A STUDY ON DATA WAREHOUSE ARCHITECTURE

A.R.Arunachalam, S.Srigowthem,
1, 2 Assistant Professor, Department of CSE, BIST, BIHER,
Bharath University, Chennai.

Abstract: Data Warehouse is the large collection of data. It consists of repository of information. This paper discusses about the data warehouse multilevel architecture and its components.

1. Introduction

A data warehouse data is subject oriented, non-volatile, time variant and integrated data[1]. Data Warehouse consists of collection of data for decision support system. Data Warehouse architecture can be represented as one-level, two level and three level architecture[2]. In a data warehouse, at the one level architecture data is stored once. In two level architecture it contain features of one level as well as middleware between source data as well as analysis tools.[3-8] In three level architecture it contains operational and derived data. The first level contains data from different sources of data. At the second level it contains transformed data. In the third level it contains data for decision support.[9-14]

2. Data Warehouse Architecture

The purpose of the Data Warehouse architecture is to provide integrated environment from the database. Data Warehouse doesn’t produce any data but it can store the data from various databases ie from external resources. Data Warehouse consists of 4 layer Database, Data Staging, Data presentation and various data mining tools.[15] Database is nothing but a source for the data which consists operational source systems and from other sources like Oracle, SQL and from SQL. Data staging area is an Data Warehouse environment includes ETL which is an Extraction, Transformation, and Loading (ETL) of data. Data Presentation layer includes OLAP and an Data Warehouse. Data Warehouse consists of tool like query based from the data base and tools for reporting and analysis.[16]

The Data is entered into the Data Warehouse from various operational resources. When the data enter into the Data Warehouse it cleaned and transformed then it is stored. Data transformation occurs due to large volume of data as well as the data is of different data structures. Various tools are available to transform the data, cleaning up the data. Data transformation includes replacing some default value for missing data, conversion of data into some format, removing unnecessary data.

Meta data includes information about the data warehouse. Generally metadata includes dimension details, view details, summarized data and indices. The metadata which contains information about Data Warehouse is called technical meta data. Business meta data contains information for decision making. [17][18]

Figure 1. Data Warehouse: Three Tier Architecture

In the classical architecture, it is difficult to analyze the data because data are stored from heterogeneous databases. Inmon proposes top down approach which adapts to the relational database tools. Kimball proposes bottom up approach. Data is redundant and it is flexible. In the top-down approach the database is centralized. Redundancy is directed to some extent. It is easy to maintain the source data but the data is inflexible in this approach. OLAP On Line Analytical Processing provides various analytical capabilities as well as operations. It supports slicing, dicing, drill-up and drill-down. OLAP is the tool to analyze the data. MOLAP and ROLAP are the various types of OLAP.
3. Components Of Data Warehouse

Data Warehouse supports multidimensional database. Data Warehouse consists of Data Mart which is a small database consists of fact table as well as dimension table. Data Warehouse supports Multidimensional schema such as Star Schema, Snowflake schema and Fact constellation table. In a Star schema , it consists of single centralized fact table and one or more dimensional tables. In a snowflake schema, it consists of one fact table as well as normalized tables. Fact constellation schema consists of more than one fact table. In the multidimensional data model the are viewed in the form of data cube. In multidimensional model the data can be structured in the form of dimensional modelling. Multidimensional data model has the components such as Dimension, Dimension hierarchy, Summary measure and summary function. Summary measure may be distributive or Algebraic.

4. Conclusion

This paper describes the features of three level architecture and the various components of Data Warehouse and its components. Data Warehouse integrates large volume of data as well as it support both structured and unstructured data.

References
