Understanding the Concept of Data Warehousing and Challenges in Its Implementation

Monika Soni
President, Auricle Global Society of Education and Research,
Bikaner, Rajasthan (India).
president@agser.org

Abstract

The aim of this paper is to understand the concept of Data ware housing and how it is implemented. It is related to the data analysis of the data in an organisation. It facilitates and makes the analysis process easy for the workers of the organisation. The paper will also explain two approaches that are followed in data ware housing. The process of implementation of data ware house will also discussed further in this paper. There are certain challenges to create data ware house.

Keywords: - Data ware house, Process, Benefits, Challenges.

Introduction: - Data ware house is the concept of analysis of data and is used for reporting the data in an organisation. The data is collected and stored from various resources and made available in a ware house which can be used by the workers. The data ware house helps to store the old data as well as new data. The data can also be modified. There is a lot of information in an organisation which comes from various sites, it is important to refine and store it in such a way that it can be used in future. It provides the facility to study the relationship patterns in the data and helps to report. There are two approaches which are used in data ware housing: - [1]

➢ ETL Data ware house approach: -

In this approach Extract-transform- load method is used to store the data and made available for the users. First of all, the data is extracted from various sources using the staging layer. It is then cleansed and arranged in proper form. Then with the help of the integration layer, it is integrated and stored at one place. The data can be arranged based upon its type into either listing form, hierarchal form. The access layer provides this data for the users need. Hence in this approach, the data is extracted from various resources, then it is cleansed and integrated in proper forms and the made available for the use of workers in the organisation. They can perform reporting and analysis of data for market research, for studying patterns between data etc.

➢ ELT Data ware house approach: -

In this method, Extract-load-transform technique is used. First of all, the necessary data is extracted from all the sources which is required for the business. It is then uploaded in the data ware house as it is in the same form as it is extracted. Then within the data ware house the data is cleansed, refined, and then arranged in the proper manner in tables and rows and columns etc. It can then be used by the users to serve their daily needs of reporting and analysis.

Characteristics of Data ware House: -

➢ Specific Data: - The data ware house helps to store information related to specific subjects. It is not related to the business data but data linked to specific subject which can be used to analyse it for business requirements.

➢ Time effective: - The data stored in the data ware house can be used for studying the different trends in the market. With the help of it, the time used to extract and read data from different resources is less. It provides all the necessary information at one place which can be accessed at any point of time.

➢ Integrated data: - The data extracted from different sources is cleansed and then integrated according to the category of the data in the data ware house. The integrated data which is sorted on the basis of the type of data is easy to access.

➢ Intact data: - As the data ware house can store the old data as well as the newly added data, the data will always be available for the business study at any point of time. It gives the facility to access the historical data as well as the newly added data.

Data ware House Implementation life cycle: -
There are 8 major steps followed to implement the data ware house: -

1. Requirement Analysis: -
In this stage the designer of ware house will understand the business goals and aims so as to understand which type of data should be stored in the ware house. There are many departments in an organisation. Hence, the data ware house designer should sit with each department present in the organisation so that he can understand the goals of each department individually and then he creates the requirement specification which will explain in detail the goals of the business.

Aim of first stage: -
➢ Documentation of all the requirements of each department of business.
➢ Understand the goals of the business organisation.
➢ Understand the security threats to the ware house.
➢ Finding solutions in case of ware house failure.

2. Software and Hardware decision: - Based on the result of the first stage, the designer will decide the hardware and software need of the data ware house. It will decide that which type of hardware will be best for the data ware house implementation. Once it is decided then the integration of the hardware and software will be done using integration elements like servers, software tools etc.

3. Modelling and designing: - In this stage, the designer will start designing and modelling the data ware house of the business. In this stage it is decided that how the data will be arranged in the ware house and how will it look like, which data and information will go in which column and row etc. This stage will give visualization of the data stored in the data ware house.

4. Data ware House approach selection: - In this stage the approach used to implement the data ware house is chosen. ETL is the best solution for this stage. First the designer will find from where to extract the data. It is then transformed in to the desired form of data as per requirement analysis of the business and then it is made available in the data ware house. A lot of tools are available in the market to perform ELT.

5. Designing OLAP cubes: - OLAP means multidimensional data cubes. In this stage the designer will obtain the data from related and unrelated sources. That is why it is called multidimensional data. It gives the facility to obtain data from multiple resources and provides the data in the OLAP cube which is structured data according to the business goals and aims. Before uploading the data in the OLAP data base, the data is refined and cleaned and then arranged in the OLAP data base.

6. User Interface Modelling: - In this stage the graphical user interface will be planned and designed. This is to see as how the data will be seen by the user. This is external modelling of the data ware house; it should be planned in such a way that it should be easily used by the end users. The user interface should be such that the data can be easily accessed by the end user.

7. Maintenance: - In this stage the specifications of the data ware house can be modified. If there are any changes to be made in the data so it can be done in this stage. The data ware house should be such that it keeps the record of the modifications in the ware house data.

8. Testing: - In this stage, the data present and arranges in the data ware house is tested. Since the data in the ware house is present from different resources, the tester will test the scope of the queries of the users. Which means the tester will first of all test whether the data ware house is according to the business requirements and meeting all the aims and goals of the organisation. The testers will also test whether the data ware house is responding with the correct results of the queries of the users. If there is issue with some resource then it should be fixed in this stage.

9. Deployment: - The data ware house is finally made available for the users use in this stage. Once the testers are satisfied by the designed data ware house then the ware house will be deployed in live environment which can be used by the user.

All the above-mentioned stages are involved in creating and implementing the data ware house. Each and every detail of each step is mandatory to perform to get efficient working data ware house for the end-users. The user will request a query like study of the latest market trend then the data ware house will respond with all the necessary information it will contain in its data base in the form of tables, columns etc. It will also give the graphical representation of the trends which makes it easy to understand.

Figure 1 Data ware House Process.

Design Methods of Data Ware House: - [4]

Following are the three design methods commonly used to design the data ware house: -

- ➢ Bottom-up Design Approach: - In this approach the data is collected from various resources in once place based on the business goals. It is then integrated in the data ware house for the use of the end users. It is integrated so it is the data from two or more resources.
- ➢ Top-down Design approach: - In this approach, the data is collected in detailed form and is related to the complex data type and large volume of data. It is arranged in data ware house which can be easily accessed by the various department of the business
- ➢ Hybrid Design approach: - This is the combination of first two design methods. It supports multiple dimensional data gathering from multiple resources. It is used to reduce the data redundancy. It is hybrid model which uses the best concept of first two approaches to provide best possible solution to the users.
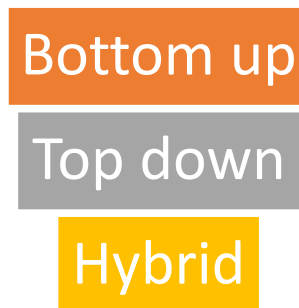

Figure 2 Design methods.

Advantages of Data ware House: - [5]

1. Improved quality of data: - There is large amount of data present in each department of the organisation. It is difficult to perform data analysis. But with the help of the data ware house, this process is eased as the data stored in it is loaded after the cleansing of the data. Hence, the scope of quality of data increases because of the data ware house data.

2. Improved efficiency: - The efficiency with which the user or employee gather information from multiple resources in order to read trends is very low. It takes a lot of efforts and time-consuming task to do so which result in low performance rate of the employees. But as the data is present in structured form based on the type of the data, it is very easy for the employees to access data as per their needs. Hence the efficiency with which the employees perform data analysis and reporting improves a lot using data ware house.

3. Access to Old Data: - The data ware house provides the facility to even access the old and previous data's which is not possible in the operational data base. In some scenarios the employee needs historic data to perform many tasks like training purpose etc. In such cases data ware house is prove to be a boon to the business.

4. Enhanced security: - The data ware house uses several security techniques which improves the security of the unauthorised data access.

5. Higher performance: - The employees of the business have many tasks and responsibilities to perform due to which they will not be able to do proper data analysis and takes a lot of time for the queries and answering them. If the business will have separate dedicated data ware house which

can handle queries fast and efficiently then there will not be unnecessary pressure on the employees and the response to the queries can easily be taken care buy the data ware house which will improve the overall performance of the business.

6. Accuracy rate: - Since the data stored in the data ware house is already refined and cleansed before it is uploaded and made available for the use of the clients, the accuracy rate to respond to a query and the result will improve. Businesses with data ware house gives accurate results to the users.

7. Increases the revenue: - The overall revenue generated by the business will improve with the use of the data ware house. As data ware house has clean data, it will help the employees to perform better data analysis and helps them to make fast and correct decisions which is useful for the business as they can make correct decisions as where to invest and generate higher revenue.

8. Facilitates Cloud: - The modern data ware house can also be implemented cloud based which in turn will improve the performance of the business.

Challenges of Data ware House: -

1. High maintenance cost: - If a small business gets large project for which large volume of data is required, then the cost of the implementation of the data ware house increases.

2. Chances of failure: - If there is some issue in the resource system then the data ware house will not be able to give best possible solution to the query raised by the employees which may end up in data ware house failure.

3. Difficulty in time estimation: - It is very difficult to determine how much time it will be taken to gather data from various resources and then clean and refine it and then made available in proper format.

Conclusion: - It is seen that the data ware house is very effective as it helps in faster data analysis and reporting which helps the business to achieve its goals and aims in less time. It follows certain steps in order to implement the data ware house. There are three types of design approaches which helps in easy implementation of the data ware house.

References: -

1.https://en.wikipedia.org/wiki/Data_warehouse

2.https://www.xplenty.com/blog/the-ultimate-guide-to-data-warehouse-design/

3.https://www.javatpoint.com/data-warehouse-implementation

4.https://en.wikipedia.org/wiki/Data_warehouse#Design_methods

5.https://www.datamation.com/big-data/top-10-benefits-of-a-data-warehouse/