GJCRT: Global Journal of Computer Science Research & Technology
ISSN (online): 2349- 0608

Digital India: Perspective view to Relational Databases

Deepak Kumar

Research Scholar, Department of Computer Science, Punjabi University, Patiala, Punjab, INDIA

dmehta1001@gmail.com

Abstract

Online Analytical Processing and Data warehousing are two of the most significant new technologies in the business data processing area. A data warehouse can be defined as a 'very large' origin of historical data concerning to an organization. The concept of Online Analytical Processing has been developed to meet this challenge. OLAP refers to the technique of performing complex investigation over the information stored in a data warehouse. Data cube is the most important component of OLAP, which is a multidimensional database model which has developed for speed-up of analyzing and processing large amount of data with the usage of various techniques. In this paper, we design a data warehouse that specifically targets storage, querying and analyzing requirements to the multidimensional cubes of organization data within efficient timely manner.

Keywords: - Business Intelligence, OLAP, Data Warehouse, Multidimensional Cubes, Star Schema

----------***----------

1. INTRODUCTION

Data warehouses are extensively used in today's business market for organizing and analyzing large amounts of data. An important technology to exploit data warehouses is the Online Analytical Processing technology [1], which supports flexible interactive analysis of multidimensional data in different granularities. It has been widely applied to many different domains [2]. OLAP on data warehouses is mainly supported through data cubes [3].

Data warehousing and Online Analytical Processing both of them are the most important components of contemporary Decision Support Systems (DSS). They allow organizations to make effective decisions with the passionate of

both of their current and future state. All of the organizations are storing huge amounts of data related to their day to day operations. Today's markets are much more competitive than past, this huge amount of organization data should be processed and made available when required with more accuracy. The success or failure of the organizations dependent to their ability to analyze and synthesize data. So it becomes more vital to process and manage this data. Moreover, these days decision making is dependent on availability of high quality information within time variant.

2. OBJECTIVES OF THIS WORK

This proposed study has been focused upon building and querying multidimensional cubes of organization data.

1. To perform demographic analysis of a dataset and identifying various dimensions, levels of dimension and building a cube based on basic elements required for the structure of cube.
2. Analysis and evaluation of results and query performance.

3. LITERATURE REVIEW

The core of any OLAP system is an OLAP cube also called a 'multidimensional cube'. It consists of numeric facts called measures which are categorized by dimensions. The cube metadata is typically created from a star schema or snowflake schema of tables in a relational database. Dimensional modeling is the most appropriate approach to design a data warehouse [4]. The main objectives of dimensional modeling are: **(i)** to implement database structures that are easy to understand and write queries against, and **(ii)** to maximize queries efficiency [5]. Therefore the solution to building an efficient multidimensional data base is to combine all logical attributes. The combination is especially valuable since typical dimensional are hierarchical in nature. We found this approach is very useful and relevant for our work.

4. DIMENSIONAL MODELING

Dimensional modeling is the different way to view data from different dimensions. This view may be used in a Decision Support System as a part of BI in correlate with data mining duties. DSS applications often require that information obtained along many dimension. For example, if we want to list out the detail of male_employee from marketing_department and belongs to rural_area having above 15% hra allowance in their gross salary. This query requires the three dimensions: *Employee_Info, Salary, Department.* All dimensions are collection of logically related levels and then levels of each dimension again having hierarchies and attributes. Within each dimension, these entities form levels, on which various questions may be asked. The specific data stored is identified as facts and having mostly numeric data. Facts consist of measure data. The measures are the numeric attributes of the facts that are queried against data cube. DSS queries

may access the facts from different dimension, levels and hierarchies.

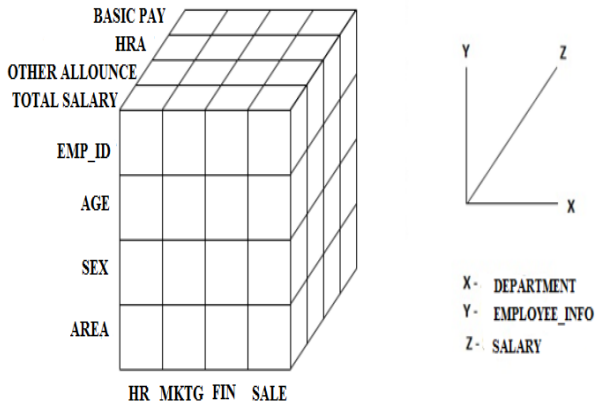

Figure 1: Multidimensional cube with different levels of each dimensions.

The levels in each dimension permit the retrieval of facts from different dimensions. The data warehouse should have summarized collection of data attributes that makes information retrieval more efficient. Facts are also called key performance pointer in data warehousing. The aggregated information of fact viewed at 0-D level, it is known as 0-D cube. This type of cube meaningless at most of time. If it contains information specific to one dimension, then it is known as 1-D cube. If it contains two dimensions, it is known as 2-D cube. Similarly if it contains n-dimensions for visualization of facts, it is known as multidimensional cube. Any combination of different dimensions with various axis while browsing the OLAP cube which helps us base level analysis of fact data for efficient decision making. Cubes are developed for mainly data analysis. Data analysis is a process to evaluate predictive information from large databases to find out relevant facts. The problem of building the multidimensional cubes can be representing in figure 2.

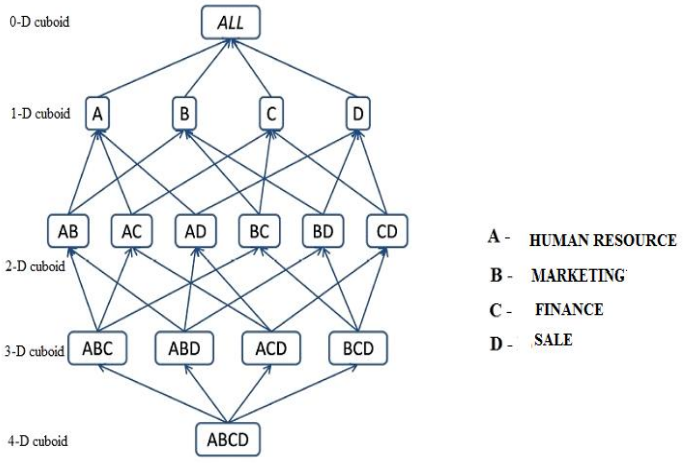

Figure 2: - A Discrete mathemathics application (Lattice Structure) of multi-dimensional cube.

5. MULTIDIMENSIONAL SCHEMA FOR DATA WAREHOUSE

Multidimensional models represent data with a 'cube' structure [6], making more attuned logical data representation with Online Analytical Processing data administration. The data can be directly queried by passing complex database queries by using different combinations of dimensions. Multidimensional models gross advantage of essential relationships in data to populate data in multidimensional matrices called data cubes. in multidimensional cube query performance is much better than relational

data. The response time of the query depends on how many levels are added at each dimension[7]. There are two types of schemas are surveyed in the process of designing multidimensional data cubes. If all the dimensions are directly joined to the fact table then this schema is called *star schema* if the dimensions are not directly joined to the fact table but through other dimension, this schema is called *snowflake schema*. Mostly snowflake schema is used when the records in the fact table is less compared to dimension table.

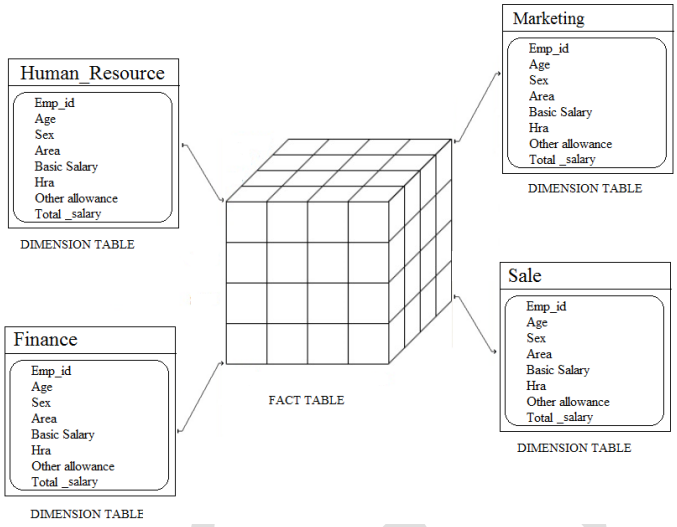

Figure 3: - Star schema of this study.

The various steps for designing a multidimensional model include

- Designing a fact table is to determine the granularity of the fact table.
- Identify the dimensions.
- Define measures which lead to analyzing and reporting for each fact table.
- Aggregating all the attributes, levels and hierarchy for each dimension of fact table.

6. OLAP FOR ORGANIZATION DATA WAREHOUSE

OLAP system is battered than relational database or online transaction processing system to provide more complex query results. . The OLAP system provides analysis of data as well as more vague nature of queries. The complex nature of Online Analytical Processing application requires a multidimensional view of data, and the type of data accessed is often a data warehouse. Online Analytical Processing tools can be classified as relational OLAP or multidimensional OLAP. In MOLAP system, viewed, data are modeled and stored in a multidimensional schema. MOLAP tools are implemented by specialized Data Base Management System and software systems capable of supporting the multidimensional data. With Multidimensional Online Analytical Processing the view of cube data is stored as an n- dimensional array, this approach require high storage space and then indices may be used to speed up processing of data. The parallel computing can be used to overcome this limitation, and then such cube requires less processing time.

In this Organization data warehouse, we persistently use multi-dimensional cube to store the data. Every cube is predominantly its own value because all these make possible to drill-

down and roll-up operations with other cubes. There are several types of Online Analytical Processing cube operations that which support to response more complex queries within timely and efficiency manner.

Slice: This Slice operation performed by selecting one particular dimension from a given cube and provides a sub-cube.

SELECT ALL (Total_Salary) FROM ALL_EMPLOYEE_INFO WHERE Study_Area='HR' having Gorss_Salary>=7500;

Dice: This Dice performed by slice on one dimension and then rotating the cube to select on second dimension.

SELECT Emp_Id FROM ALL_ EMPLOYEE_INFO WHERE Study_Area='Sale' and Salary_Gross<=50000;

Roll-up: This allows asking queries that moves up an aggregation hierarchy. Instead of looking one fact we look at all the facts.

SELECT ALL(Employee_info) FROM MOLAP_Cube WHERE Study_Area=HR,'MKTG','FIN','SALE';

Drill-down: This operation allows user to navigate lower in the aggregation hierarchy. In this user get more specified results.

SELECT Total FROM AGE WHERE Study_Area='SALE';

7. CONCLUSION

In this paper we have applied notions of dimensional modeling. We have identified major organization attributes and place these attributes in the dimensions of the cube. Further we developed a star schema for data cube and identified the primary keys to associate each of the dimensions with fact table of the star schema. The fact table gathers information from each of dimensions and then, the values of decision coefficient is computed and stored in the fact table, correspondent to each cell of the cube. Many data mining tools and techniques may be applied to this multidimensional database, which possibly responds to more complex queries. This paper strongly advocates the use of predictive data mining techniques to retrieve more sophisticated information from organization data warehouse. We will find our work meaningful if an agent based software could be developed for the purpose of data warehousing and mining of organization attributes.

References

[1] S. Aggarwal , A. Gupta and R. Ramakrishnana, ,*On the computation of multidimensional aggregates*, in VLDB, 1996, pp. 506–521.

[2] R. A. Hankins and J. M. Patel, *Efficient aggregation for graph summarization*, in SIGMOD ’08: Proceedings of the 2008 ACM SIGMOD international conference on Management of data, New York, NY, USA, 2008, ACM, pp. 567–580.

[3] L. Chen and R. Ramakrishnana, *Prediction cubes*, in VLDB ’05: Proceedings of the 31st international conference on Very large data bases, VLDB Endowment, 2005, pp. 982–993.

[4] Kimball, R., Reeves, L., Ross, M. and Thornthwaite, W. The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, developing, and Deploying Data Warehouses. John Wiley & Sons, 1998.

[5] M.A.R Kortnik, D.L. Moody, *From entities to stars, snowflakes, clusters; constellations and galaxies: A Methodology for data warehouse Design*, 18th International-conference on conceptual modeling, industrial track proceedings, ER'99.

[6] R.Kimball, *The Data warehouse toolkit*, John Wiley & Sons 1996.

[7] W. H. Inmon. *Building the Data Warehouse*. John Wiley & Sons, Inc., 1992.