy mathematical programming into flexible programming, fuzzy stochastic programming and mathematical programming with the fuzzy coefficients.
M S Khan, M Quaddus, A Intrapairot3 and A Chongl,2000[33]	Fuzzy Cognitive Map	The process of building the FCM (Fuzzy Cognitive Map) for simulating the data warehouse diffusion scenario has been analyzed. The analyzed results are presented and compared with the corresponding results obtained by using the system dynamics methodology for modeling complex systems.
Dr. James F. Smith and Robert D. Rhyne,2000[26]	Fuzzy membership functions	Described scheduling of electronic attack, resources distributed over many platforms is also under this process. The functional form of the fuzzy membership functions for the root concepts that will be Selected heuristically and will generally carry one or more Free parameters
Rohit Ananthakrishnal Surajit Chaudhuri and Venkatesh Gant,2002[67]	Data Warehouse	Developed an algorithm for eliminating duplicates in dimensional tables in a data warehouse, which are usually associated with hierarchies to increase high quality, scalable duplicate elimination algorithm, and evaluate it on real database from an operational data warehouse. The users exploits dimensional hierarchies in data warehouses to increase high quality, scalable, and efficient algorithm for detecting fuzzy duplicates in dimensional tables
Tang Jiafu Wang Dingwei, Richard Y K Fung And Kai-Leung,2004[85]	Fuzzy optimization	Described an extensive study on fuzzy optimization, which leads to the following concluding remarks that the basic procedure of fuzzy optimization problems is to transform a fuzzy model in to a crisp one, and the most important thing is how to make this transformation to
Owner kaser,2006[53]	Fuzzy	Visualization should provide easy Understanding of the result for fuzzy queries in the fuzzy data ware house.
Hua-Yang Lin, Ping-Yu Hsu and Gwo-Ji Sheen, 2007[20]	Data warehouse	Used systematic procedure which is based on the fuzzy set theory and has been proposed to select among the alternative with several decision criteria. The applicability of this procedure is illustrated through a case study of data warehouse system selection for the Bar code Implementation Project for Agricultural Products in Taiwan.

Table 2: Different approaches of fuzzy logic with data warehouse (cont.)

Author(s)	Purpose(s)	Description(s)
Lior Sapir, Armin Shmilovici, and Lior Rokach, 2008[37]	Fuzzy Data Warehouse	A data warehouse is a special database used for storing business oriented information for future analysis and decision-making. In business scenarios, where some of the data or the business attributes are fuzzy, it may be useful to construct a warehouse that can support the analysis of fuzzy. The users can make more intuitive and easy to understand queries in a natural like language
Lior Sapir and Armin Shmilovice,2008[37].	Fuzzy Data warehouse.	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 warehouse can be extended to the construction of a fuzzy data
Daniel Fasel, 2009[7]	Fuzzy Data warehouse	Used a fuzzy data house approach to support the fuzzy analysis of the customer performance measurement. The potential of the fuzzy data warehouse approach is illustrated by using a concrete example of customer performance measured for hearing instrument manufacture only A few for summary can be guaranteed by using this approach and the data
Fasel, D. and Shahzad, 2010[12]	Fuzzy Data warehouse	Fuzzy data warehouse model facilitated smooth transition between classes, have been proposed. By using the fuzzy Data warehouse model, data can be classified both fuzzily and sharply. Because of this, the FDWH supports qualitative and quantitative analyses without affecting the core
A. Prema and Dr.A.Pethalakshmi 2012[57]	ETL with Fuzzy	Proposed an algorithm to design data mart, which improves the decision making processes. To do so, we use Extraction, Transformation and Load (ETL) tools for better performance. In addition to that, the membership function of fuzzy is used for summarization
A.Prema and Dr.A.Pethalakshmi 2014[63]	Fuzzy optimization	Projected decision making methodologies to increase the sales promotion in data mart and located best decision making method by using fuzzy optimization technique.
A.Prema and Dr.A.Pethalakshmi 2014[58]	Fuzzy optimizaton	Estimated decision Matrix methodology to boost the sales endorsement in data mart using fuzzy optimization technique. This incorporated approach which improves efficiency of Hyper ETL and the decision making processes for better

Hua-Yang Lin et al. proposed the systematic procedure which is based on the fuzzy set theory and has been proposed to select among the alternative with several decision criteria. The applicability of this procedure is illustrated through a case study of data warehouse system selection for the Bar code Implementation Project for Agricultural Products in Taiwan. The procedure used objective structure, fuzzy set theory and fuzzy algebraic operation to solve the decision-making problem of choosing among DW alternatives, using ranking based on linguistic assessment. Although the case study is related to a specific software system and industry the same concept can be applied to other software products and industrial sector. The use of fuzzy set theory improves the decision making procedure by considering the vagueness and ambiguity prevalent in real-world systems. The author also found the using triangular fuzzy number made data collection, calculation and interpretation of the result easier for decision makers. Further proposed method can be computerized,
by implementing fuzzy linguistic assessments on computer, decision makers can automatically obtain the ranking order of alternatives and proposed a fuzzy multi-criteria decision making procedure, to facilitate data warehouse system selection,
with consideration given to both technical and managerial criteria [20].
M S Khan, et al.described the use of an FCMs is given, and the process of building the FCM for simulating the data warehouse diffusion scenario has been analyzed. The analyzed results are presented and compared with the corresponding results obtained by using the system dynamics methodology for modeling complex systems. Fuzzy cognitive maps (FCMs) have been used recently for representing and

analyzing complex systems evolving with time. Results of such analysis can be used for decision support. The work presented in this paper is aimed at exploring the effectiveness and reliability of an FCM in this regard by comparing its performance with system dynamics, which is a well-known modeling methodology. Compared with the systems dynamics methodology, an FCM had added the attraction of relative simplicity and ease of development [33].

Dainel Fasel demonstrated the uses of a fuzzy data house approach to support the fuzzy analysis of the customer performance measurement. The potential of the fuzzy data warehouse approach is illustrated by using a concrete example of customer performance measured for hearing instrument manufacture. Only a few for summary can be guaranteed by using this approach and the data ware house concepts can retain flexibility. Using a fuzzy approach in data warehouse concepts improves information quality for any company. It provides broader possibilities to create indicators for customer performance measurement as in the example given for a hearing instrument manufacture. The proposed approach does not include fuzzy linguistic concept directly in to hierarchical structure of dimension or into fact tables of the data ware house model but explains how the fuzzy concepts can be aggregated over dimensions without having redefined the fuzzy sets in every degree of granularity [7].

Visualization should provide easy understanding of the result for fuzzy queries in the fuzzy data ware house. Owen Kaser et al., described to apply the business intelligence techniques of the data ware housing and OLAP to the domain of text processing. A literary data ware-house is a 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 built by the data warehouse. 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 [53].

Lior Sapir et al. suggested that 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 the outlined Kimball's methodology for the design of a data warehouse can be extended to the construction of a fuzzy data warehouse. A case study demonstrates the visibility of the most commonly used methodology today which is Kimball's. It describes the process of translating business data and prose in to a dimensional model. It has also 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 of human concept which are actually more realistic. The fuzzy dimensions also allow more flexible and interesting filtering of the facts. The author has demonstrated that fuzzy measures used with fuzzy aggregation operators allows the user to better understand his business and data ware house measures [37].

Tang Jiafu et al. described an extensive study on fuzzy optimization, which leads to the following concluding remarks that the basic procedure of fuzzy optimization problems is to transform a fuzzy model in to a crisp one, and the most important thing is how to make this transformation to have an appropriate and reasonable interpretation. During the transformation, the first thing to do is to understand the problem and interpret the optimal solution. And then try to find an appropriate interpretation, and also propose some concepts and theory to support the interpretation, finally transform the fuzzy model in to a crisp one. The interpretation and formulation are the key constituent parts of the approaches, and they also bridge
the gap between the fuzzy optimization and the application in solving practical problems. This summary is made on the aspects of modeling and fuzzy optimization, classification and formulation for the fuzzy optimization problems, models and methods [85].

Kankana Chakrabarty et al presented an attempt with examples on real life problems. The occurrence of union/intersection of two fuzzy sets in two different universe is very common in many real life problems. This paper generalized Zadeh's notion of union and intersection in this work [32].

James F. Smith et al. described scheduling of electronic attack, resources distributed over many platforms is also under this process. The functional form of the fuzzy membership functions for the root concepts that will be selected heuristically and will generally carry one or more free parameters. Finally, fuzzy logic based multi-sensory association should prove effectiveness in its ability to form high quality conclusions faster than the standard of Bayesian algorithm because it allows linguistic data to be shared easily between the resource manager and the multi-sensor association algorithm [26].

James F. Brule's demonstrated that the fuzzy system is an alternative to traditional notions of set membership and logic that has had 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 [2]. 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, and laboratory water level controllers for ROBOT are welders, feature definition controllers for ROBOT vision, graphics controller for automated police sketchers and so on. 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.

The study on the theory and methodology of the fuzzy optimization has been active since the concept of fuzzy decision and the decision model under fuzzy environments were proposed by Bellman and Zadeh in 1970's.Various model and approaches to fuzzy linear programming [10,11,16,17,25,68,65,83,95,94],fuzzy multi-objective programming [72,73], fuzzy integer programming [81], fuzzy dynamic programming [31], possibilistic linear programming[8,35,66,69,82] and fuzzy non linear programming [40,87,88,92]have been developed over the past few years by many researchers. In the meantime, fuzzy ranking, fuzzy set operation, sensitivity analysis [52] and fuzzy dual theory [93], as well as the application of fuzzy optimization to practical problems also represent important topics.

The surveys on other topics of fuzzy optimization like discrete fuzzy optimization and fuzzy ranking have been conducted by Chanas [6] and Bortolan[5] respectively. The classification of uncertainties and of uncertain programming has been made by Liu [39,38]. The latest survey on fuzzy linear programming is provided by Inuiguchi & Ramik [24] from a practical point of view which is The possibilistic linear programming approach using example.

Recently many methods have been proposed for classifying fuzzy mathematical programming. Zimmermann [100] has classified the fuzzy mathematical programming into symmetric and asymmetric models. Luhandjula [41] has categorized the fuzzy mathematical programming into flexible programming, fuzzy stochastic programming and mathematical programming with the fuzzy coefficients. Inuiguchi and Ramik [24] further have classified the fuzzy mathematical programming into the following three categories in view of the kinds of uncertainties involved in the problems such as fuzzy mathematical programming with vagueness, i.e. flexible programming, fuzzy mathematical programming with ambiguity, i.e. possibilistic programming and fuzzy mathematical programming with vagueness and ambiguity, i.e. robust programming. In author's opinion, the formulation and classification of the fuzzy mathematical programming problems depend on what and where the fuzziness are involved.

Classification of the fuzzy linear programming has some problems owing to the simplicity of linear programming formulation and the existence of some developed software for optimization. Linear programming has been an important and most frequently applied for Operations Research technique for real life problems. Since the introduction of fuzzy theory into traditional linear programming problems by Zimmermann [98] and the fuzzy decision concept proposed by Bellman and Zadeh[3], the fuzzy linear programming (FLP) has been developed in all directions with successful applications. It has been an important area of the fuzzy optimization.

Symmetric approach is an important approach to the fuzzy optimization problems, especially for FMP1. The word 'Symmetric' used here comes originally from the symmetric model by Zimmermann. The symmetric approaches here cited by many researchers [41] usually refer to the approaches proposed by Bellman and Zadeh [3], Tanaka [84] and Zimmermann [98] to FMP1 firstly, and they are then extended to represent a type of approach to symmetric mathematical programming models in the sense that the goals and the system of constraints involved in the problem are dealt with in a symmetric way with regard to fuzziness. It means that the scope of the symmetric and the asymmetric approach is made from the perspective of the ways in which the goal and the system of constraints are treated, and not from the view point of the problem itself. The symmetric/asymmetric way in which the goals and the system of constraints are treated is understood to be the same concept assymmetric/asymmetric model. In this sense, the symmetric or asymmetric approach is named according to the symmetric or asymmetric model, and not to the symmetric or asymmetric problem.

A.Prema and A.Pethalakshmi presented a Fuzzy Data Mart model that imparts the exile interface to the users and also extends the Data Warehouses for storing and managing the fuzzy data along with the crisp data records. They proposed, an algorithm to design data mart, which improves the decision making processes. That proposed work is implemented in a linear programming problem through an assignment problem in terms of quantity [57].

A.Prema and A.Pethalakshmi projected decision making methodologies to increase the sales promotion in data mart and located best decision making method by using fuzzy optimization technique. This paper has compared the various methodologies by using fuzzy optimization technique and observed that the decision matrix approach is the best methodology to improve the performance of sales data mart rather than other Decision Model [63].

III. DATA MART, DATA WAREHOUSE AND DECISION MAKING METHODOLOGIES

Decision making can be regarded as the cognitive process resulting in the selection of a belief or a course of action among several alternative possibilities. Every decision-making process produces a final choice that may or may not prompt action. Decision-making is the study of identifying and choosing alternatives based on the values and preferences of the decision maker. Decision-making is one of the central activities of management and is a huge part of any process of implementation.

Maxim Likhachev et al. described a new planning algorithm, calledMCP(short for MDP Compression Planning),which combines A* search with value iteration for solving Stochastic Shortest Path problem in MDPs with sparse stochasticity. They present experiments which show that MCP can run substantially faster than competing planners in domains with sparse uncertainty; these experiments are based on a simulation of a ground robot cooperating with a helicopter to fill in a partial map and move to a goal location, planning algorithm designed for deterministic worlds, such as A* search, usually run much faster than algorithms designed for worlds with uncertain action outcomes, such as values forces us to use the slower algorithms to solve them, interspersed with a small number of sensing actions which have uncertain outcomes [46].

Jason D. Williams et al. displayed how a dialogue model can be represented as a Partially Observable Markov Decision Process with observations composed of a discrete and continuous component. The continuous component enables the model to directly incorporate a confidence score for automated planning. Using a tested simulated dialogue management problem, this paper shows how recent optimization techniques are able to find a policy for this continuous which outperforms a traditional MDP approach. Further a method is presented for automatically improving handcrafted dialogue managers by incorporating the belief state monitoring, including confidence score information. Experiments on the test bed system show significant improvements for several example handcrafted dialogue managers across a range of operating conditions [27].

Mausam et al. defined the concurrent MDP problem and described two algorithms to solve them, Pruned RTDP relies on combo-skipping and combo-elimination with an admissible initial value function, it is guaranteed to converge to an optimal policy and is faster than plain, labeled RTDP on concurrent MDPs. sample RTDP performs backups on a random subset of possible action combination; when guided by our heuristics. it converges orders of magnitude faster than other methods and produces optimal or close-to-optimal solutions. It is believed that the author's sampling techniques will be extremely effective on very large, concurrent MDP problems. They believe, the methods will extend easily to solve concurrent MDP with rewards non-absorbing goals and other formulation. And also to prove error bounds on S-RTDP and to modify it so that its convergence is formally guaranteed. They also hope to extend their methods to include durative actions, and continuous parameters [45].

Patrice Perny et al. presented an algebraic approach to note Markov Decision Processes (MDPs), which allows an unified treatment of MDPs and includes many existing models (quantitative or qualitative) with particular cases. In algebraic MDPs, rewards are expressed in a semi ring structure, uncertainty is represented by a decomposable plausibility measure valued on a second semi ring structure, and preferences over policies are represented by a generalized expected utility. This paper recasts the problem of finding an optimal policy at a finite horizon as an algebraic path problem in a decision rule graph where arcs are valued by functions, which justifies the use of the Jacobi algorithm to solve algebraic bell-man equations. In order to show the potential of this general approach, they exhibit new variations of MDPs, admitting complete or partial preference structures, as well as probabilistic or possibilistic representation of uncertainty. The author has introduced a general approach for defining solvable MDPs in various contexts. The interest of this approach is to factorize many different positive results concerning various rewards system, uncertainty and decisbraic model. Once the structure on reward, the representation of uncertainty and decision criteria have been chosen, it is sufficient to check two semi rings on V and P and that

conditions (C1) through (C5) are fulfilled to justify the use of an algorithm "a la Jacobi" to solve the problem. It is likely that this result generalizes to the infinite horizon case, provided a suitable topology is defined on the policy valuation space [56].

Finale-doshi-velez presented the infinite POMDP, a new model for Bayesian RL in partially observable domains. The iPOMDP provides a principled framework for an agent to posit more complex models of its world as it gains more experience. Despite the complexity of the model to the agent's experience, the agent is not forced to consider large uncertainties-which can be computationally prohibitive near the beginning of the planning process, but it can later come up with accurate models in the world when it requires them. An interesting question may also apply to these methods to learn large MDP models within the Bayes-Adaptive MDP framework. Recent work in Bayesian reinforcement learning has made headway in learning POMDP(iPOMDP) model that does not require knowledge of the state space; instead, it assumes that the number of visited states will grow as agent explores its world and only models visited states explicitly and demonstrated the iPOMDP On several standard problems [14].

Patrice Perny and Paul Weng presented the search of the best compromise solution in MMDPs with use distance. Despite this non-linear criterion the author has provided an LP-solvable formulation of the problem. Experiments have shown the practical feasibility of the approach on difficult instances specially designed to exhibit conflicting criteria. In all the experiments, the Tchebycheff criterion significantly brings the out performance on weights sum concerning the quality compromises. Interestingly enough, this way of incorporating non-linear function in MMDPs could be extended to other non-linear criteria. For instance, our approach can be applied to multi-agent problems with a non linear social welfare function to determine polices that fairly share rewards among agents. The users feel that this notion of optimality depends on the initials state. It appears that the best compromise policy cannot be found by a direct adaptation of value iteration and they observed in some situations, the optimal solution can only be obtained with a randomized policy. To overcome all these problems the paper proposes a solution method based linear programming and give some experimental result [55].

Planning under uncertainty can be approached according to (fully observable) Markov Decision Processes (MDP) or a partially observable Markov Decision (POMDP), and both of these techniques have been applied to dialogue the management. The application of MDPs was first explored by Levin and Pieraccini (1997).

Esther Levin and Roberto Pieraccini [9] provided a formal treatment of how a MDP may be applied to dialogue management, and Singh et al. (2002)[88] show application to real systems. However, MDPs assume the current state of the environment (i.e., the conversation) is known exactly, and thus they do not naturally capture the uncertainty introduced by the speech recognition channel.

Partially observable MDPs (POMDPs) extend MDPs by providing a principled account of noisy observations. Roy et al. (2000)[49] compare an MDP and a POMDP version of the same spoken dialogue system, and find that the POMDP version gains more reward per unit time than the MDP version. Further, the authors show a trend that as speech recognition accuracy degrades, the margin by which the POMDP outperforms the MDP increases.

Zhang et al. (2001) extend this work in several ways. First, the authors add "hidden" system states to account for various types of dialogue trouble, such as different source of speech recognition errors. Second, the authors use Bayesian networks to combine observations from a variety of source (including confidence score). The authors again show the POMDP-based methods outperform MDP-based methods. In all of these proposals, the authors have incorporated confidence score by dividing the confidence score metric into regions, often called confidence buckets". For example, in the MDP literature, Singh et al. (2002) [74] tracks the confidence bucket for each field as "high, medium, or low" confidence. The authors address neither how to determine an "optimal" number of confidence buckets, nor how to determine the "optimal" thresholds of the confidence score metric that divide each bucket.

In the POMDP literature, Zhang et al. (2001) [97]used Bayesian networks to combine information from many continuous and discrete sources, including confidence score, to compute probabilities for two metrics called "Channel Status" and "Signal Status". Thresholds are then applied to these probabilities to form discrete and binary observations for the POMDP. However, it is not clear of how to set these thresholds to maximize POMDP return.

Table3 summarizes the various decision making approaches with data warehouse.

Table 3: Decision making approaches with data repository concept

Author(s)	Purpose(s)	Description(s)
Esther Levin and Roberto Pieraccini 1997[9]	Markov Decision Process	Planning under uncertainty can be approached according to(fully observable) Markov decision processes (MDP) or a partially observable Markov decision (POMDP), and both of these techniques have been applied to dialogue the management
Patrice Perny, Olivier Spanjaard and PaulWeng[56]	Markov Decision Process	Provided with algebraic approach to note Markov decision processes (MDPs), which allows a unified treatment of MDPs and includes many existing models (quantitative or qualitative) with particular cases. In algebraic MDPs, rewards are expressed in a semi ring structure, uncertainty is represented by a decomposable plausibility measure valued on a second semi ring structure, and preferences over policies are represented by a generalized expected utility.
Singh et al. (2002)[74]	Decision Making	Tracks the confidence bucket for each field as "high, medium, or low" confidence. The authors do not address neither how to determine an "optimal" number of confidence buckets, nor how to determine the "optimal" thresholds of the confidence score metric that divide each bucket.
Mausam and Daniel S. Weld,2004[45]	Decision Making	Described two algorithms to solve them, Pruned RTDP relies on combo-skipping and combo-elimination with an admissible initial value function, it is guaranteed to converge to an optimal policy and is faster than plain, labeled RTDP on concurrent MDPs
Maxim Likhachev, Geoff Gordon and SebastianThrun,2004[46]	Markov analysis	Proposed a new planning algorithm, called MCP (short for MDP Compression Planning), which combines A* search with value iteration for solving Stochastic Shortest Path problem in MDPs with sparse stochasticity
Jason D. Williams, pascal Poupart and Steve Young,2005[27]	Markov Decsion Process	Displayed how a dialogue model can be represented as a Partially Observable Markov Decision Process with observations composed of a discrete and continuous component. The continuous component enables the model to directly incorporate a confidence score for automated planning. This paper show how recent optimization techniques are able to find a policy for this continuous which outperforms a traditional MDP approach
Jose L. Salmeron and Florentin Smarandache,2007[30]	Decision Matrix	Proposed the neutrosophic decision matrix method as a more realistic tool for decision making. In addition, a de-neutrosophication process is included. Numerous scientific publications address the issue of decision making in every fields. But, little efforts have been done for processing indeterminacy in this context
Zack, M. H.2007[96]	Decision Support System	For academics and practitioners concerned with computers, business and mathematics, one central issue is supporting the decision makers. In that sense, making coherent decisions requires knowledge about the current or future state of the world and the path to formulating a fit response
Finale-doshi-velez ,2009[14]	Markov Decsion Process	The iPOMDP provides a principled framework for an agent to posit more complex models of its world as it gains more experience. The complexity of the model to the agent's experience, the agent is not forced to consider large uncertainties-which can be computationally prohibitive-near the beginning of the planning process, but it can later come up with accurate models in the world when it requires them. An interesting question may also apply to these methods to learn large MDP models within the Bayes-Adaptive MDP framework
Patrice Perny and Paul Weng,2010[55]	Morkov Model	Presented the search of the best compromise solution in MMDPs with use distance. Although this non-linear criterion the author have provided a LP-solvable formulation of the problem. Experiments have shown the practical feasibility of the approach on difficult instances specially designed to exhibit conflicting criteria.
D. Ashok Kumar and M. C. Loraine Chalet Annie,2012[1]	Decision Making	Explained modern electronic health records that are designed to capture and render vast quantities of clinical data during the health care prone. Utilization of data analysis and data mining methods in medicine and health care is sparse. Medical data is one of the heavily and categorical types of data.
A.Prema and Dr.A.Pethalakshmi 2014[58]	Decision Matrix	Estimated decision Matrix methodology to boost the sales endorsement in data mart using fuzzy optimization technique. This incorporated approach which improves efficiency of Hyper ETL and the decision making processes for better performance in Data Mart.

Looking outside the (PO) MDP framework, Paek and Horvitz (2003) suggest using an influence diagram to model user and dialogue state, and selecting actions based on "Maximum Expected [immediate] Utility." This proposal can be viewed as a POMDP with continuous observations that greedily select actions _ i.e., which selects actions based only on immediate reward. By choosing appropriate utilities, the authors show how local grounding action can be automatically selected in a principled manner. In this work the authors are interested in POMDPs as they enable planning over any horizon. This paper makes two contributions. First the paper shows how a confidence score can be accounted for exactly in a POMDP-based dialogue manager by treating confidence score at a continuous observation. Using a test bed simulated dialog management problem, the paper showed that recent optimization techniques produce policies which outperform traditional MDP-based approaches across a range of operating conditions. Secondly they show how a hand-crafted dialogue manager can be improved automatically by treating it as a POMDP policy. And then it is shown how a confidence score metric can be easily included in this improvement process. This paper illustrated the method by presenting three handcrafted controllers for the test bed dialog manager, and shows that the technique improves the performance of each controller significantly across a variety of operating conditions.[91].

D. Ashok Kumar and M. C. Loraine explained modern electronic Health records that are designed to capture and render vast quantities of clinical data during the health care prone. Utilization of data analysis and data mining methods in medicine and health care is sparse. Medical data is one of the heavily and categorical types of data. A Dichotomous variable is the 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 [1].

Traditional optimization techniques and methods had been successfully applied for years to solve problems with a well-defined structure/configuration, sometimes known as hard systems. Such optimization problems are usually well formulated by crisply specific objective functions and specific system of constraints, and solved by precise mathematics. Unfortunately, real world situations are often not deterministic. There exist various types of uncertainties in social, industrial and economic system such as randomness of occurrence of events imprecision and ambiguity of system data and linguistic vagueness, etc. which come from many ways[77], including errors of measurement, deficiency in history and statistical data,

insufficient theory, in complete knowledge expression, and the subjectivity and preference of human judgment etc. As pointed out by Zimmermann[99],various kinds of uncertainties can be categorized as stochastic uncertainty and fuzziness.

Stochastic uncertainty relates to the uncertainty of occurrences of phenomena or events. Its characteristics lie in that descriptions of information are crisp and well defined however they vary in their frequency of occurrence. Systems with this type of uncertainty are the so called stochastic systems, which can be solved by stochastic optimization technique using probability theory. In some other situations, the decision-maker (DM) does not think the commonly-used probability distribution is always appropriate, especially when the information is vague, relating to human language and behavior, imprecise/ambiguous system data, or when the information could not be described and defined well due to limited knowledge and deficiency in its understanding. Such types of uncertainty are categorized as fuzziness which can be further classified into ambiguity or vagueness.

Benoit Bagot discussed whether people decide rationally or irrationally has elicited many interesting results, but did not result in any final answer. This remains true today, a big advantage of objectifying decision lies in the possibility of using strategies systematically in a repeatable and even automated process. The relief that results from this can free up more capacities to search for new strategies used in a genetic problem for the optimalization of an automation gear box, this tool helps to conciliate numerous, partly opposing criteria, in order to emphasize a unique final solution [4].

Jose L.Salmeron and Florentin Smarandache proposesd a renewed decision matrix method as a methodological support. The

author has used neutrosophic logic. This emerging logic extends the limits of information for supporting decision making for academics and practitioners concerned with computers, business and mathematics, one central issue is supporting decision marks. A generalization of logic is proposed and it emerges as an alternative to the existing logic and it represents a mathematics model of uncertainty and indeterminacy. This paper proposes the neutrosophic decision matrix method as a more realistic tool for decision making. In addition, a de-neutrosophication process is included. Numerous scientific publications address the issue of decision making in every fields. But, little efforts have been done for processing indeterminacy in this context. But this paper shows a formal method for processing indeterminacy in decision matrix method and includes a de-neutrosophication process. The main outputs of this paper are two-folds: it provides a neutrosophic tool for decision making and it also includes indeterminacy in a decision tool [30].

For academics and practitioners concerned with computers, business and mathematics, one central issue is supporting the decision makers. In that sense, making coherent decisions requires knowledge about the current or future state of the world and the path to formulating a fit response (Zack, 2007). [96]

The authors proposed a generalization of Decision Matrix Method (DMM), or Pugh Method as sometimes is called, using Neutrosophic logic (Smarandache,1999). The main strengths of this paper are two-folds: it provides a more realistic method that supports group decision with several alternatives and it presents a de-neutrosophication process. It is proposed that this is a useful endeavour Decision Matrix Method (DMM) which was developed by Stuart Pugh (1996) as an approach for selecting concept alternatives. DMM is a method (Murphy, 1979) [48] that allows decision makers to systematically identify and analyze the strength of relationships between the sets of information. This technique is especially interesting for looking at large numbers of factors and assessing each relative importance. Furthermore, DMM is a method for alternative selection using a scoring matrix. DMM is often used throughout planning activities to select produce/service feature and goals and to develop process stages and weight options.

A.Prema and A.Pethalakshmi estimated Hyper ETL with decision Matrix methodology to boost the sales endorsement in data mart using fuzzy optimization technique. This incorporated approach which improves efficiency of Hyper ETL and the decision making processes for better performance in Data Mart. The objective of the paper is to find out an effective decision making and to get better performance of ETL process through attaining high Scalability, CPU utilization, hroughput, Reliability, Execution speed than an existing ETL. This Paper suggested the design of

Hyper ETL with Decision Matrix method and Fuzzy optimization technique used to formulate right decision making to raise the sales promotion[58].

IV. SUMMARY

The Extraction Transformation and Load plays a vital role in Data Mart. The performance analyses of various approaches for Data Mart in the context of decision making methodologies were reviewed for different data sets.

REFERENCES

[1] Ashok Kumar.D and M. C. Loraine Chalet Annie, “Decision Making onDichotomous Medical Data using Novel Clustering approach”, National conference on Signal and Image processing (NCSIP),2012.

[2] Baldwin.J.F, “Fuzzy systems- logic andreasoning in fuzzy applications” London:Academic press, 1981.

[3] Bellman R.E and L. A. Zadeh, “Decision making in a fuzzy environment”, Management Science, 17: B-141-164,1970.

[4] Benoit Bagot “The Harmonic Decision Matrix: a group of operators for the fuzzy-logic, multi-objective decision and optimization”.

[5] Bortolan.G, R. Degani, “A review of some method for ranking fuzzy subsets”,Fuzzy Sets and Systems,1985,15: 1- 19.

[6] Chanas.S, D. Kuchta,”Discrete fuzzy optimization, in Fuzzy Sets in decisionAnalysis” Operations Research and Statistics-The Handbook of Fuzzy Sets Series(Slowinski, Ed.,) Kluwer Academic, Dordrecht, 249-276,1998.

[7] Daniel Fasel, “A fuzzy data warehouse approach for the customer performance measurement for a hearing instrument manufacturing Company”, Sixth International conference on fuzzy systems and knowledge discovery,2009.

[8] Dubois.D,”Linear programming with fuzzy data, in J. C. Bezdek Ed., Analysis of fuzzy Information, CRC Press, Boca Raton, FLa, 241- 263,1987.

[9] Esther levin and Robert Pieraccini,”A Stochastic Model of Computer-Human Interaction for Learning Dialogue Strategies”, Eurospeech,Greece,1997.

[10] Fang.S, C. F. Hu, et al.,"Linear programming with fuzzy coefficients in constraints",Computers and mathematics with Applications, 37: 63-76,1999.

[11] Fang.S.C, and G. Li," Solving fuzzy relation equations with a linear Objective function", Fuzzy Sets and Systems,103: 107-113,1999.

[12] Fasel, D. and Shahzad, K "A Data Warehouse Model for Integrating Fuzzy Concepts in Meta Table Structures" IEEE digital Xplore, Mar 2010.

[13] Fasel.D and D. Zumstein.,"A fuzzy data warehouse approach for web analytics", A Web Science Perspective, volume 5736 of Lecture Notes in Computer Science, 276-285. Springer, 2009.

[14] Finale-doshi-velez,"The infinite partially observable markov decision process",Advances in neural information processing system, 22-2009.

[15] Gregory S. Nelson and Cary in "Planning for and Designing a Data Warehouse: A Hands on. Workshop,May 19, 2007- 1. Paper 111-2007.

[16] Hamacher.H, H. Leberling and H. J. Zimmermann,"Sensitivity analysis in Fuzzy linear programming", Fuzzy Sets and Systems,1(1): 269-281,1978.

[17] Han.S, H. Ishii and S. Fujii,"One machine scheduling problem with fuzzy Due dates", European Journal of Operational Research, 79: 1-12,1994.

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

[19] Huamin Wang,"An ETL Services Framework Based on Metadata", Second International Workshop on Intelligent Systems and Application, May 2010.

[20] Hua-Yang Lin, Ping-Yu Hsu and Gwo-Ji Sheen, "A fuzzy- based decision-making procedure for data warehouse system selection" Expert Systems with Applications, Elsevier 2007.

[21] Inmon W.H. , Derek Strauss and , Genia Neushloss,"DW 2.0: The Architecture For the Next Generation of Data Warehousing" Morgan Kaufman Series in Data Management Systems, Aug 2008.

[22] Inmon, William,"Data Mart Dose Not Equal Data Warehouse",2000.

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

[24] Inuiguchi.M, J. Ramik,"Possibility linear programming: a brief review of Fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem", Fuzzy Sets and Systems, 111: 3-28, 2000.

[25] Ishibuchi.H, N. Yamamoto, T. Murata and H.Tanaka, "Genetic algorithms And neighborhood search algorithm for fuzzy shop scheduling problems", Fuzzy Sets and Systems, 67(1): 81-100, 1994.

[26] James F. Smith and Robert D. Rhyne "A fuzzy logic based algorithm for Optimal allocation of Distributed resource",Naval Research Laboratoty, September 2000.

[27] Jason D.Williams,Pascal Poupart and Steve Young.,"Partially observable Markov decision processes with continuous observations for dialogue Management",6thSIGdial Workshop on Discourse and Dialogue Lisbon, Portugal Sep 2- 3,2005.

[28] Jeffrey R. Bocarsly, "Complex ETL Testing-A Strategic Approach",2014.

[29] Jeremy, Jean King and Andreas Schindler,"Understanding of Business Intelligence: ETL and Data mart Practices".

[30] Jose L. Salmeron and Florentin Smarandache."Redesigning decision matrix method with an indeterminacy-based inference process", International Journal of Applied Mathematics and Statistics, Volume 13, NumberM08,March2008.

[31] Kacprzyk.J, and A. O. Esogbue,"Fuzzy dynamic programming:Main developments and applications", Fuzzy Sets and Systems, 81(1): 31-46,1996.

[32] Kankana Chakrabarty, Ranjit Biswas and Sudarsan Nanda,"union and intersection of fuzzy sets".

[33] Khan.M S, M Quaddus, A Intrapairot3 and A Chongl., "Modelling data warehouse diffusion using fuzzy cognitive maps -a comparison with the system dynamics approach", Third Western Australian Workshop on Information systems Research(WAWISR) 2000.

[34] Kimball.R and M. Ross,"The Data Warehouse Toolkit",Wiley Publishing, Inc.,2002.

[35] Lai.Y.J and C. L. Hang,"A new approach to some possibilistic linear programming problems", Fuzzy Sets and Systems, 49:121-133,1992.

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

[37] Lior Sapir, Armin Shmilovici, and Lior Rokach,"Methodology for the Design Of Fuzzy Data Warehouse" IEEEXplore,2008.
[38] Liu.B,"Theory and Practice of Uncertain Programming",Physica-Verlag,Heidelberg, 002.
[39] Liu.B,"Uncertain Programming",Wiley & Sons, New York,1999.
[40] Liu.J, and SC Fang,"Solving nonlinear optimization problems with fuzzy Relation equation constraints", Fuzzy Sets and Systems, 119: 1-20, 2001.
[41] Luhandjula. M. K, "Fuzzy optimization: an appraisal", Fuzzy Sets and Systems, 30(3): 257-282,1980.
[42] Lunan Li, "A framework study of ETL processes optimization based on Metadata repository", International Conference on Computer Engineering and Technology, April 2010.
[43] Marko Banek, Zoran Skocir and Boris Vrdoljak,"A Tool for Data Warehouse Design Form Xml Sources".
[44] Master Data anagement- An Oracle, White Paper September 2011.
[45] Mausam and Daniel S.Weld,"Solving concurrent Marko Decision Processes",American Association for Artificial Intelligence, 2004.
[46] Maxim Likhachev, Geoff Gordon and Sebastian Thrun.,"Planning for Markov Decision Processes with Space Stohasticity",Advances-in-neural-information- processing-systems-17-2004.
[47] 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.
[48] Murphy, K. R.,"Comment on Pugh's method and model for Assessing environmental effects", Organizational Behavior and Human Performance, 56-59,1979.
[49] Nicholas Roy, Joelle Pineau and Sebastin Thrun,"Spoken Dialogue Management using Probabilistic Reasoning", Annual meeting of the Association for Computational Linguistics,2000.
[50] Nitin Anand, "Application of ETL Tools in Business Intelligence",International Journal of Scientific and Research Publications , Vol 2, Issue 11,November2012.
[51] Osama E.Sheta and Ahmed Nour Eldeen ,"The Technology Of Using A DataWarehouse to Support Decision-Making In Health Care" ,International Journal of Database Management Systems (IJDMS) Vol.5, No.3, June 2013.
[52] Ostermark.R,"Sensitivity analysis of fuzzy linear programs: An approach to parametric interdependence", 16: 113-120,1987.
[53] Owner kaer, Steven Keith and Daniel Lomire,"Data ware housing with literature", September 11, 2006.
[54] PanosVassiliadis and Alkis Simitsis "Extraction, Transformation and Loading"Encyclopedia of Database Systems, 1095-1101,2009.
[55] Patrice Perny and Paul Weng, "On Finding compromise Solutions in Multiobjective MarkovDecision Processes" ECAI 2010.
[56] Patrice Perny, Olivier Spanjaard and PaulWeng. ,"Algebraic Markov Decision Processes", International joint conference on Artificial intelligence.
[57] Prema.A and A. PethaLakshmi "An Approach to construct the fuzzy Data Mart using Soft computing" in International Journal of Computer Applications(IJCA), Vol 58, Issue 20,Nov2012.
[58] Prema.A and Pethalakshmi.A "An approach to Formulate Intelligent Repository", International Journal of Computer Science Trends and Technology (IJCST) -Volume 2 Issue 3, May-Jun 2014.
[59] Prema.A and Pethalakshmi.A "An Improved Enhancement of Decision-making Analysis to increase the sales promotion using Hyper ETL in DataMart" International Journal of Computational Intelligence and Informatics(IJCII) Vol 2,No 4, Mar 2013.
[60] Prema.A and Pethalakshmi.A "Novel Approach in ETL" IEEE Explore Digital Library ISBN: 978-1-4673-5843-9, Feb 2013.
[61] Prema.A et al "A Comparative analysis of ETL and Hyper ETL" International Journal of Emerging Trends and Technology in Computer Science (IJETTCS)-Vol 2, Issue 6, Nov-Dec 2013.
[62] Prema.A et al"An Approach of Quantitative Analysis for Optimizing the Movements of Sales Items Using Markov Decision Model." European Journal of Scientific Research, Vol.119,No1,Feb2014, 77-84.
[63] Prema.A et al "Formulate Competent Decision Making in Data Mart using Fuzzy optimization

technique",International Journal of Computer Science and Information Technologies (IJCSIT) Vol 5(3), pp 3196-3202, May2014.
[64] Radha Krishnan and Sree Kanth, "An Object Oriented modeling and Implementation of Web based ETL process" IJCSNS, International Journal of Computer Science and Network Security, Vol 10 no. 2, February 2010.
[65] Ramik.J and H. Rommelfanger,"Fuzzy mathematical programming based on Some new inequality relations", Fuzzy Sets and Systems, 81(1): 77-87,1996.
[66] Ramik.J, and J. Rimanek, "Inequality relation between fuzzy numbers and its Use in fuzzy optimization", Fuzzy Sets and Systems,16: 123-138, 1985.
[67] Rohit Ananthakrishnal Surajit Chaudhuri and Venkatesh Ganti "Eliminating Fuzzy Duplicates in Data Warehouse" Proceedings of the 28th VLDB Conference,Hong Kong, China, 2002.[2.36]
[68] Rommelfanger.H,"Fuzzy linar programming and applications", European Journal of Operations Research, 92(3): 512-527, 1996.
[69] Rommelfanger.H, Hanuschek R. and J. Wolf,"Linear programming with fuzzy objective", Fuzzy Sets and Systems, 29: 31-48, 1989.
[70] Saagari.S, P. Devi Anusha, Ch. Lakshmi Priyanka, V. S. S. N. Sailaja, Data Warehousing, Data Mining, OLAP and OLTP Technologies are Essential Elements to Support Decision-Making Process in Industries ", International Journal of Innovative Technology and Exploring Engineering (IJITEE),Volume-2, Issue-6, May 2013.
[71] Sabir Asadullaev,"Data Warehouse Architectures - III SWG IBM",2009.
[72] Sakawa.M and H. Yano,"An interactive fuzzy satisfacing method for multi-objective nonlinear programming problems with fuzzy parameters", Fuzzy Sets and Systems, 30(10): 221-238, 1989.
[73] Sakawa.M and H. Yano,"Fuzzy dual decomposition method for large-scale multi objective non-linear programming problem", Fuzzy Sets and Systems, 67:19-27, 1994.
[74] Satinder Singh, Diane Litman,Michael Kearns and Marilyn Walker,"Optimizing Dialogue Management with Reinforcement Learning: Experiments with the NJFun System", Journal of Artificial Intelligence,Vol.16, 105-133,2002.
[75] Shaker H. Ali El-Sappagh, Abdeltawab M. Ahmed Hendawi, Ali Hamed El Bastawissy "A proposed model for data warehouse ETL processes"Journal of King Saud University - Computer and Information Sciences, Volume 23, Issue 2,91-104 July 2011.
[76] Simitsis, A Vassiliadis, P. Sellis, T.,"State-space optimization of ETL workflows", IEEE Transactions on Knowledge and Data Engineering, Vol 17, Issue 10, Oct 2005.
[77] Simon.F,"Uncertainty and imprecision: modeling and analysis",Journal Of Operation Research Soc.,46: 70-79,1995.
[78] Srikanth.K, N. V. E. S. Murthy, J. Anitha "Data Warehousing Concept Using ETL Process For SCD Type-3" in International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 2, Issue 5, 142-145,Sep - Oct2013.
[79] Stephen Overton, ," Developing a Flexible ETL Process to Let SAS® Capture Data Changes Efficiently in a Data Warehouse and Clean Up the Mess",Data Management, Paper 112-2012.
[80] Steven Scherma and Stephen Bolivar,"Data mart Use For Complex Data Retrieval in an ArcIMS Applications in Tech",Whitepapers 2010.
[81] Stoica.M, et al.,"Fuzzy integer programming Fuzzy Sets and Decision Analysis", 123-132,1984.
[82] Tanaka.H and K. Asai, Fuzzy programming problems with fuzzy numbers, Fuzzy Sets and Systems,13: 1-10,1984.
[83] Tanaka.H, and K. Asia, "Fuzzy solution in fuzzy linear programmin",IEEE Trans. on SMC., 14(2): 325-328,1984.
[84] Tanaka.H, T. Okudu and K. Asai,"On fuzzy mathematical programming", Cybernet,3: 37-46,1947.
[85] TANG Jiafu WANG Dingwei, Richard Y K Fung and Kai-Leung., "Understanding Of Fuzzy Optimization Theories and Methods",Journal of Systems Science and Complexity, Vol. 17 No. 1 Jan., 2004.
[86] Tang Jun,Shenyang,Feng Yu,Tong Gang "Research & Application of ETL Tool in Business Intelligence Project" Information Technology and Applications,2009.

[87] Tang.J and D. Wang,"A non-symmetric model for fuzzy nonlinear programming problems with penalty coefficients",Computers & Operations Research, 24(8):717-725,1997.

[88] Tang.J, D. Wang and Richard Y K Fung," Model and method based on GA for non-linear programming problems with fuzzy objective and resources", International Journal of Systems Science, 29(8): 907-913,1998.

[89] Tec Ying Wah, Ng Hooi Peng, and Ching Sue Hok, "Building DataWarehouse", 24th South East Asia Regional Computer Conference, Bangkok, Thailand Nov 2007.

[90] Teori kontra praktik Ann Josefsson and Isabel Zitoun, "ETL-processen", Uppsala Universitet,2010.

[91] Tim Paek and Eric Horvitz ," On the Utility of Decision - Theoretic Hidden Subdialog". In Proceedings of International Speech Communication Asociation (ISCA) Workshop on Error Handling in Spoken Dialogue Systems,Switzerland,2003.

[92] Trappey.J.F.C, C. R. Liu and T. C. Chang, "Fuzzy non-linear programming: Theory and application in manufacturing",International Journal of Production Research, 26(5): 957-985, 1988.

[93] Verdegay.J.L,"A dual approach to solve the fuzzy linear programming problems", Fuzzy Sets and Systems, 14(1): 131-141, 1984.

[94] Wang.D and S. C. Fang, "A genetic-based approach for aggregate production planning in fuzzy environment", IEEE Trans. on SMC (Part A), 12(5): 636-645,1997.

[95] Wang.D,"An inexact approach for linear programming with fuzzy objective and resource", Fuzzy Sets and Systems, 8(1): 61-68,1997.

[96] Zack, M. H.,"The role of the decision support systems in an indeterminate world", Decision Support Systems,. 2007.

[97] Zhang.B, Q, Cai, J. Mao, E. Chang, and B. Guo,"Spoken Dialogue Management as Planning and Acting under Uncertainty",Eurospeech,Denmark,2001.

[98] Zimmermann. H-J.,"Description and optimization of fuzzy system", International journal of General System,2: 209-216,1976.

[99] Zimmermann.H.J,"Fuzzy Set Theory and Its Applications",1991.

[100] Zimmermann.H-J,"Application of fuzzy set theory to mathematical programming", Information Science,36: 29-58,1985.