Novice Evaluation and Comparative Survey on Database Management System, Data Warehousing and Data Mining

Mohd Muntjir
Department of Information Technology
College of Computers and Information Technology
Taif University Taif Saudi Arabia

ABSTRACT
A Data management is the progression and execution of frameworks, approaches, processes and measures in order to control the information lifecycle requires for a business in a creative method. A Database Management System, Data Mining and Data warehouse which intrinsically concentrate on the management and maintenance of data in modern world. Different enterprise services with various methods are available in Database Management System. Henceforth; a database system overcomes with an important role in this real world. In this paper, it is mentioned about the overview of database management system, data mining and data warehousing in DBMS, along with on their modern needs. For management of data extraction a bank end tools are being discussed here. Furthermore, it is emphasized to cleaning and uploading data into a data warehouse system; and front-end client tools for querying; data evaluation server extensions for proceedings the efficiency of queries, several mechanisms for metadata organization and for control the warehouse. This latest summary provides the fundamentals comprehension of different database tools and procedures. We will have a look about the database integration; data security with data segregation and storage.

Keywords
DBMS, Database, Data warehouse, Data Mining, Data management, Architecture

1. INTRODUCTION
A DBMS is a computer Software that collections, locates, indexes and maintains its reliability, and storage of data, and output in the form required by a user. A Relational database management systems (RDBMS) appliances the relational model of tables and associations. Inconsistent to a data processing system a DBMS performs only nominal scientific procedures. The database Management System is essential because without the presence of certain kind of instructions and regulations it is not possible to uphold the database.A Database Management System can be an exceptionally composite set of software programs that operates the organization ,storage and retrieval of data in a database system.It also accomplishes the security and varacity of the database.A DBMS states requetsts for data from the application program and initiates the operating system to allocate the suitable data[1].The major solution of Database Management System is to construct and control the data, to make it achievable for the users.
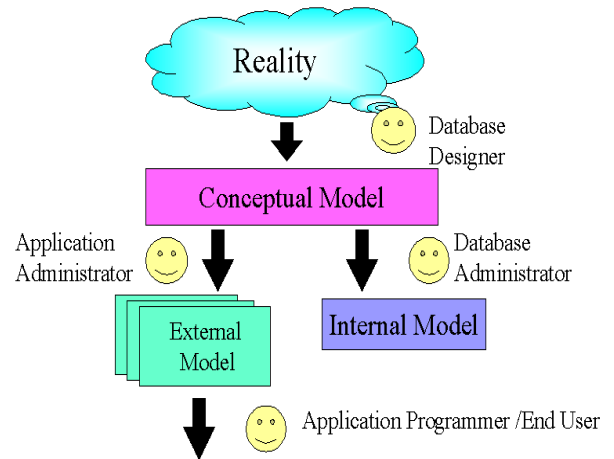

Fig 1:Database Model

1.1 The Main Benefits Of Database Methodology

1.1.1 Control on Data Redundancy:
Database Management System supports to construct a background in which end users have better approach to more and better-managed data. These types of approaches make it conceivable for end users to react speedily to alterations in their environment [1].

1.1.2 Improved data security
The more users approach with data, the greater the threats of data security violations. Organizations invest significant aggregates of time, determination, and money to confirm that business data are used appropriately. A Database management system delivers a schema for better implementation of data confidentiality and reassurance strategies.

1.1.3 Improved data integration.
 Larger approach to well-managed data endorses an integrated opinion of the organization's progressions and a stronger opinion of the big depiction. It develops much easier to see how arrangements in one section of the company distress other divisions.

1.1.4 Reduced data inconsistency
Data irregularity occurs when dissimilar types of the similar data perform in unusual spaces. For example, data discrepancy happens when a company's sales department stocks a sales representative's name as "Scott Steris" and the company's personnel department stores that same person's name as "Michal S. Steris," or when the company's regional sales office shows the price of a product as $49.90 and its countrywide auctions office demonstrations the uniform

product's price as $45.90. The opportunity of data irregularity is significantly summarized in an applicably measured database [2].

1.1.5 Improved data approach
The Database Management Systems constructs it conceivable to create fast responses to ad hoc queries. Since a database perception, a query is a precise application delivered to the Database Management Systems for data management—for example, to recite or modernize the data. Basically place, a query is a problem, and an ad hoc query is a spontaneous issue. The Database Management Systems guides back an answer to the request. For example, an end users, once dealing with large quantities of sales data, might want rapid answers to inquiries (ad hoc queries) such as:
– What was the dollar capacity of sales by product during the past six months?
– What is the sales bonus figure for each of our salespeople during the past six months?

1.1.6 Improved decision-making
Well-managed data and developed data approach make it conceivable to produce better-quality communication, on which better conclusions are constructed. The excellence of the information produced differs on the value of the fundamental data. Furthermore, Data characteristic is a inclusive methodology to endorsing the precision, strength, and suitability of the data. Although the Database Management Systems does not promise data importance, it distributes a schedule to allow data quality innovations.

1.1.7 Increased end-user productivity
The accessibility of data, collective with the tools that alter data into functioning evidence, inspires end users to make rapid, well-versed conclusions that can make the modification between achievement and disaster in the universal budget.

1.2 Framework Of Datababe Management System
Three levels Architecture recommended by ANSI/SPARC. It constructed a provisional report in 1972 trailed by a final report in1977 [2].

The goal of the three level architecture is to isolate the users' interpretation,

- It permits independent modified user views: All user should be capable to retrieve the same data, however have a different customized view of the data in database. These should be autonomous.

- It covers the physical storage details from the users.

- The database administrator should be capable to modify the database storage arrangements without influencing the users' views.

- The inner construction of the database should be unpretentious by alterations to the physical features of the storage.
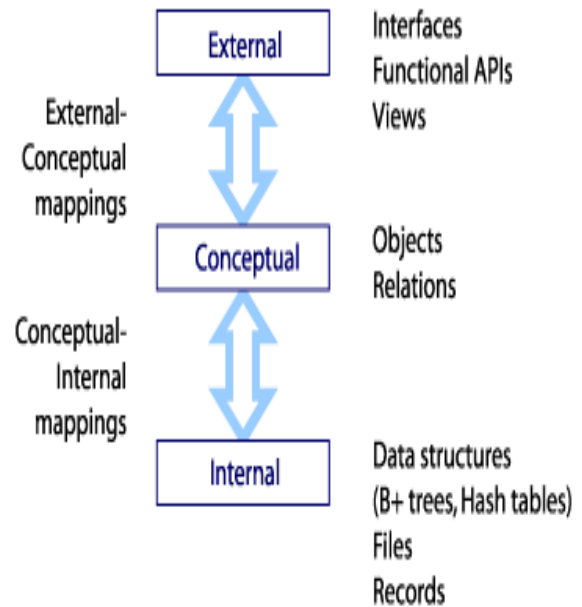

Fig 2 Levels in 3-tier architecture

The Three Level Architecture has the objective of allowing users to retrieve the same data but with a personalised view of it.

1.2.1 External Level:
A user's construal of the database designates a part of the database that is applicable to a specific user. It eliminates inappropriate data as well as data which the users are not permitted to methodology.

1.2.2 Conceptual Level:
The conceptual level is an approach of explaining what data is stored within the entire database and how the data is interconnected. The conceptual level does not postulate how the data is actually stored.

Particular valuable facts about this level are:

1. DBA works at this level.

2. Explains the structure of all users.

3. Only DBA can describe this level.

4. Worldwide view of database.

5. Autonomous of hardware and software.

1.2.3 Internal Level:
The internal level concerns how the database is actually denoted on the computer organization. It defines how the data is actually stored in the database and on the computer hardware.

1.2.4 Physical Level:
It is the fourth level of 3-tier architecture. It describes the physical storage of data.

2. DESCRIPTION ABOUT DATAWAREHOUSE
A data warehouse also recognized as an enterprise data warehouse is an arrangement used for reporting and data study.A data warehouse is a relational database that is developed for query and exploration rather than for transaction handling. A data warehouse poses historical data consequential from transaction data; however it can contain

data from additional bases. It split up analysis assignment from transaction workload and allows an involvement to organize the data from some different bases. In competes to a relational database system, a data warehouse background contains an extraction, transformation, transportation, and loading (ETL) clarification, client exploration implements, and additional presentations that accomplish the procedure of gathering data and providing it to professional customers [3].

2.1 Features Of Data Warehouse

The data warehouse is that quantity of an inclusive Architected Data Framework that performs as the single unified source of data for administering information. A data warehouse has exclusive appearances that comprise the following [4]:

2.1.1 *Integrated:*

The data warehouse offers one-stop shopping and comprises information about a diversity of businesses. A data warehouse integrates data from multiple data sources. For example, source A and source B may have different ways of identifying a product, but in a data warehouse, there will be only a single way of identifying a product.The data warehouse offers one-stop shopping and comprises information about a

2.1.2 *Subject-Oriented*:

Information is accessible according to particular topics or areas of concentration, not basically as computer records. Data is operated to deliver information about an individual topic. According an example, a specific database is not basically made manageable to end-users, however is delivered organization and systematized conferring to the precise essentials.

2.1.3 *Time-Variant***:**

Comprising an antiquity of the topic, as well as existing information. Historical information is an imperative constituent of a data warehouse.

2.1.4 *Non-Volatile***:**

Steady information that doesn't amendment each time an operative procedure is implemented. Information is dependable irrespective of when the warehouse is retrieved.

2.1.5 *Process-Oriented*

It is significant to observe data warehousing as a procedure for distribution of information. The protection of a data warehouse is continuing and iterative in environment.

2.1.6 *Data Mart:*

A data structure that is augmented for retrieve. It is created to enable end-user evaluation of data. It naturally supports a particular, organized application used by a different set of employees.

2.1.7 *Accessible:*

The main resolution of a data warehouse is to specify freely available information to end-users.

2.1.8 *Operational Data Store:*

An assortment of data that reports operative requirements of several functioning elements. It is not a section of a data warehousing framework, but a resolution to functional requirements.

2.1.9 *Staging Area:*

Every data store that is intended mainly to obtain data into a warehousing framework.

2.1.10 *Hypercube:*

A means of visually signifying multidimensional data in a framework.

2.1.11 *Star Schema*

A means of collecting data established on a set of recognized dimensions. It stows data multidimensionally in a two dimensional RDBMS such as Oracle Database.

2.1.12 *Snowflake Schema:*

An expansion of the star schema by means of applying extra dimensions to the dimensions of a star schema in a relational atmosphere.
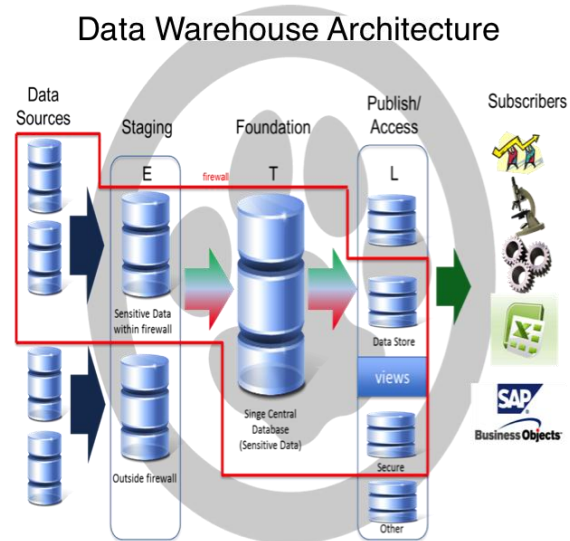
2.2 Data Warehouse Framework

Fig 3 Data warehouse Framework

2.2.1 *Data warehouse Framework:*

The model revealed above is the hub-and-spokes Data Warehousing framework that is prevalent in many administrations. In brief, data is stimulated from databases used in working systems into a data warehouse staging area, therefore into a data warehouse and conclusively into a set of imitated data marts. Hence data is copied from one database to another with a technology called ETL (Extract, Transform, Load).The ETL Technology is an imperative element of the Data Warehousing Framework. It is used to copy data from Effective Applications to the Data Warehouse Staging Area and from the Data Warehouse Staging Area into the Data Warehouse and conclusively from the Data Warehouse into a set of adapted Data Marts that are approachable by decision makers[3].

The ETL software extorts data, alters values of unpredictable data, and cleanses bad data, sieves data and loads data into a target database. The Data Warehouse Staging Area is provisional location where data from source systems is derivative. A staging area is primarily essential in a Data Warehousing Framework for timing argues. In brief, all needed data must be accessible before data can be incorporated into the Data Warehouse. Data in the Data Warehouse can be either persistent or transient .The intention of the Data Warehouse in the overall Data Warehousing Architecture is to assimilate commercial data. It comprises the single version of truth for the association that has been sensibly created from data stored in dissimilar internal and external operational databases. The quantity of data in the Data Warehouse is enormous[4]. Data is amassed at a very

gritty level of notify. Conflicting to prevalent view, the Data Warehouse does not comprise all the data in the association. Its persistence is to stipulate key professional metrics that are required by the association for calculated and planned conclusion making.

2.3 Data Mining

Data mining is the statement of compulsorily searching enormous provisions of data to determine patterns and trends that go beyond easy analysis. Data mining uses complicated mathematical algorithms to section the data and assess the possibility of future proceedings. Data mining is also recognized as Knowledge Discovery in Data (KDD). The authentic data-mining task is the instinctive or semi-automatic evaluation of large measures of data to extract formerly unidentified, exciting patterns such as groups of data proceedings, unusual proceedings, and needs. This typically includes using database procedures such as exceptional manifestations .The patterns can then be realized as a kind of summary of the input statistics, and may be expended in further evaluation. Data mining is usually expended in research ,businesses , and government defense .The inclusive goal of the data mining process is to obtain information from a data set and convert it into an comprehensible arrangement for additional procedure. Separately from the raw evaluation phase, it comprises database and data management features, processing model and implication deliberations, interestingness metrics, and complication reflections post-processing of revealed structures, conception, and online modernizing. Data mining tools calculate future trends and behaviors, accessing enterprises to make practical, information driven decision [5].

Data mining models can be functional to particular scenarios, given as following:

- Forecasting: Calculating sales, forecasting server loads
- Risk and probability: Collecting the finest customers for targeted mailings, formation the possible breakeven point for risk scenario, transferring, possibilities to diagnose or other outcomes.
- Recommendations: Identifying which products are probable to be sold together, creating and maintaining acknowledgements
- Finding sequences: Understanding customer selection in a shopping cart and assuming next same proceedings
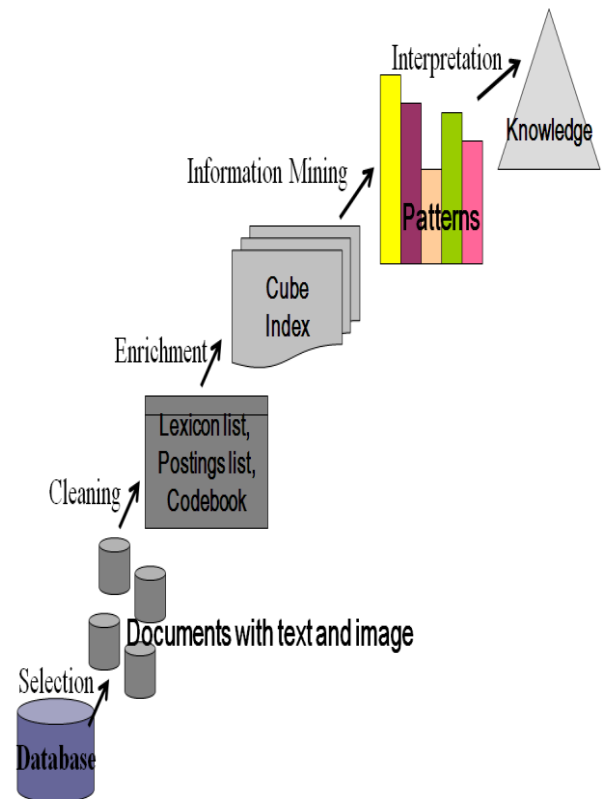- Grouping: extrications of customers or events into cluster of associated items, assuming and predicting affinity

Fig 4 Knowledge Discovery Process

Table1: Data Mining Procedures and Related Techniques

Operations	Data Mining Techniques
Predictive Modelling	1.Classification
	2.Value Prediction
Database Segmentation	1.Demographic Clustering
	2.Neural Clustering
Link Analysis	1.Association Discovery
	2.Sequential Pattern Discovery
	3.Similar Time Sequence Discovery
Deviation Detection	1.Statistics
	2.Visualization

2.3.1 Predictive Modelling:

Predictive modelling is a procedure used in predictive analytics to generate a statistical model of future behaviors. Predictive modelling influences statistics to predict consequences. Mostly the occurrence one needs to predict is in the future, although predictive modelling can be pragmatic to any type of unidentified occurrence, irrespective of when it happened. According to an example, predictive models are regularly used to detect crimes and identify doubts, after the crime. In many cases the example is selected on the basis

of detection theory try to predict the possibility of an outcome given a set quantity of input data. Models can use more classifiers in trying to regulate the probability of a set of data fitting to another set. Depending on definitional limitations, predictive modelling is identical with or basically overlapping with the field of machine learning, as it is usually indicated to in academic or research and improvement circumstances. When implemented commercially predictive modelling is often indicated to as a predictive evaluation. Predictive analytics is the part of data mining apprehensive with forecasting possibilities and analytics. In predictive modelling a data is composed for the appropriate predictors, an arithmetical model is calculated, predictions are completed and the model is validated as an extra data becomes accessible. The model may utilize a simple linear equation or a complex neural network mapped out by classy software. Predictive modelling is basically used in information technology. Predictive modelling is sometimes used to recognize the possibility that a given message is spam. Further applications of predictive modelling comprise customer relationship management change management and planning, security management, recovery, engineering, city planning and meteorology.i.e. risk examples can be established to unite member statistics in composite ways with demographic and lifestyle information from outward sources to progress endorsing[6]

2.3.2 There are two procedures related with predictive modelling:

classification and value prediction that are recognized by the nature of the variable being prophesied.

2.3.2.1 Classification in predictive Modelling-

It is used to create a precise scheduled class for each record in a database from a finite set of likely, class values.

In classification; all applicable instructions are applied and a score is calculated based on rule assurances and occurrences..In a classification problem, we normally have historical data and unlabeled example. Each labeled example consists of numerous predictor attributes and an objective attribute .The significance of the target feature is a class label. Furthermore the unlabeled examples consist of the predictor attributes only. The goal of classification is to create a model using the historical data that precisely predicts the class of the unlabeled examples. A classification assignment begins with construct data for which the objective values are known. Different classification algorithms use different techniques for finding associations between the predictor attributes' standards and the target attribute's standards in the construct data. Hence, these relations are summarizing in a form, which can then be useful to new cases with unidentified target values to predict target values. In a classification model the application of a new data is called applying the model and the data is called apply data. Furthermore; applying data is regularly called scoring the data. Classification is used in business modelling, customer segmentation, credit analysis, and many more another. There are two specialties of classification: tree induction and neural induction [7].
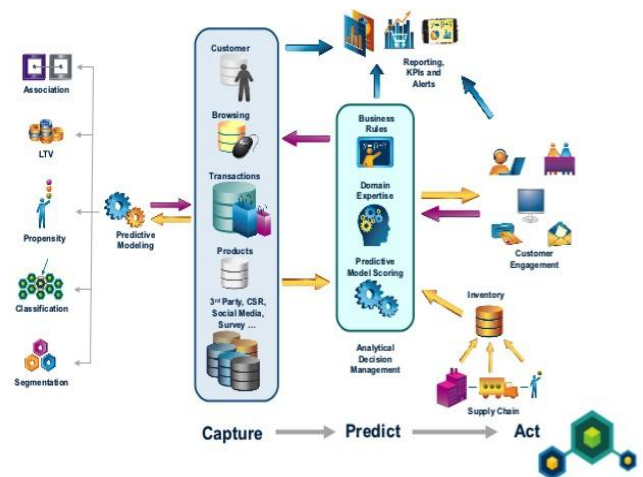

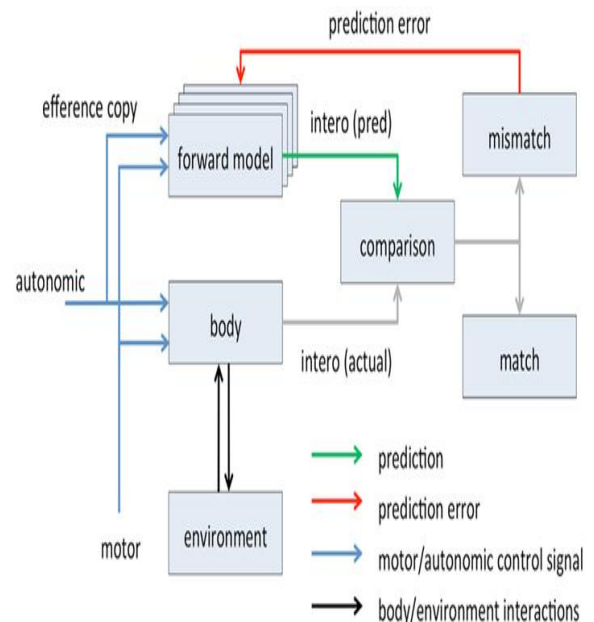**Fig 5 Predictive Modelling- Classification**

Fig 6 Predictive Modelling-Neural Induction

2.3.2.2 Value Prediction in Predictive Modelling

It is used to calculate a constant numeric value that is related with a database record and the conventional statistical techniques of linear regression and nonlinear regression. This technique is relatively easy-to-use and comprehends. Linear regression efforts to suitable a straight line through a plot of the data, that the line is the best demonstration of the average of all interpretations at that point in the plot. Drawback for this technique is that, this technique only works fine with linear data and is delicate to the existence of outliers. While nonlinear regression evades the main problems of linear regression, and it is still not elastic enough to control all conceivable shapes of the data plot. Data mining needs statistical methods that can contain non-linearity, non-numeric data and outliers. Presentations of value prediction incorporate credit card counterfeit detection or target emailing list association.

2.3.3 Database Segmentation

The words segmentation sometimes called clustering and both words are used interchangeably in data mining. Database segmentation is grouping a set of data objects into clusters. The aim of database segmentation is to partition a database into segments of analogous records that is records that stake a number of properties.It procedures unofficial learning to determine homogeneous sub populations in a database to recover the precision of the summaries. It is less accurate than other procedures thus less delicate to redundant and inappropriate structures. Sensitivity can be condensed by flouting a subset of the attributes that define each example or by allocating an increment factor to each variable. Presentations of database segmentation comprise customer describing, direct advertising, and cross vending. Hence it is a main feature of cluster that a Cluster is a collection of data objects that is similar to one another within the same cluster but it dissimilar to the objects in other clusters[10].

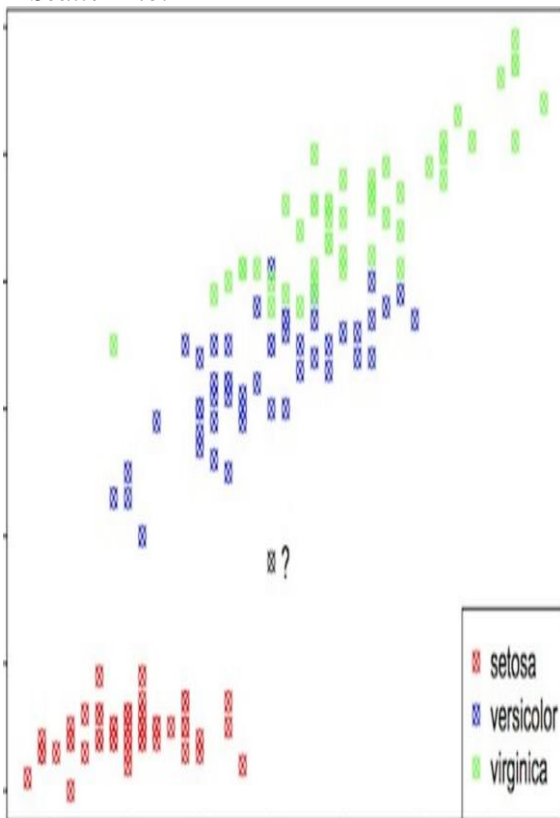### 2.3.4 Example of Database Segmentation using a Scatter Plot

Fig 7 Database Segmentation using Scatter Plot

2.3.5 An Association with demographic or neural clustering technique:

That is illustrious by Permissible data inputs, procedures and used to compute the distance between records, demonstration of the subsequent segments for exploration.

2.3.6 Link Analysis

Link analysis applies to establish associations between records and database. Applications comprise product similarity analysis, uninterrupted marketing, and stock value association.

2.3.7 Deviation Detection

In Deviation Detection, Comparatively an innovative operation in terms of commercially accessible data mining tools. Regularly a foundation of true innovation because it classifies outliers, which precise divergence from some formerly known anticipation and norm, it can be achieved using statistics and visualization. Applications comprise fraud detection in the use of credit cards, defects tracing, and quality control [9].

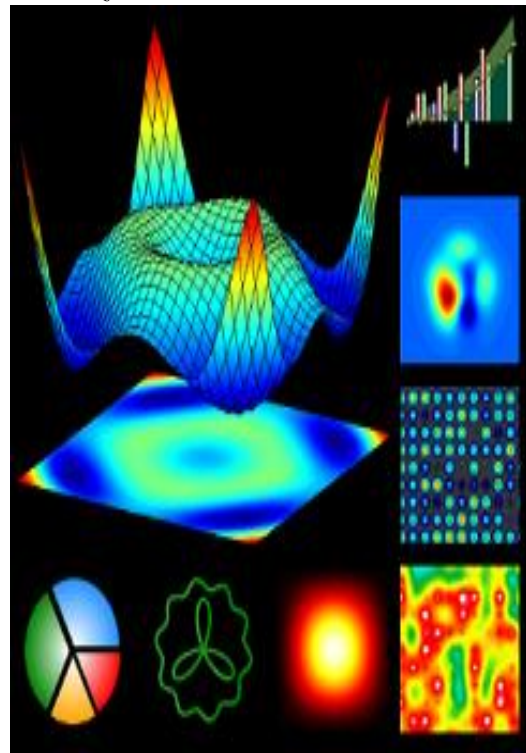### 2.3.8 Example of Database Segmentation using Visualization

Fig 8 Database segmentation using Visualization
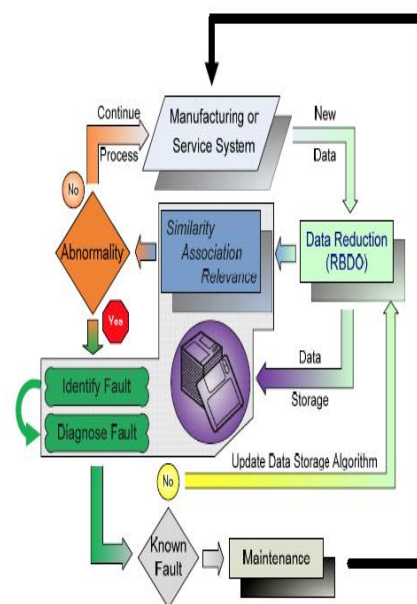
2.3.9 Data Mining Framework

Fig 9 Data Mining Framework

2.3.10 Tight Coupling

In the tight coupling data mining architecture basically three tiers is used.

2.3.11 Data Layer:

Data layer can be a database system or a Data warehouse system separately. This layer is an interface for all databases. Data mining outcomes are stored in the data layer so it can exist to end users in form of reports or other kind of conception.

2.3.12 Data mining application layer:

It is used to recover data from databases. Some conversion routine can be completed here to preferred organization. Then data is administered using numerous data mining algorithms.

2.3.13 Front-end layers:

It delivers instinctive and friendly user interface for end user to cooperate with data mining systems. Data mining outcomes are accessible in conception procedure to user in the front-end layer.

3. SIMILARITY BETWEEN DATABASE MANAGEMENT SYSTEM, DATA WAREHOUSING AND DATA MINING

A DBMS means Database Management System. A DBMS is the software that manages and handles the data on physical storage devices in database. This software delivers the proficiency to collect, retrieve and transform the data in a database. This software also offers a set of functionalities to observe and handle the presentations on those activities alongside the data in a database. Some Examples of DBMS would be SQL/Server, Oracle, Informix and DB2 in the relational database management systems[8][9].

A data warehousing is a framework that stipulates a vehicle to provide expressive information to the professional community to facilitate operative procedures omission and enterprise decision support systems. A big example of data warehousing is a social networking website that is face book that everybody can communicate to be what face book organizes. It collects all of your data –yours friends, comments, likes and your stalk etc., and then accumulates data into the fundamental depository [4].

Data Mining: A warm catchword for a class of database enterprises that observe for concealed configurations in a set of data in database. For example, data mining software can help wholesale companies find customers with common attentions. The tenure is generally distorted to pronounce software that offers data in new behaviours.A real data mining software doesn't just modify the arrangements, but truly realizes formerly unfamiliar associations between the data in database.Subsequently this data expended by anyone who wishes to mining the some valuable data narrates to field in database[5][6].

4. FUTURE SCOPE

Hence in this research paper, it is introduced to the concept for summary of database management system, Data mining and data warehouse with additional analysis about framework. Dbms is an imperative subject from employments point of view because it can act as a alternate for tough progrmaing languages. There are many achievable futures of a Database Management System.The tendency is similar throughout computing and may seem self marked, it has consequences for the future trends of database management system.There are many numbers of visions about the future of data storage. An ideal resolution of a Database Management System is that can relate the constraints of any individual model to the fundamental data, allocating obtainable framework to achieve at current levels while stipulating native and natural support for new models as demanded.On these days ,it is a buzz that there is not much scope for Data warehousing and data mining as Big data is impending very fast. Transforming everything to digital makes the significance of the data warehousing 100 times more. Data mining originates its name from the similarities between searching for valued business information in a largeDatabase.Any given databases of appropriate size and eminence, data mining technology can generate new commercial prospects by delivering the attainable proficiencies.However the verity is relational database is assumed by many as a main stream database system. The main scope of data warehousing and data mining is very good. Furthermore,there are several tools used for data warehousing but the most acknowledged ETL tool is Informatica.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Stack overflow, "Difference between a database and a data warehouse", http://stackoverflow.com/questions/3419353/what-is-the-difference between -a-database-and-a-data-warehouse

[2] Bcanotes,"DBMS",http://www.bcanotes.com/Download/ DBMS/Rdbms/database.pdf

[3] PrigrannerInterview.com,"What's the difference between data mining and data warehousing?" http://www.programmerinterview.com/index.php/database-sql/data-mining-vs-warehousing/

[4] Inman, W.H., Data Warehouse. John Wiley, 1992.

[5] Zentut," Data Mining Architecture", http://www.zentut.com/data-mining/data-mining-architecture/

[6] M. S. Chen, J. Han, and P. S. Yu. Data mining: An overview from a database perspective

[7] Zentut,"What is Data Mining", http://www.zentut.com/data-mining/what-is-data-mining/

[8] Navdeep,"Advantages and disadvantages of data mining "http://navdeep19.blogspot.in/2012/04/advantages-and -disadvantages-of.html

[9] Ramandeep Kaur,Amanpreet Kaur,Sarabjeet Kaur,Amandeep Kaur,Ranbeer Kaur," An Overview of Database management System, Data warehousing and Data", International Journal of Advanced Research in Computer and Communication Engineering,2013

[10] Srinivasan Parthasarathy, Data Mining overview. srini@cse.ohio-state.edu.