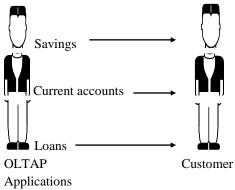# AN INTRODUCTION TO DATAWAREHOUSE AND ITS METHODOLOGY

Kavita Rawat, Jyoti Hazrati
Dronacharya College of Engineering, Gurgaon

Abstract- **According to Kamber, "Data warehouse is a structured repository of subject-oriented, time-variant, historical data used for information retrieval and decision support. The data warehouse stores atomic and summary data". That means it provides a way to represent management information system, executive information system, business intelligence system, analytic application, decision support system which helps an organization to get the information in a way they want. In this manuscript, I have tried to get a brief introduction about why data-warehouse is being so important in today's scenario and what are its methodology.**

I. INTORDUCTION

There are some possibilities that we have heard about datawarehousing- a little fuzzy on exactly how it is being used in large organization. It is being widely used within largest and most complex business in the world. The basic idea of datawarehousing is simple. Data is being extracted periodically from the applications and transfer into the detonated computers where it can be validated, reformatted, reorganized, summarized and supplemented data. The resultant data will be saved and used for further process. Here is an example of how it actually switch between applications and customer.

OLTAP   Customer
Applications

II. WHY YOU NEED A DATAWAREHOUSE?

Earlier, it was very difficult to build datawarehouse without knowing the correct technique. Many early adopters found it to be costly, time consuming, and resource intensive. Over the years, it has earned a reputation for being risky. This is especially true for those who have tried to build data warehouses themselves without the help of real experts. Fortunately, it is usually no longer necessary to custom build your own data warehouse. The heavy lifting has already been done by others. Prebuilt solutions are now available that dramatically reduce the effort and risk. As a result, the time has come for organizations to develop a thorough understanding of data warehousing. The datawarehousing typically:-

• Reside on computers dedicated to this function
• Run on a database management system (DBMS) such as those from Oracle, Microsoft, or IBM.
• Retain data for long periods of time
• Consolidate data obtained from many sources
• Are built around a carefully designed data model that transforms production data from a high speed data entry design to one that supports high speed retrieval.

III. WHAT IS A GOOD DATAWAREHOUSE?

A best datawarehouse inquiries about the data which it holds and the type of reports generated. This is done by developing strong metadata and there is some technique involved in knowing what kinds of metadata will be useful in support of reporting and analysis. The best data warehouses include a rich variety of useful metadata fields. The best data warehouses perform some prejudging of the raw data in anticipation of the types of reports and inquiries that will be requested. But difficulty always occur in any form and anywhere. Difficulty comes in making data warehouse is the design of data model around which it will built. There has to make a decision in naming each field and checking whether each data field need to be formatted and what field should be calculated, manipulated and

added. It should be taken under consideration what data need to be added to the model. Once a Data

Model is made operational, it is important that it should remain stable. Also, if there any changes made should become updated and new data should be rewritten overtime and doesn't require reports to be rewritten.

IV. HOW TO HOLD DATA IN DATAWAREHOUSE?

- *Generating reports*: Creating reports regarding the manipulation with data increases the accuracy and consistency of data and also reduces cost. Such creation directly accomplished by business environment.
- *Packaging of analytical applications*: Analytical applications are in demand in market. It provides reports that organization used to measure their performance.
- *Ad-hoc reporting and analysing*: Since datawarehouse eliminates need for BI tools so, it helps users to generate and analyse it easily. While doing this, it not only helps in retrieving the data which is well accurate.
- *Dynamic presentation*: The growing number of managers want an access to a vision of maintain data. Sophisticated displays that show real time information in creative, highly graphical form are often called dashboards.
- *Capability*: The leading BI systems all allow users to drill down into the details underlying the summaries in reports and dashboards. The presence of a data warehouse makes it practical to use this capability as much as need.
- *Supportive*: Such a well-designed datawarehouse provides and ensures that data is retained in datawarehouse for as long as is required.
- *Creation of metadata*: The creation of metadata helps datawarehouse to build the data and organize it carefully because datawarehouse act as a source between source applications and BI tools. There is a need to prejudge some data. It can include something as simple as an average. Data warehouses can be used to create and store a great deal of metadata of potentially great value.
- *Support for operational processes*: The creation of a sound BI infrastructure is often the best way to meet certain ongoing business needs. The most common example is to facilitate the consolidation of financial results within complex organizations,

especially those whose divisions use different software systems. Meeting regulatory reporting requirements is another common situation.
- *Data mining*: The outstanding software tools that can sift through mountains of data and uncover hidden insights work best on a data warehouse.
- *Security*: Datawarehouse makes it very easier to provide security to those who have legitimate need to particular data and to exclude others.

V. IS DATAWAREHOUSE NECESSARY?

Yes the big question arises when we think of whether it is mandatory or not. Following are the points that describes why it is necessary:-
- Every data is organized in tabular format to calculate data entry and performing validation activities and making them easy for retrieval and analysis.
- There is no other good way to interpret data from other resources into the database of a particular application.
- With the help of datawarehouse, we can easily develop and maintain metadata.
- Transaction processing always provided with the priority. Reporting and analysis functions tend to perform poorly when run on the hardware that handles transactions.
- Using datawarehousing, there is always a risk maintained that users can corrupt or misuse the data.
- There may certain condition occurs when some changes make it very difficult to create and maintain reports that summarize data originating within more than one release.
- Sometimes, it becomes very difficult to encrypt and decrypt the database. We can easily performs the encryption from plain to cipher text but it may create problem while decryption process. And data may get decrypt improperly.

VI. CONCLUSION

In this manuscript, I have prepared the manuscript where we have discussed the concept of datawarehousing in organization. We have obtained various ways in which datawarehousing will be properly and efficiently be deployed in working environment as well as I have cleared the concept of metadata. The case for obtaining a BI solution based on a datawarehouse has become compelling, even for businesses struggling with layoffs and drastic cost cutting. Without one it is very hard to determine how to rebuild a business model around current levels of demand.

REFERNCES

[1]
https://www.google.co.in/#q=topics+on+data+ware
housing+ppt
[2]   http://www.engpaper.com/white-papers-
computer-science-data-warehousing.htm
[3] http://www.wikipedia.org\datawarehouse