4.1 Studies Related to Data warehousing

A number of researchers have conducted research in data warehousing and different tools of data warehousing.

Ajit et al (2011) discussed about analyze historical data in to the data warehouse in the Traditional way. Now a day the organization used data warehouse for supporting the features of real time decision making process for the day to day life of the organization. The authors stated that the improvement of function of the organization in warehouse for fire any query and also update of data is a challenges of a data warehouse. They suggested to use better query and making the data or information in to the warehouse in proper stored for function of data warehouse specially when we have entered new information in it. Now a day modern organization always used different level of system to manage their day to day operation. These systems some time call operational system and assemble different hardware and software from the different vendors for function of it. The combination of the types of system creates a complex system for the organization especially for the EDP manager and Analyst. The authors give the options to the end users that instead of complex system they can use data warehouse software for solved the integrating data in the operational system. The main purpose of data warehouse is that the information can be stored in to the data warehouse for business analysis and used operational system for that. The basic reasons to used operational data in to data warehouse that the analysis of data and implementation of new system with new data into data warehouse in essay. Finally the authors stated that basic principals of data warehouse system for the organization to build the successful system for analysis of
data and making the decision for the organization and also they proposed new way to build the new system for the organization.

**Anshuman (2011)** stated that the data of data warehouse can be used for different purpose such as operational and traditional and it also comes in different source it can consist different sensitive information for the organization, these information can used by the organization for decision making and also used for analyzed the data for development of the organization. The author said that these information required security for prevention of data which can be access by unauthorized end user. The authors also explain the different methods of security of data of the data warehouse; these methods can be used for the organization in different level such as business level, conceptual level, logical and physical level. The author show the different strategies for the security of data from the data warehouse for the organization, they suggested different security solutions for the data warehouse, they provide the solution for different level of business such as logical and physical level, business level, as well as conceptual level of the data warehouse design and implementation for the organization. The security of data warehouse can be implemented at the access level when the data warehouse is going to design for the important data for unauthorized access by the end user. The authors suggested that the security of data warehouse should be important for the organization after the implementation of data warehouse in to the organization. The EDP manager or system analyst always takes the responsibilities for the security of important of data of the data warehouse for the organization.

**Ashadevi (2012)** discussed about the materialized views with relations of different source. They stated that when designer of data warehouse start the design of that they are facing different problem in selection of materialized view. The authors give the different survey and different methodologies for selecting of materialized view and also how they can use effectively in data warehouse. They also give different methods of critical analysis such as evolutionary optimization methods and features of this method. When designer start designs the desired data warehouse they are face general problem called materialized view selection and also selecting appropriate sets of view for the end users.
Now a days there are different research issues for researcher in data warehouse regarding selection of one and most important challenge. The authors suggested data structures for the desired view selection for the data warehouse designer. For this purpose the designer can use relational algebra, directed acyclic graph etc. The designer can use different methods to reduce the cost of answering queries because for the organization point of view it is very large and complex and also it is use for decision support system into short time. The author also suggest to solved this problem through the help of different types of constraints such as aggregation and grouping constraints, currency constraints, as well as time constraints etc. Finally, to further improve of selection of materialized view the designer can used cost model formulation and critically analyzed into the present and past methods to solve the desired problem.

**Kamal et.al (2009)** discussed data warehousing with statistical mining for the different organization. They authors suggested that the data warehouse methods can be used for preprocessing step to implement different schema for prediction of data mining method and also prediction of data mining result. The data warehouse technique can also used for correlation between different analysis of best system approach and develop best model for the organization. They approach new model called marketing model for study the different approach of the marketing for the system. The data warehouse can be combine of different statistical data mining method for different plat form of the data mining. Finally the authors gave the best way for modeling the relationship that show different sales respond of various advertisement and also used regression methods to compared about the system. The authors used local patterns which are used between the modeling process of the system and also behavior of the system. The data warehouse can be combine the suitable statistical data mining methods to represents an affective system for data mining for different level of organization.

**Kishori et. al(2011)** discussed about role of data warehousing and data mining in E-governance. They have define the E-Governance as being accessible electronically to all end users with publically of relevant data or information form different department of government, the government use all system of which is used
electronic resource and different technology for their business and share all information to the end users. In the e-governance they user ICT called information and communication technology for access all information and as well exchange the data and information, and services to different department to the citizens and business. This information or data or resource can be done by the government from anywhere of the world. The Governments have lots of information and data they store all information in such a way to access all information in effectively for decision making for the any organization. The e-governance provides the facility to implement any queries for analysis the data for decision making process for the organization. The authors focus of different scope and utilities of data warehouse and data mining in e-governance such as Government to Citizen (G2C), Government to Government (G2G) as well as Government to Business (G2B). The functionality of Government can be improved by the Data Warehouses and Data Mining. Finally the authors define the data warehouse such as way it is a collection of IT based information and data which is successful implementation in different department of government. It supports different tools for access and extracts information from government department for all level of information as well as complex query and also provides quick, accurate and meaning information for the all end users. The e-governance user data mining for different query and generation of reports. The data mining provide the facility for end users for implement queries and different level of reports generation. It also supports systematic way to fine the hidden information and also help out the client behavior. The basic use of data mining and data warehouse in e-governance is to user different techniques for decision making process for the organization.

Khan et al(2011) discussed about cloud base data warehouse Data Warehouse plays a important function in organization for any decision making. The different environment supports an extremely beneficial decision making platform and smooth function and maintenance for the organization. The version of data warehouse 2.0 support enhancements approach that help the organization to organize their data in different physical boundaries. The Data Warehouse is used for strategic decision making for different level of organization. When data is more in the size, then we can devolved different model and different shape for decision making for the organization and also it helps us to make decision for past and feature strategy.
It required more data storage device to implement the data warehouse for the organization. The system analyst can use classical architecture of data warehouse for de-normalized different format with help of different traditional entity-relationship modeling methods. They can also use different storage methods which are introduced by Ralph Kimball for data modeling and star schema. When we want to maintenance of such a big data repository it is required hard work and good knowledge in the database. It is required divide and conquer responsibilities for maintenance of such data and also used split these data in different section and cloud them in maximum output for the organization. Finally the authors in this research paper suggested a model with explanation to that how we can store data and split it and also what type of data we can store in local and server for the cloud database. The author also suggested the way for the organization and top level executive to maximized the data analysis and stay any type trouble free from the database and make the decision for the organization.

Meenakshi et. Al(2011) discussed about Schema Evolution for Data Warehouse with consideration of core component of decision making process for the organization. The all organization used data warehouse for major decision making and it also support function of daily transaction of the organization. It can be used at the time of design and implementation of data warehouse in dynamically and subjectively. The authors suggested that this dynamic and subject nature this can be reflect the different evaluation of data warehouse. They also suggested that evolution of data warehouse always focus on schema evolution, schema versioning and view maintenance for the organization. It can always change according to data and their structures of the data warehouse. The schema can be changed according to structure of data warehouse, software as well as end user of data warehouse and types of organization such as small, medium and large organization. They suggested that schema evolution in data warehouse can be combine different level such as structural level, conceptual level, and behavioral level of the organization. The authors main focus on schema evolution and also compare different approaches of the schema. Finally the authors propose different models for creation and evolution of aggregate table for schema for the organization with the deferent level of the organization such as structural, conceptual, and behavioral level of schema.
They define the different criteria such as dimension updates, hierarchy updates, instances updates, fact updates etc for the schema evolution for the different level of the organization. The requirements of end user the schema will be change and according to types of organization the schema versioning and schema maintenance scheme will change.

**Pushpa et. Al(2011)** discussed about object oriented data warehouse with help of narrower data model. We can offloaded data from the operational system for any purpose such as report generation. They suggested that the end user can transfer all data or selected data by the operational data store for different purpose in to the data warehouse for report generation for the organization. We can combine the object oriented data warehouse with different source of underlying data. They also suggested that the narrower compassed model can be implemented in data warehouse for object oriented. The oracle 10g specially can be used for this purpose. The object oriented data warehouse support different complex objects such as different types of user define and automatic object type. The different algorithms can also be used to maintain data for insert, delete and update operation in data warehouse as well as source database. Finally the authors gave main focus on relational data models for data warehouse with the help of object oriented data warehouse and also oracle 10g programming language can be used for database management and other purpose such as insertion, deletion and updating of data in object oriented data warehouse. The different model can also be used for views the performance and security of data warehouse.

**Sankalap et.al(2011)** discussed about An Efficient Indexing Technique Used In Data Warehouse for the organization. They stated that the data warehouse is a house where traditional data and information is store and managed in different numerical format. When the end users store textual and non transactional information then data warehouse loss these data and information. The data warehouse define by the different end user and different types of end user work and store data in it, then they encounter different text and number into the data and information. For complex query, different updation such as delete, insert and update operation require indexing
techniques as the time of design of data warehouse instead of tradition data warehouse. In the data warehouse it is compulsory to that it should be flexible and efficient queering on the data or information for extracting the information and generate different level of report for the organization. They authors stated that indexing technique for data warehouse for the textual data and information. When we use indexing technique on textual data it can increase the efficiency of the data warehouse for the end users. They also focus on Full Text indexing technique and also do some observe the result on different size of data and information on the time and space complexity. Finally the authors stated that the indexing techniques are use full in text base data and information with the help of clustered indexing on the large data warehouse and database. The full text indexing methods is one methods that can be used to reduce the storage of database, index maintenance and also reduced the administrative cost for the organization.

Srinivasa et. al (2010), discussed on hidden stored of unstructured data. The unstructured data is used in critically in decision making for the top level manager, making relationship between other business function etc. The authors also suggested text mining method to management of unstructured and semi structured data such as XML and MML for retrieve and analysis of data and also text analysis. It can help to predict business results with help of business intelligence and data warehouse. The authors put attention on text mining which is use in documents and also discover different hidden patterns of text. The text analysis can also used to enhance data retrieval into different form such as search and enabling clustering into text mining. The authors stated about unstructured data which is used into business application for analysis data and find solution in the future. The some freeform test create complicate into query, searching data, extracting information as well as integration of other data. The authors give the idea about test analysis which is used full in decision making and also break the limitation of unstructured and structure data for implement transaction based application. Finally, the authors leave one important point regarding text mining, structure data can use to unstructured data by text tagging and also with annotation which is used in integration with other data.
Ashadevi and Navaneetham (2010) stated new technology which is use to access information on timely, accurately, it can also use for management information system for decision support system for the top level manager of the organization. The authors specify the main aim of data warehouse to select appropriate set of views to reduce total cost of materialized views. The authors also stated about the framework of selection of problem in view which is used reduce the cost of low maintenances, and combine the different query results. The cost is based on query execution frequencies, and base relation update frequencies etc. This framework can also use to constraints in storage space. The authors discussed about important issue about views of materialize of data warehouse. They specify the cost effective views for materialization through new framework. This can also be use to maintenance of query, storage cost of query, as well as query processing cost for optimization of data warehouse of the organization.

Mutaz (2011) discussed about different challenge to implementation of data warehouse in the desired organization. The authors also classify and find out the different challenges about it. The authors give more importance in practical as well as theoretical implementation of data warehouse into the organization. They stated about qualitative research methodology to find out the different challenge of data warehouse in the organization. The authors also give the idea about data analysis and storage and utilized from bottom-up methods which is given the complete concept of different technique of data analysis. There are different challenges into data warehouse for the organization such as organizational, technological as well as managerial issues. The authors also discussed about different subject related to data warehouse in this paper such as team skills, architecture, enterprise schema, data integration and data quality, security of data, risk management and network management for the organization. When designer start the design of data warehouse it is very difficult to design as well as it is highly risk cover in the success of data warehouse. It’s required lots of amount in investment, times in development, as well as also requires technical and managerial skills of the staff in the organization. And some social aspects and shaping technology in designing of data warehouse.
When data warehouse designer used latest technology and social context then it can be success of data warehouse design and implementation of the organization.

Ramakrishnan and Padmanathan (2011) stated that Data warehousing project failures not appropriate for managing every situation or even every kind of project. Project management does apply to great many situations. It is not only in large scale infrequent undertakings that project management applies. It’s in all kinds of smaller more frequent activities as well. As long as the activities are directed to somewhat different unfamiliar goals that require co-operation from several parties’ project management might be appropriate. The chaos is basically Greek mythology and very old concept which is use to give the world preceded cosmos information. The authors define it is a dynamic systems which is generated by deterministic way and does not have any noise source and basically provide the unique solution in mathematical problem. There are various aspects of the software development divided into three categories, People-related, Process-related, and Product-related. Software engineering flaws assume many forms, including absence of a standard software development methodology, inefficient/flawed execution of solution steps, accidental omission of solution steps due to hasty schedule, failure to follow to raise sufficient human and other resources. Each of these flaws can induce disorder in software development process. Failure to identify and remove the root causes of the chaos inducing factors at the point of their origin itself can bring disorder in the system development. As chaos increases, various system development activities get derailed. The system development process then enters a normally hidden zone of chaos and the system development fails.

Umm et. Al (2011) discussed about Enhanced Architecture of a Web Warehouse based on Quality Evaluation Framework to Incorporate Quality Aspects in Web Warehouse Creation for the organization. The authors stated that the world wide web become the main information resource in different aspect and all areas of interest. In the web relevant information extraction is become very difficult for any one and also it is time consuming for the end users. The authors giving advantage of data warehouse function and improve the retrieving information from the web site for the end users.
The web warehouse can be developed to store relevant information or data and give the opportunity to end users to extract information for them. The main function of web warehouse is to extract and store the different information for the end users for any organization. This can be used for decision making of different dimensions for the organization. The authors provide the complete architectures of web warehouse and also discussed about it. They suggested that if any existing architecture is available the end user can enhance the current system and develop quality evaluation framework. The end users can add three different layers to existing system to enhanced architecture for the organization such as source assessment, query evaluation and data quality layers. Finally authors suggested that during the web warehouse creation phase the report from one phase can be used input for next phase, so it can increase the quality of web warehouse. They also suggested the enhanced architecture of web warehouse for quality of warehouse and proposed the new architecture for the organization. The main architecture of a web warehouse is used to enhanced the embedding the different layer such as source assessment layer, query evaluation layer and data quality layer. Data quality layer is used for quality of data before uploading it into the system. Source assessment layer is used for checking validity, relevancy and other quality attributes of different source for the warehouse of the organization.

Vishal et. al (2011) discussed about the feature implementation of data warehouse in the organization. Now a day the data warehouse is the most important technology for the different types of organization for planning, management, and forecasting any policy and making the decision for the organization. The authors basic aim how to maintenance the data warehouse for the different level of organization. When the end user not taking attention to maintenance of data warehouse the software will be failure and the organization not able to take any decision for future so every user must take attention for maintenance of data warehouse. The other system such as operational systems need not require more attention for maintenance but the data warehouse it require qualify staff, good management team and also support from others staff from the organization to maintenance of the data warehouse. When organization implement data warehouse system the need staff attention about data
extraction, data loading, network management, training and communication within staff, different level of view and others things for the organization. The author’s main focus on impact of selected factors, dimension and application for the organization. Finally authors stated that the data warehouse system is not only a application software but it is complex process to developed IIS for the organization. When end user adopt the software it required mass capital expenditure, time and other things. The different factors must take into consideration when organization want to implement the data warehouse. The ETL design must be carefully before the adaptation of system. The end user also understand the basic requirement of software, types of business and different types of data require for the data warehouse. It is not only reporting software but it is also way to handle the business, thee need and requirement of the organization. The software provide fast view of the report so end user must aware about the functioning of data warehouse. The end user must understand the view materialization strategy for different query.

4.2 Studies Related to Data Mining

Asth a et. Al (2012) discussed about data mining technique in cloud computing, they suggested that the data mining can be use for extracting the data from the database and also useful information from data. They can also be used for extracting the data form large data volume and provide data into different techniques for data management. They support different technique such as clustering, classification, neural, genetic etc for support to find the hidden information from the large database. They suggest that the cloud Computing can be used in web-based technology for access different resources and also help them to shared information for the different department of the organization. When the business is large and large database management is required the cloud data center can be form for the organization. The cloud data center provide the data in very low cost and also data mining and cloud computing technique can be access data from them. The both technique help the end users to cut cost and also maximized the profit for the organization. They also suggested different way to implement data mining technique such as Google App Engine and Cloud SQL.
Finally they suggest that K-Means algorithm for clustering algorithms for analyze basic problem for the organization. It is very efficient technique for large database management and also different level of the organization.

**Christoph et. al (2011)** discussed about data mining driven manufacturing process optimization for the organization. Now a day very high completion in different organization in globally specially in manufacturing company. They always try to take efficient, effective and continuously development process in particular sector for that organization. The manufacturing company some time uses existing analytical process and also provides the execution systems for better improvement for the organization. They can use all technique in very short time in considering limiting continuous process of improvement. Especially the manufacturing company did not used data mining system for find out the hidden information and data for the company. The authors stated that the indication based and pattern based manufacturing paces of optimization is the most important for the advance manufacturing company for different analytical information and decision making process. They used use cases and find out the suitable data mining methods and also implementation of data mining system for the organization. They suggested indication based manufacturing optimization as best data mining system for advanced company for analytical process and decision making process. Finally the authors give the information about indication based optimization and OLAP system for different purpose such as extracting data and decision making. Some time company used holistic data the company used data mining as use case to identify the hidden information and data for the organization and generate the different level of reports. They also give information about metric oriented dashboard in different industry for find the analysis of presents indication of process and also used in improvements or the organization. They introduced the further use case for indication based optimization and define the formalization methods for manufacturing optimization patterns and develop different optimization methodology for the particular organization.
Pandey et. al (2012) Discussed the attracting the students to get admission in Higher Education Institutions (HEI). The authors provide the solution of student’s admission which is start from contact students with the help of different level of advertisement. The authors also state different types of advertisement methods which can be used by the institution for getting students admission. The institution can make advertisement by different methods because of different constraints and also due to time constraint and money constraints. The authors suggested for the institution to take the confidence method of advertisement so the students will get admitted in the institution. The beginning of study the institution directly contact to students without taking any extra effort and using time. But now a day it is very difficult to institution to directly contact to student because of finance and money management of the institution. The authors suggest about the marketing of the institution for attract student to take the admission in the institution, the institution can show their product, infrastructure and quality of education to the consumer. The authors also suggested about term enrollment management and strategic planning which is provide the very specific direction to the top level manager of the institution to prepared clear chart of the course of the institution. In this paper author give different methods for HEI forget admission of students, such as:- Data collection, Data Analysis, Support and confidence Analysis, Cosine analysis etc. Finally, to implement any method for get admission of students in HEI, try to use cost factor associated with each advertisement which help an HEI to make an economic and effective advertisement method.

Kodinariya and Seta (2012) stated that to build the Data Warehouse that containing all the information related to election to increase awareness of voting. Using this authors suggested different interesting patterns that are extracted and represented using Visual Data mining to arrange awareness program. The Authors specify different approach such as Data Preprocessing, Data Warehouse Creation, Task-Relevant Data Extraction, Data Mining and In Data Warehouse Creation phase, warehouse with vote as measure in fact and voter Gender, voter age, voter education, candidate, Religion, time, session, word as dimension. In the proposed system, authors focus on only Data Cleaning and data Transformation. Some of information related to voter and candidate are missing, which was represented by constant “UNKNOWN” in data cleaning process.
We can obtain result of all these rules through cube editor but it is not possible to represent all these rules at once, to do so we need to use the visualization techniques like Pixel-Oriented, Icon-Based, Hierarchical, Graph-Based and Geometric Technique. Visualization techniques are used mainly for explorative analysis, confirmative analysis and presentation of facts. Technique of visualization which is appropriate is pixel oriented technique as ours is an 8-dimensional data. Each dimension of the cube can be represented as a pixel. Visualization of the output generated using visual data mining techniques facilitates decision makers to make decisions like fluctuation in voting; voting ratio of male to female; voting rating of different party in a city/state. This type of analysis helps to arrange appropriate awareness program in different regions.

Acharya and Madhu (2012) discussed about the knowledge discovery in the different filed such as mathematics, AI and statistics. It helps the user to extract the knowledge about the system. The authors also discussed about different phase of knowledge discovery. It can be used in KDD for extracting information from different patterns and also with correlated application which is always store in data set. The authors specify the features of knowledge discovery in database which is used to analysis the data and predict the future plans and area where it can be used in the organization. The authors define the system which is used to analysis and prediction of placement of the students based on the different historical data from the database and also generates different result about the placement for the students and top level management. It can also specify the different rules for the placement of the students and it will very benefit to them. The authors specify the role of knowledge discovery and data mining application in different fields such as extracting information and generate different result from the very big data sets. They also developed new model for the placement department and also prepared the database for them. This database can be use by institution for finding the different patterns of placement and analyzed them and make a future plan. Finally to further development of this model they can include features of dynamic update of the database by the end user and also apply different algorithm and patterns on desired data. And data mining can also used it with some constraint and generate different constrains for the end users.
Kundu (2012) discussed Web usage patterns are an important aspect to discover hidden and meaningful information. It will be big challenge in web mining when the volume of traffic is large and the volume of web data is still in the growing phase. The author highlights different challenge an intelligent approach of web traffic analysis. There are different processing stages of mining normally data collection, data preprocessing, and pattern discovery and pattern analysis. The data source selected in this approach is from the Web traffic data generated by the ‘Webalizer’ Web access log file analyzer. The author specifies the model which is used to obtain accurate analysis and meaningful information. This model also used to remove the irrelevant and noisy data at the beginning step from the processing task. It is also used to access data which is sorted with higher value and last access data is always store in bottom with lowest value. The designer point of view this process in very complicated due to time variant and variables dependence on web pages. Finally, to further improve of web mining system to incorporate cluster diagnostics and verifcation tool with the help of agent communication language which will helps in making the web mining system not only efficient but also an intelligent system.

John and Meshram (2008) discussed about the prediction of awaited monsoon in the Indian sub continent. The monsoon plays a deep impact on our country economic and social development so the prediction of monsoon is important. The authors specify the onset of monsoon prediction which is based on data which is extracted from satellite. The data may be image or text. They develop model for monitoring and depth studying of monsoon prediction which is effect globally. The data or image is always pictorial representation of electromagnetic radiations which is collected for the satellite. The data may also be infra red, far infra red and near infra red etc. The authors use data mining technique in prediction of monsoon with some algorithm such as K-nearest neighbor algorithm etc. The authors proposed model which can be predict monsoon before 10 to 30. They match all distance of country and predict the monsoon. The authors also discussed on different subject which is affected the monsoon onset such outgoing long wave radiation and quantitative precipitation estimate etc.
Batra and Alam (2012) discussed the Privacy-Preserving Data Mining (PPDM) which has become a significant subject in most recent years. Generally privacy means “keep information about person from being available to others” but, the real worry is that their information not be mishandle. The data mining techniques enable users to extract the hidden patterns which may lead to leakage of sensitive data. So the main concern is to secure the data mining result with the help of PPDM. In this paper authors provides a framework to preserve privacy in data mining results by manipulating SEMMA analysis cycle. Privacy-preserving data mining (PPDM) refers to the technique that provides shelter to insightful information from unwanted or unauthorized revelation. Privacy preservation key objective is to protect data records from any type of disclosure. PPDM provides high quality information to end users without exposing personally identifiable information. PPDM employs a variety of practices to alter either the original data or the data generated using data mining methods. Organizations must consider different dimensions before designing their privacy policies and access control rules. These dimensions are: Purpose, Data categories, User groups, Actions, Obligations, Data object context. Finally, authors suggest a modeling method of PPDM by SEMMA methodology. PPDM techniques help in ensuring the quality delivery of data mining results to the end users without affecting the one’s basic details. SEMMA methodology is an analysis cycle; it filters the datasets from large pool of records and also discloses hidden patterns that help in achieving data mining objectives.

Mande et. al (2011) discussed the different technologies and investigation about map of the criminal. This has given scope for the development of new methodologies in the area of crime investigation using the techniques based on data mining, image processing, forensic, and social mining. In this paper, authors present a model using new methodology for mapping the criminal with the crime. This model clusters the criminal data basing on the type crime. When a crime occurs, based on the eye witness specified features, the criminal is mapped. In this research paper authors propose a novel methodology that uses Generalized Gaussian Mixture Model to map the features specified by the eyewitness with that of the features of the criminal who have committed the same type of the crime, if the criminal is not mapped, the suspect table is checked and the reports are generated.
Generalized Gaussian Mixture Model (GGMM) is used for crime data values intensities in the entire crime data obtained by crime data values intensities is a Random Variable and follow a Finite Generalized Gaussian Mixture Distribution. Finally, to further improve of crime mapping required more research tried to identify the criminal by mapping the criminal using the Generalized Gaussian mixture model.

Mala and Dhanaseelan (2011) discussed about efficient knowledge discovery and its applications. The EKD is useful in traffic modeling, online data processing etc which is store lots of data stream or information. The authors also specified the EKD and future prospective in research with the help of data mining technique. The authors also differentiate between traditional data and data streams which are used for high speed data access and also store large amount of data for the processing. The authors develop new model of data mining rules for data stream. It require multiple scans of entire data sets can not put in one stream instead of that use one stream of data at a time and wait for another requested of data stream. It increases the processing speed of the system and also reduces the response time. This model requires mining frequent patterns as well as association rules of data stream. The authors stated different point regarding the design of stream data association rule of mining technique and also they discussed about latest trends about the technique. For fourth development of existing model which is specified by authors require user friendly mining techniques for the end user.

Sankari (2012) discussed Ant Colony Optimization (ACO) algorithm has evolved as the most popular way to attack the combinatorial problem. The ACO algorithm employs multi agent called Ants that are capable of finding optimal solution for given problem instance. In this paper, the author proposes new updation mechanism based on clustering techniques, which is aimed at exploring the nearby solution region. They also report in detail the impact on performance due to integration of Data mining, cluster and ACO. The author specifies the ant colony optimization algorithm which is always depends on agents for natural behavior of ant such mechanisms of cooperation and adaptation. Data clustering is one of the most important human activities which is used to underlying data such as discovering groups and pattern/classes of objects in such a way that the objects within the group are very
similar and the objects across the group are quite different. STING (Statistical Information Grid) is a grid based clustering technique. The grid based approach divides the spatial area into rectangular cells. The rectangular cells are partitioned recursively and hierarchical structure is employed to represent them. The hierarchical structure represents several layers of rectangular cells and each layer represents the data at different level of resolution. Finally, to further improve of ACO require research in two important directions. They can be ant miner with continuous attributes and second investigate the performance of heuristic function of strategy.

Vydehi (2012) discussed different framework of mining web log files. The authors stated different challenging of real life web usage mining such external data and user profile. The web usage mining is the process for data mining technique which is use to find the usage patterns of web data. It can helps the end user to get information from user profiles and extract information and fire different query to implement of web log data. The authors also give the latest information about web usage mining such as academic and industrial research and also commercial activities. This technique provides panoramic view of mass usage of modes and also provides specific information for the end users. The authors specify different objective of this research such as validation strategy, quality of mined profiles as well as behavior of techniques. The authors discussed different activities of the research such as view paged, search engine queries, and develop different inquiring for the organization. To improve the access of dynamic web site it is require using scalability and webbing click stream for data stream and also do some mapping of new session and updating the profiles.

Jain et. al (2011) discussed about security issue for database. We are storing large information into database so security of data is most important issues for the database administrator in client server technique and unauthorized access. The authors develop privacy preserving data mining model for privacy of data of statistical database and data mining. They suggested association analysis method for find out the relationship which is always hidden in big database. The model can use data modification and hiding rule for security of data. The authors suggest different objectives of this model such as proposed association rule for data hiding, association
data mining rule of algorithm. This algorithm is used to increase or decrease the support of different rules. The association rule always work with left hand side or right hand side items which is used to increase the association rule mining algorithms. The author’s proposed model of algorithm is always compare with ISL and DSR algorithm for check the hiding of rule for accessing the database. The basic features of data mining are the extracting hidden rule for data from database. The authors specify the different objectives of this research such as preserving data mining; hide some confidential data with help of data mining techniques. This algorithm is basically depends of hiding approach of previous algorithm. Finally to improve the functions of algorithm develops an integrated secure association rule and protect data from unauthorized access. The research work can also be done on clustering, fuzzy set and classification tools.

Kalra and Gupta (2011) discussed about data mining and use of data mining as a tool. The data mining provide important data from big database for the top level manager or executives for making decision in effective manner. The authors specify the features of data mining are analysis tools for collect raw data from database and produce the valuable information for the manager. They also specify the functioning policy of data mining especially in banking and risk management area. It can also be use in market segmentation, direct mail marketing, and fraud detection etc. The banking sector can be use data mining technique to find out the hidden pattern of group of data and also use to find out the relationship between them. The data mining in banking sector can be use in following field such as, market segmentation, customer churn, trend analysis and direct mail marketing etc. Now a day banking sector found different competition so they can use this technique to fight more creatively to gain market shares. The authors stated that it can be use by banking sector to get knowledge of their excellence. Data mining can be used in various fields of banking like Market segmentation by which banks can segment their customers into different groups, direct mail marketing can help the banks to improve their marketing strategy and to increase their business, customer churn to increase the rate of retention of the customers, risk management to reduce the various risks like creditworthiness and fraud detection to reduce the number of fraudulent.
Data mining has wide application domain in almost every industry where bulky data is generated and that is why it is consider as one of the most important and promising developments in Information Technology.

Ramachandran et. al (2012) discussed about effective methods of classifying process of knowledge data discovery (KDD). This process is used decision tree algorithms in different field such as medical and store medical data and also analysis the medical data for future use. The advent of newer and easy to use technologies, compact storage devices and development of software indirectly boosted the development of data mining. Data, which is the base of data mining, started following at higher rate and ignited the need to find newer pattern and store them instead of huge data set. These patterns produced further lead to pattern mining. Classification has been identified as important problem in data mining. Data mining on medical data can help in simple classification to highly accurate predictions. The advantage over using classification on medical data would be to get over all idea of the data based on various attributes, So that the complexity can be reduced and detection of anomalies becomes easier. Cancer data has higher complexities due to various types of cancer and various method of diagnosis. In this paper authors give the find pattern using classified data. There is further ways to find cancer pattern to avoid it spread which medical is a miraculous outcome. Cancer is the disease that has been for years and yet incurable due to lot of medical needs and lack of technical advancement. But in this age of higher technology and greater potential data mining can be helping had only if it can help not only finding newer pattern but using predictive methodologies to predict the future work. Data mining has evolved from an experimental analysis to a predicting methodology due to highly precise algorithms and high performance data mining tools. Knowledge discovery in database has been used to predict survivability and diagnosis of diseases in the field of medicine. This can prove helpful for prevention of epidemics. Finally, to improve finding pattern of cancer they have to future work on this can be suggested that not only cancer pattern in southern part of India but cancer pattern around the world can be deeply review and researched to find out any anomalies or cause of the cancer or at least determine the flow of it in the way.
Sankari. A (2012) discussed about data mining and clustering using ACO algorithm. This algorithm can be used for solved any problem as well as combination of problem. This algorithm also call multi agent algorithm because it can be used more than one program or modules in the system at a time. They suggest that ACO algorithm can used for iteration process, solution converges to find out the optimized the solution for the organization. The author suggest the update mechanism based clustering to solved the problem in exploring way and give the performance in the system due to integration of Data mining, cluster and ACO. The author gives novel methods to update the pheromone trial. The update methods can be combine with cluster methods for give better performance for the organization. The Anit-Miner technique is give the different classification of data and behavior of the real life problem and also give the complete concepts and principals of data mining for the end user. Finally the author discussed about continuous attributes of ant-miner and different performance of update mechanism for decentralized of data for the organization.

4.3 Studies Related to OLAP and OLTP

Reddy et. al (2010) discussed about different tools of data warehouse and data mining such as OLAP and OLTP. They also are exploring the features and application as well as architecture of data warehousing. The OLAP is basically used in data warehouse to improve the performance of the system. In the traditional database management system use OLTP for functional requirement on the operational databases. The authors also specify the features of OLAP such as interactive analysis on multidimensional data with different tools. Both are used for decision making by the top level executives and basically focus on database industry. The manager, executives and analysts always use OLAP in market oriented data analysis for the decision making process. The authors also stated about functioning of OLAP such as it is leading technologies for storage of data, database management and desired extraction of data. Basically both are complementary of each other. Where data warehouse manage and store the desired data and OLAP doing transforms of data from data warehouse to produced meaningful information. Both are used for calculation, browsing, and navigation of complex modeling.
They authors also specify the characteristics of OLAP which is enables to managers to use less time and make appropriate decision for the organization.

**Bagdi and Patil (2012)** discussed about online analytical processing and it used in medical and healthcare industry. The traditional database management system does not support analysis features for medical industry. The authors stated that the OLAP also can used in healthcare and medical industry for collection of raw data and produced the valuable information for that industry. It can also be use for advance technologies and data mining for the decision making process by the top level management. This research paper authors suggested DSS with help of OLAP and data mining which help the top level manager to predict the future condition of medical industry and also use useful information for decision making. The authors specify the utilities of doctors and patients in the model, the doctors can used data mining to predict patient’s problem and the OLAP can give the information about patient’s historical data for making the appropriate decision about the patients. Both can use in clinical decision support system to make the decision for health industry. The doctors can use OLAP and data mining into existing process and produce the suitable patterns for the patients. This system more useful in find out the hidden patterns from the desired data and also increase real time indicators and visualization of information. Further work can be done to enhance the system.

**Hassan (2011)** discussed about backbone features of decision support system with help of data warehouse and on line analysis processing and also many of others applications. Working to discover important information and appropriate depending on the analyzing results of the multidimensional database. In this paper author give different fragmentation feature which is the most important that characterizes the Distributed database system (DDBS) to improve OLAP performance during fragments the Multidimensional database (MDDB) which it the most popular data model for DW. The fragmentation will be depending on the geography situation. The author specify the function of data warehouse, it can be analyzed the data with help of on line analytical processing and provide the facilities to find out the trends, patterns of behavior and also find hidden dependencies of data. Hybrid fragmentation uses a combination of horizontal and vertical fragmentation to generate the fragments we
need. The mixed fragmentation was implemented on the multidimensional database. Horizontal Fragmentation on the Countries Dimension table and Vertical Fragmentation on the other Dimension tables. If user want to analysis the different query then they can use the on line analytical processing to find the solution. Oracle 9i can be use to implement all features of OLAP. Finally, to improve the system performance which is depending on OLAP based model especially the E-business, the fragmented technique can be used for multidimensional database depending on the fragmentation feature which is the most significant of the distributed database. After running the OLAP query, they founded the efficiency of the system increased whenever the amount of database increased and thus it reflected on efficient the making right decision for the future work.

4.4 Studies Related to Multidimensional Data Warehousing

Moudani et. al (2011) discussed on problem with MOLAP, which is always load big tables into main memory. It reduces the system performances and increases the cost of model. The authors introduced the new model which is called BTC which is used in multidimensional data warehouse. Basic used of BTC to increase the access time of data; it is also based on B. Tree. The OLAP and data warehouse which is also called analysis tools which is used to find the appropriate solution through business intelligence. The multidimensional data model is based on dimension, fact, and OLAP operators which is used to analysis big data with the help of multidimensional model. The authors specify the different characteristics of multidimensional style which is always use to access data very easy without implement any complex query. The authors also specify the features of hypercube with the help of MOLAP to select data into intersection of dimension; it is also use to decrease the dense or scarcity of cube. The authors develop new model namely BTC (BTreeCube) for access multidimensional data into data warehouse. It access different big relational database and increase the process f cube. Finally to solve the problem of binary tree and get balanced binary tree to reduced the complexity of order O (log n) which his use for appropriate solution.
Singh et. al (2011) discussed about the design process and technique for data warehouse, which is develop for big organization. The authors specify the existing Meta model for quality management of data warehouse and architecture of them. The authors also define the special process and different operators for quality control. They specify the different application of model on the evolution of data warehouse. They stated that data warehouse is basically used in daily life of business for the organization, it used to OLTP system for decision making. The authors provide details information about Meta model with their framework to quality control of database management system. Finally data warehouse maintenance provides up to date information for the end users and also give status of data warehouse. To improvement of function of data warehouse some research require in details of different types of data warehouse, and also refined and extend the new aspects of data warehouse process.

4.5 Studies Related to Distributed Data Mining

Kavitha and Sasipraba (2011) stated numerous current data mining tasks can be implemented effectively only in a distributed data mining. Thus distributed data mining has achieved significant importance in the last decade. The authors’ proposed distributed data mining application framework is a data mining tool. This framework aims at developing an efficient association rule mining tool to support effective decision making. Association Rule mining focuses on finding interesting patterns from huge amount of data available in the data warehouses. In order to build strong association rules, it depends on the extraction of association rules by Apriori algorithm, AprioriTID algorithm, AprioriHyprid algorithm, FP growth etc. The efficiency of the distributed data mining framework is determined based on the selection of the algorithm. The object oriented implementation has enabled the system to be platform independent. The use of self defined database format gives an upper hand for the system by operating efficiently without any need for third party database drivers. The mined results can be compared and graphically projected. Finally, concentrate on the performance factors while working with large number of algorithms, databases and clients.
Decision making would be easier by increasing the modes graphical representations such as bar graphs, pie charts etc. By adding additional data mining models like clustering, classification etc., can be converted into a full-fledged data mining framework for mining real large databases.