

Impact of Data Warehousing in Decision Making: Pre and Post Covid-19

¹Jignesh P. Shah and ²Dr. A.J.Patel

¹Research Scholar, Department of Statistics, Gujarat University, Ahmedabad, Gujarat (India)

²Head & Associate Professor, R. H. Patel Arts and Commerce College, Ahmedabad, Gujarat (India)

ARTICLE DETAILS

Article History

Published Online: 15 September 2020

Keywords

Data warehousing, data mining, decision support system, Staging Layer, Data Marts, Operational Data Store, Knowledge Discovery.

ABSTRACT

World environment which is now post Covid-19 give users to access their data, but it does not solve all the problems of user. Every user has started using data online approximately more than 90%. So users must have privilege to access the data but do not guarantee the integrity of the data and adequacy of response time. So in this Pandemic situation Huge Data warehousing and Big data solve the above problems and provide technology which enables the user or decision maker to process the huge amount of data in a short amount of time. With the help of big data warehousing, user extracts the data as knowledge based in a real time. Common people, Students, and many companies want to use that data for other own purposes and has no other solution and has to depend on data only. So data mining techniques are evolved for extracting new knowledge from data warehouse. Data warehousing and provides the right foundation for building decision support. Different system tools which help to measure the progressing speed of data usage now has become organization's goal and it now is the duty for ISP (Internet Service Provider) to help companies to reach towards their goal.

Data warehousing and provide a technology that enables the user or decision-maker in the corporate sector, education sector, government sector etc to process the huge amount of data and make decisions which are useful for whole organisation. This paper will help to explore the overview, advantages and disadvantages of data warehousing with suitable diagrams. In this paper, roles and responsibilities of organizational members of data warehousing are also discussed. As a concluding point, how "Date Warehouses" can be used in different areas for different users, how these data help in decision making and allow the data manager to perform more accurate, substantive and consistent analysis.

1. Introduction

The concept of Data Warehousing in pre Covid-19 was very popular as a business information management tool where it is expected to disclose knowledge structures that can guide decisions in conditions of limited certainty. The usage of data before Covid-19 was slightly limited to the Business only, but lot of software teaching and learning methods were about to become popular. Such data were available in data warehouse supports [1]. Now, decision-making

methods in post Covid-19 has become a challenge for data managers. By creating an enterprise-wide integrated database, it integrates data from multiple, incompatible sources. By transforming data into meaningful information, and a data warehouse allows the manager to perform more substantive, accurate and consistent analysis. The figure below shows the difficulties faced by the data mangers for Huge Data Warehouse Storage.

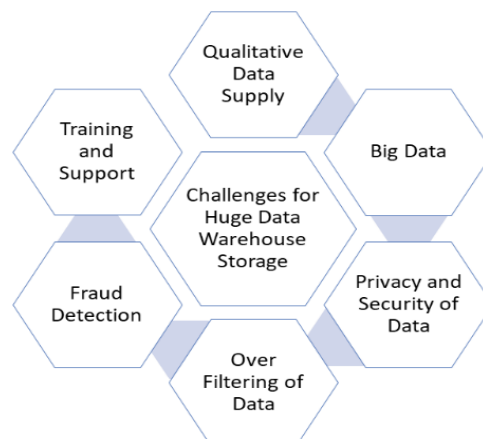


Figure: Decision Making Trouble

The data warehouse is not the normal database, as we understand the term “database”. The main difference is that the traditional databases hold operational-type most often, transactional type data and that many of the decision-support type applications put too much strain on the databases intervening into the day-to-day operation (operational database). A data warehouse is of course a database, but it contains summarized information. Data warehouse refers to database that is maintained separately from an organization operational database. Normally, Data Warehouse holds read-only-data. Data processing, also called Knowledge-Discovery in Databases or Knowledge-Discovery. Through Data processing, the extraction of hidden predictive information from large databases, works as a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data processing tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. Data processing tools can answer business questions that traditionally were too time consuming to resolve.

environment which tries to satisfy the customers by supplying the required data. Data Warehouse plays a vital role by providing the ability to the organizations to perform information processing effectively, strongly, and securely. It is a collection of integrated, subject-oriented databases designed to support the DSS function, where each unit of data is non-volatile and relevant to some moment in time. Numerous roles and responsibilities will need to be acceded to in order to make data warehouse efforts successful and generate return on investment. For the technical application programmer, system administrator, database administrator, it is recommended that the following roles be performed full-time by dedicated personnel as much as possible and that each responsible person receive specific Data Warehouse training. Such kind of training is must after Post Covid-19. The data warehouse team needs to lead the organization into assuming their roles and there by bringing about a difference between Pre and Post Covid-19. Management also need to make actionable plans out of these directives. Following are the team, team members and their responsibilities to make data warehouse make effective and helpful to user and organization:

2. What Data warehousing (DWH): Post Covid-19

A data warehouse is a collection of integrated databases designed to support a DSS. Data Warehouse is an

Team Member (Pre Covid-19)	Role of each Team Member before and after Covid-19
Manager	The data warehouse manager ensures support for the data warehouse program at the highest levels of the organization and understands high level requirements of huge data. Principles for data warehousing has changed in Post Covid-19 which is maintained and managed by the manager.
Project Manager	Project managers deliver commitments on time. Project managers maintain highly detailed plan and caring about progress on it. Matching team member’s tries to issue list of tasks to Chief Architect.
Architect Designer Manager	The Manager of data warehouse will need to rely on a Architect Designer Manager as one of his/her direct reports, to work on complex issues of architecture, modeling, and tools. Architect Designer would have significant interface with the internal clients and increase their confidence in the data warehouse organization in this Covid-19 pandemic situation.
Final/End User	Data warehouse is made to meet end users requirements. Data warehousing is used to answer the end users queries and generate reporting. End user receive ID and password on the data warehouse system and provide feedback to the data warehouse team like performance, functionality, data quality, meta data quality and completeness.
Database Administrator (DBA)	Database administrator (DBA) works on division of roles and responsibilities between the support group and the user community. Database administrator has many responsibilities like database maintenance, backup and recovery, data replication, Performance Monitoring and Summary table creation. It also becomes a responsibility for maintaining Backup and Recovery very fast during Covid-19.
Application Programmer Specialist	The Data Warehouse Application Programmer is responsible for applying transformation rules as necessary to keep the data clean and consistent.
System Administrator	Responsibilities of system administrator are: Installing and maintaining the Database Management, monitoring the performance, architecting the data warehouse architecture. The Data Warehouse (DWA) System Administrator is responsible for the performance of data transfers and post a query for data replication or synchronization

Data users are trying to use Huge Data from Data Warehouse (DWA) during Covid-19 requires following data support techniques:

Integration: All inconsistencies regarding naming convention and value representations must be removed.

Nonvolatile: Data must be stored in read-only format and do not change overtime because huge data usage required non-volatility.

Time Variant: During Pre Covid-19, Data were not current and were normally time series. But during Post Covid-19, Data is both Current and Time Series.

Summarized: Pre and Post Covid-19 have Operational data that are mapped into a decision-usable format only.

Large Volume: After Covid-19, now a days since number of users have increased a lot, Time series data sets are normally quite large amount.

3. Architect and Working of Data Warehousing (DWA)

Data Warehouse are designed in such a way to overcome the problem of inaptness of Informational and Operational Transaction System. The EDP and MIS are designed to satisfy the frequently, incompatible requirements of the business. Data warehouse is a database used for

reporting and analysis. It is a place where data is stored by integrating different Data bases. It can be used for storing current and historical data. With the help of historical and current database new prediction can be drawn. The following diagram shows different compounds of data warehouse.

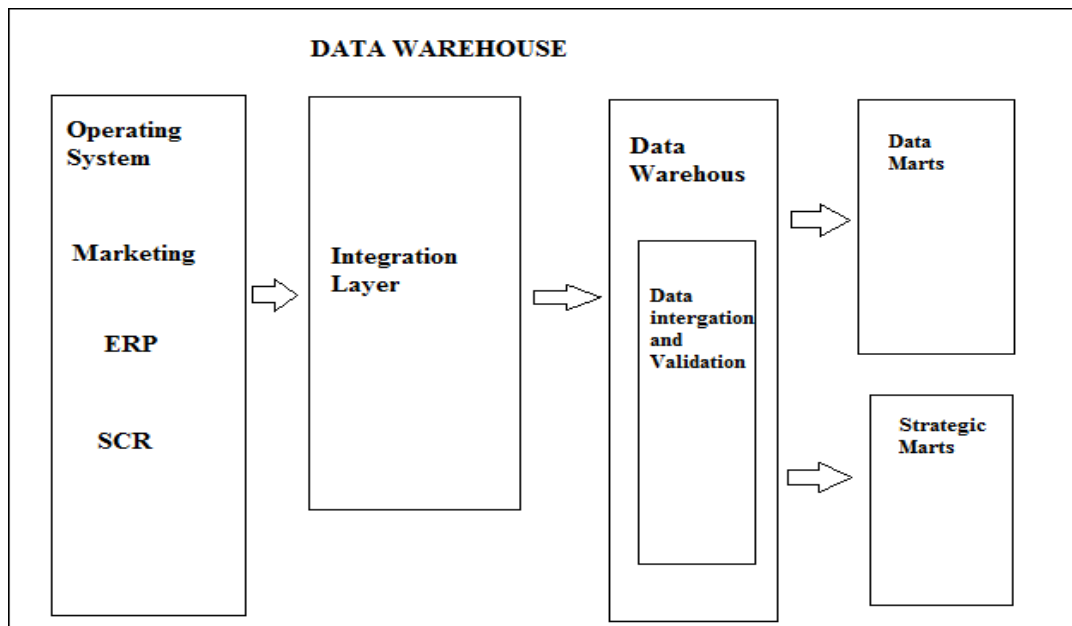
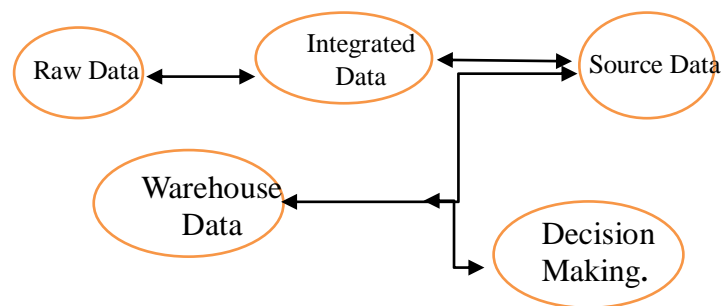


Figure: Components of Data warehouse (DWA)

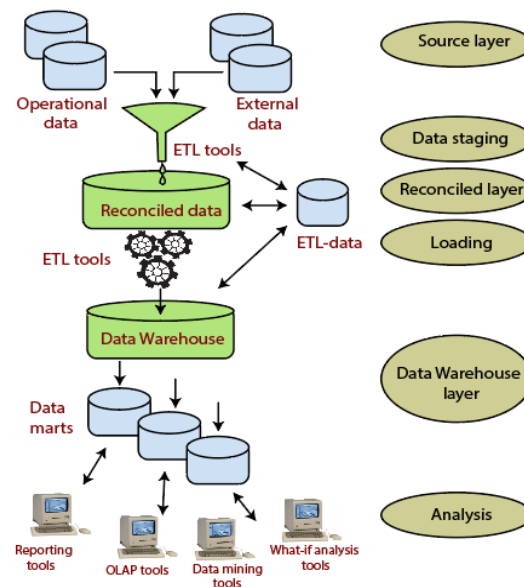
Data warehouse is a database used for reporting and data analysis. It is a central source of data which is created by integrating data from one or more different sources. The data stored in the warehouse are received from the operational systems. The staging layer stores raw data collected from each of the different source data systems. The integration layer integrates the disparate data sets by transforming the data from the staging layer often storing this transformed data in an operational data store database. A Data Mart (place from where data can be purchased) is a

small data warehouse concentrated on a specific area of interest. Every new user during Covid-19 now uses Data Warehouses, so Data Warehouses can be subdivided into data marts for improved performance in use. The company can have one or more data marts towards a larger and more complex enterprise data warehouse. The Data Warehouse saves time of business user and helps to generate the reports quickly. Education users and Business users (Commercial users) can quickly use these reports on one place and can take decisions quickly.



Business users and Education users won't waste their precious time in collecting data from multiple sources. They all during Covid-19 will try to take help of data warehousing, can query the data themselves and saves money and time because during pandemic situation every one was told to work from home. So, in this situation the pressure of using

data warehouse has increased tremendously, and hence the reliability was everything on DWH only. The following figure shows the operation of data by the users of different kind from home during pandemic situation. Different levels of data churning from Data Warehouse are shown.



Three-Tier Architecture for a data warehouse system

4. Data processing

The data processing applications are available on all size systems for Main frame, client/server, and PC platforms. Data processing means it is a process that aims to use existing data to invent new facts and to uncover new relationships.

Data processing on the above said situation includes several steps:

- problem analysis.
- data extraction.
- data cleansing.
- rules development.
- output analysis.
- review.

Data processing sources are typically flat files extracted from on-line sets of files/application, from Data Warehouses or other datasource. Data may however be derived from almost any source. Users during Pandemic situation use data only on-line. So, 95% of data was to be used only from Data Warehouse. Whatever the source of data, data processing will often be an iterative process involving these steps:

1. Uniqueness of Data

Before user starts using data, he/she be clear on what hope to accomplish with the analysis. Know in advance the business goal of the data processing. Establish whether or not the goal is measurable.

2. Choice of the Data

After user defines his/her goal, the next step is to select the data to meet this goal. This may be a subset of users Data Warehouse that contains specific product information. It may be your customer information file.

Here are some key issues like:

1. How current and relevant are the data to the business goal/education goal?
2. Is the data stable—will the churned data attributes be the same after the analysis?

3. Compilation/Processing/Churning of the Data

Once you've assembled the data, you must decide: -

- a) Which attributes to convert into usable formats?
- b) Which input of domain experts/creators are there?
- c) Which users are there of the data?
- d) Establish strategies for handling missing data.
- e) Find extraneous noise and outliers.
- f) Decide on a log or square transformation.
- g) Determine the distribution frequencies of the data.

4. Evaluate the Data:

Evaluate the structure of your data. What is the nature and structure of the database? What is the overall condition and distribution of the dataset?

5. Choice of Appropriate Tools:

Two important factors for the selection of the appropriate data-mining tool business objectives and data structure. Both should guide you to the same tool. No single tool is preferred to answer the queries.

6. Prepare indented the Solution:

Find out the answers of some questions like: What are the available format options? What is the goal of the solution? What do the end-users need graphs, reports, code?

7. Prepare the desired Model:

When the data process begins, user splits data into sets, construct and evaluate the model according to their requirement. They will use data depending upon their work/interest. The generation of classification rules, decision trees, clustering sub-groups, scores, code, weights and evaluation data/ error rates takes place at this stage.

8. Check and Validate the Findings:

Then comes how to and when to share and discuss the results of the analysis with the business client or working experts or extreme user or domain expert. Ensure that the

findings are correct and appropriate to the all types of objectives. Find out the answers of many queries like:

- Do the findings make sense?
- Do data match the requirement of user?
- Do data churns properly?
- Do data give proper results?
- Do data organize result properly?

9. Reporting the Findings:

Prepare a final report for the business unit or client. The report should document the entire data mining process

including data preparation, tools used, test results, source code, and rules. This report helps in decision making and plays important role in the growth of organization.

10. Combine components to integrate the solution:

Share the findings with all interested end-users. User might incorporate the results of the analysis into the company's business procedures. Although data processing tools automate database analysis, they can lead to faulty findings and erroneous conclusions if administrator is not careful since data usage is very heavy.

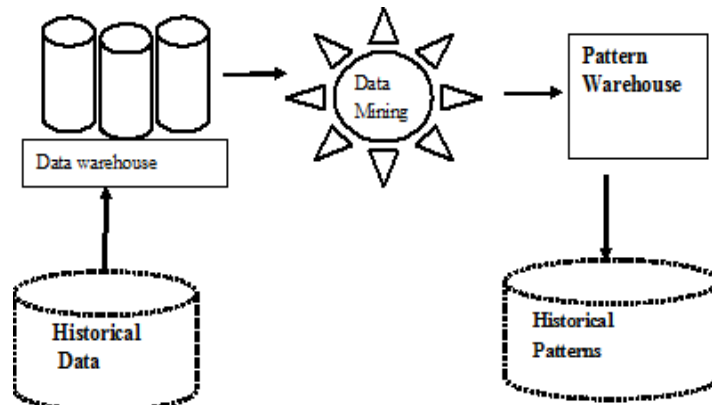


Figure: Working of Data processing within Data warehouse

Data processing can be applied to operational databases with individual transactions. Both private and public sectors such as education, banking, insurance, pharmaceutical manufacturers, health care providers, manufacturers of sanitizers and mask, PPE kit manufacturers and retailers are using data processing for a variety of purpose to reduce costs, enhance research, predict the effectiveness of a proper medicine, security, accuracy and increase in sales. Data processing is used to predict future trends, customer purchase habits and help in decision making. Online marketing using this Data processing improved company revenue and lower costs.

Data processing is also used in analyzing the market and find out the frauds. Since during Pandemic situation, since everything has to do from home, Data processing have privacy and security issues. During Covid-19 situation, Data processing sometimes is costly at implementation stage. Data processing along with privacy issues also has an issue of misuse of information due to huge use of data. Data processing sometimes cannot promise perfect results, cannot explain why an outcome occurs, and cannot help in correcting problems in your data. User may sometimes get absurd results, fraud cases, and data hacking.

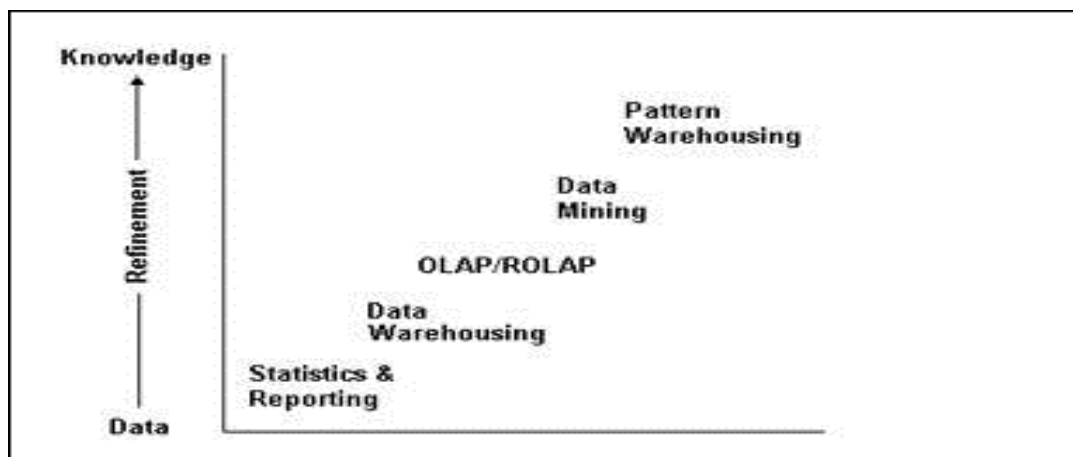


Figure: Redefining of Knowledge and Data representation technique during Covid-19

5. Role of data warehousing and data Processing in Decision

Due to huge use of data from Data Warehouse, it was not the problem to design or architecture of client-servers,

but it has been additionally introduced new demands on the environment itself. The goal of a Data Warehouse is to support decision making with data. Data processing can be used in conjunction with a Data Warehouse to help with certain types of decisions. To be successful, it needs a

skilled user who can supply the correct and accurate data with lot of security and a specialist who can make objective conclusions to show the output. Wrong supplies of data with incorrect or minimal amount of information, output will be affected and forecast will not be credible. Data Warehousing in combination with data processing plays an important role in decision making of the organization. Data Warehousing provide answers of many queries to the organization and the user to help them in decision making. There are many types of queries of the organization:

- tactical query.
- strategic query.
- update query.

A tactical query: is a database operation that attempts to determine the best course of action right now. It may happen that the data handling and processing techniques like Object oriented techniques, Middle ware, Storage and Handling of multifaceted data may be used. A tactical query provides information to rank and file elements in the field that need to respond quickly to a set of unfolding events. Tactical queries tend to produce a very small result set and usually such type of result set is designed to fit into a single window on a display screen.

A strategic query: is a query that provides the information which is necessary to make long term business decision. Such query is a database operation that attempts to determine what has happened, why it happened, and/or what will happen next. It typically accesses vast amounts of detailed data from the warehouse and ranges in complexity to simple table and works on multi-way joins and sub queries. Applications that generate strategic queries include:

- report generation
- OLAP
- decision support
- ad-hoc

An update query: is a database operation that modifies the state of a database. It updates the data according to the requirement of user. Tera data provides a set of bulk load utilities used to load large quantities of data into the database in an efficient fashion.

6. Conclusion

Data Warehousing and its processing talks about the change in business trends during this Pandemic situation. All the small and big industries are collecting and using data from various sources to identify their own business trends. During this Pandemic Covid-19 situation, Organizations understand the strengths and the weaknesses of their competitor. So they try to improve their progressing speed towards the goal and expand their business over the Data Warehouse and their processing type.

A Data Warehouse is the only solution to a business problem which has become technical problem. It constantly requires overcoming obstacles that are yet undefined and help the organization in decision making and improves the goodwill of organization. Data processing helps in securing and processing the data in to understandable chunks Data

Warehousing helps in analyzing the data in such a way that it facilitates comparison between trends and analysis of the data for the business predictions and accelerate decision making. Data Warehousing and Data processing both together includes the conversion of data from various source systems into a common format with accuracy, helps the organization in taking strong business decision in order to expand the business.

A Data Warehouse:

- Enhances Consistency of data
- Maintains Data Quality
- Standardizes the data
- Will produce results that are in line with all the departments
 - Provides relevant and organized data in an efficient manner
 - Helps in providing data from different locations that can be combined in one location.

7. Future scope (Post Covid-19)

Data processing offers an important approach to achieving values from the Data Warehouse for use in decision support. Post Covid-19 will make Data warehousing as a standard part of an organization and there will be efforts to find new ways to use the data. Data warehousing and Data processing will bring several new challenges in future like:

1. Regulatory constraints may limit the ability to combine sources of disparate data.
2. Security constraints of huge data storage and usage.
3. These disparate sources are likely to contain unstructured data which is hard to store. So requires more data integrity for converting unstructured data to a structured data for future use.
4. The internet makes it possible to access data from virtually "anywhere". So, internet will play an important role in future Data Warehouse and Data processing.
5. This just increases the disparity.
6. Today the challenge is to design Data Warehousing and Data processing applications that are reliable, easy to use and supports effective decision-making.
7. Data Warehousing and Data processing will become a valuable tool in industries/business as the amount of data increases in the future.
8. Data processing will be helpful in finding new quality data, and can help the optimize use of sales resources like manpower and marketing.
9. High performance commercial parallel computing and largest database (VLDB) processing will be used on large scale.
10. The system ability to handle, store and manipulate a variety of data Eg: Video, text, graphics, images, special time series data etc. will be used on large scale with object-oriented database systems, as if they were traditional data types.

References

1. Prof Jignesh P Shah, "The strategic role of Data Warehousing Technology in Decision making", *International Multidisciplinary Research Journal (RHIMRJ)*, Vol-2, Issue-1, 1-3, January-2015, ISSN:2349-7637.
2. EdwinMKnorr and Raymond T.N.(1998), "Algorithms for Mining Distance-Based Outliers in Large Datasets", *Proceedings of 24th International Conference on Very Large Data Bases*, York, USA.
3. Brachman R. J. and Anand. T.(1996), "The process of knowledge discovery in databases: A human centered approach", chapter 2, 37-57. AAAI/MIT Press.
4. Wirth, R. and Hipp. J.(2000)," CRISP-DM: Towards a standard process model for data mining", *The 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 29-39, Manchester, UK.
5. Monika Pathak, Sukhdev Singh, Sukhwinder Singh Oberoi, "Impact of Data Warehousing and Data Mining in Decision making", *International Journal of Computer Science and Information Technologies (IJCSIT)*, Online, 995-999, ISSN:0975-9646.
6. Frawley W., Piatetsky Shapiro G. and Matheus C.(1992), "Knowledge Discovery in Databases: An Overview", *AI Magazine*, Fall, 213-228,.
7. Daskalaki, S., Kopanas, I., Goudara, M., and Avouris, N.(2003) "Data mining for decision support on customer insolvency in telecommunications business", *European Journal of Operational Research*, Vol. 145, Issue. 2, 239-255.
8. Chen, C. and Lewis, B.(2002) , "A basic primer on data mining, *Information Systems Management*", 56-60.
9. M.J.A. Berry, G. Linoff, "Data Mining Techniques: For Marketing, Sales and Customer Relationship Management", second ed., Wiley, New York, 2004.
10. Berry, Michael J.A., and Gordon Linoff (1997), "Data mining techniques: for marketing, sales, and customer support", New York: Wiley.
11. Barry, D.(1997), "Data Warehouse from Architecture to Implementation", Addison-Wesley.
12. Fayyad U., Piatetsky-Shapiro G., and Smyth P.(1996), "Knowledge Discovery and Data Mining: Towards a Unifying Framework", *proceeding of 2nd International Conference on Knowledge Discovery and Data Mining*, Portland, pp.82-88.
13. N. R. T., Han J.(1994), "Efficient and Effective Clustering Methods for Data Mining", *International conference on Very Large Data Bases*, Santiago, Chile, pp.144-15.