EFFICIENT INFORMATION RETRIEVAL USING MULTIDIMENSIONAL OLAP CUBE

Neha[1], Kanwal Garg[2]

[1]Research Scholar, M.Tech. (CSE), Department of Computer Science & Applications, Kurukshetra University Kurukshetra, India.

[2]Assistant Professor, Department of Computer Science & Applications, Kurukshetra University Kurukshetra, India.

---***---

ABSTRACT *:* Data Warehouse is a repository to store huge amount of data, which can be further used for future decision-making process. But the most complicated question raise here is about the accuracy & efficiency of data. Many techniques & methods were proposed by many researchers, so that the knowledgeable & accurate data can be fetched from data warehouse. OLAP is one of the best data analytical techniques developed till now which gives multi-dimensional view of data to end-user which improve the quality of decision-making process. The objective of this paper is to discuss about the retrieval of efficient information by using multidimensional OLAP cube and after that perform a comparative analysis between SQL queries for relational databases and MDX queries for OLAP cube on the basis of query execution time.

Keyword: Data Warehouse, OLAP, OLTP, SSMS, BIDS, MDX, SQL

1. INTRODUCTION TO OLAP

OLAP stands for on-line analytical processing is an analytical processing tool. It mainly used for the analyzing business data together from daily transactions like health care data and sales data. OLAP is a powerful tool to support decision-making. An OLAP system permits the user to easily extract and view the data from different point of view. It also allows users to perform quick and effective analysis on huge amount of data. OLAP systems stored the data in the multidimensional form. OLAP is able to provide the summary data efficiently and enable users to access this summary data faster and easier [1].

OLAP cube (Data cube or Multidimensional cube) is a method of storing data in the multidimensional form; which allows the faster analysis of data. An OLAP cube has capability to operate and analyze the data from the different or multiple angles. The cube comprises numeric facts called measures, which are characterized as dimensions. After combining the facts and dimension get a multidimensional view of data and which is known as OLAP cube. A multidimensional cube or OLAP cube combines the data from the various sources and store these data into a form that consistent for business users. When arrange data into cube it overcomes the limitation of relational database.

2. RELATED WORK

To carry on the present research work the researcher has reviewed various research papers from 1999 to 2012 onwards. The outcome of the research is discussed in the upcoming paragraphs.

C.Sapia et.al (1999)[2] observed the requirements for a proper multidimensional model that is suitable for OLAP applications. In this paper, they choose six models according to these identified requirements and evaluate them using example of 'vehicle repair'. All the models have its specific strength and when they compare all the models then none of the model satisfied all the requirements. But the combination of all these approaches gives a resulting model and which satisfies all requirements [2]. Aparajita Suman (2004) exploring the features of data warehousing, OLAP and their application in library system. The problem of inconsistent and lengthy response time with less flexible systems can be identified and resolved by OLAP [3]. Bora Beran et.al (2008)[3] applied OLAP technology to environmental data catalogs using SQL server 2008 analysis services. And to visualize the query results they used excel and virtual earth [4]. Sellappan Palaniappan et.al (2008)[1] presents a prototype model for clinical decisions support system which combines the power of both OLAP and data mining. In this, they provide integrated architecture of OLAP & Data mining. System can predict future state and can generate useful information for good decision-making. They build OLAP cube for each disease and also diagnosed the disease by using mining functionality. For this they used clinical data of two years [5]. Constanta Zoie Radulescu et.al (2009)[2] presents OLAP cube called CUBETECH. Cube accepts queries on various dimension & hierarchies. In this they used an example of agricultural production. To perform analysis of some commercial features include crops, cropping system, fertilizers consumption & types of farmers of agricultural production OLAP operations are used. This example proves the flexibility of OLAP tool and that is suitable for the complicated analyses of multidimensional data [6]. Joseph M. Firestone (1998), conclude that not every E-R i.e. entity relationship model can represented as set of star schemas but every E-R data warehousing model which are properly constructed can be represented. The data warehousing E-R models which specifying atomic data

---

dependency relationships without fact tables provide the poor query response performance in very large databases. It often prevents execution of multi-stage analysis process. For analysis process in data warehouse the dimension model is best [7]. Srimani P.K. et.al (2011)[1] analyzed the main advantages of multidimensional model like analysis can perform better and faster on data separated based on dimension and facts. It increases performance and reduce storage costs and CPU usage which is elaborated in via a case study named as "Bharti Enterprise" [8]. Vipin Saxena et.al (2012)[1] in this, an OLAP cube was created by using object-oriented approach and with help of unified modeling language (UML). This paper used a real case study of Indian postal services. In this case study they build an OLAP cube to perform various queries and get results within few seconds. And this OLAP cube form by using snowflake schema [9].

After reviewing the above literature few issues are observed which are related to the traditional system are discussed in upcoming section.

3. PERSPECTIVE ISSUES

After reviewing the above literature it has been observed that there are various issues existing with traditional system i.e. On-line transaction processing. The OLTP systems are database used for transaction processing. It supports day-to-day operation & store large volume of data. Building an OLTP system depends on nature of IT staff's knowledge, skills & business process. The database applications become essential tool that helps entire business & without them, present business may not live. They are functionally effective from design as they collect, store & process all business data which is required to successfully perform daily operation. They provide on-line information & produce various reports to monitor and run the business. But there exist various issues with this OLTP system which is shown below-

i. It takes long time to making report.

ii. They do not support fully ad-hoc Query, analyze & summarize of calculated information which are useful for decision-making.

iii. It does not provide explorative views on data.

iv. With the huge data sets & with tricky queries there may be involved many tables.

v. It is not ready or easy to use for data analyze & for summarization.

vi. The existing historical & current data was difficult to understand & hard to use for monitoring business process.

vii. Performance issue will arise when processing analytical queries.

The following issues can overcome by designing an OLAP cube. And, the methodology for designing OLAP cube is discussed in next section.

4. METHODOLOGY FOR DESIGNING OLAP CUBE

The above mentioned issues can be resolve by using an On-line Analytical processing (OLAP) cube also known as data cube or multidimensional cube. These OLAP cube store the data in multidimensional form. The query language used to work with OLAP cube is multidimensional expressions (MDX). An OLAP cube provides access to only the data that are actually required. In this paper researcher shows how to retrieve efficient information by using multidimensional OLAP cube. To implement this firstly we need to create cube for a data warehouse. SQL server analysis services (SSAS) is one of the Microsoft SQL server 2008 component that support OLAP and data mining functionalities. SQL server includes many data management and analysis technologies. These technologies named as Database engine, analysis service, reporting services & integration services. The multidimensional data provides developers to design publish and modify a data cube. A data cube stores the data in multidimensional format. The cube data can come from relational databases, data marts & data warehouse and on the basis of cube dimensions the data is aggregated.
Tools involved in our implementation are:
When working with SSAS the management and development tools are

- Microsoft SQL server management studio (SSMS).
- Microsoft Business intelligence development studio (BIDS).

SQL server management studio (SSMS):

The SQL server management studio is a management tool to manage relational databases, analysis services databases, reporting services objects & integration services packages. Now by connecting SQL server management studio to an analysis services instance the following database management tasks can be performed- processing analysis services objects, browsing analysis services objects, constructing queries, scripting analysis services objects & managing the analysis services database.

Business intelligence development studio (BIDS):

The BIDS is a development environment for
OLAP cubes. BIDS is Microsoft visual studio with analysis projects extension. After the development is done, BIDS publishes analysis services project to an analysis services database. The database processing can be performed both from BIDS and SMS. The components of BIDS that are used

during the development of Analysis services projects are: Analysis services solution explorer, analysis services designer, analysis services menus and analysis services tools/option. Now, to store the cube data the various storage options has been discussed in upcoming section [10].

5. DATA STORAGE IN OLAP CUBE

To store the multidimensional data, various ways have been proposed by various researchers which are named as- Multidimensional OLAP (MOLAP), Relational OLAP (ROLAP) and Hybrid OLAP (HOLAP). MOLAP stores the cube data in multidimensional format and provide fast querying of database. It requires Precomputation of data and storage of the data in the cube. The data transferred from data source into multidimensional database and then data is aggregated. Since the calculation of summary data is already done it allow OLAP queries to be faster. ROLAP stores the data in relational table. If ROLAP used as storage mode then there is no need to transfer data from relational to any other relational systems. Therefore ROLAP is abstraction level over the relational data and that support for multidimensional query. The major argument related to RDBs is that querying the big database with SQL to obtain summary data resulted in complex queries [1]. HOLAP is a popular which stands between ROLAP and MOLAP. HOLAP is stands between MOLAP and ROLAP. Depending on the Designer facilities of OLAP storage, in HOLAP a developer can choose which part of the data store in the relational format and which part store in multidimensional format. This HOLAP allow the developer to utilize the fast query response of MOLAP and scalability of ROLAP at same time. In SSAS some predefined storage settings are developed to help users. But there are manual configuration is also possible to configure the storage settings in SSAS. The storage setting can be configured separately for any measure group or for any dimension within cube. The predefined storage setting in SSAS is MOLAP.

The implementation of all above mentioned points are discussed in upcoming section.

6. IMPLEMENTATIONAL RESULTS

In this paper, a technique have been used which help in efficient retrieval of data and in this direction Microsoft SQL server management studio (SSMS) & Microsoft business intelligence development studio (BIDS) tool is consider to be more appropriate for fetching the more specific information requested by the user. The Sales_ DW database was used for the implementation purpose [11].
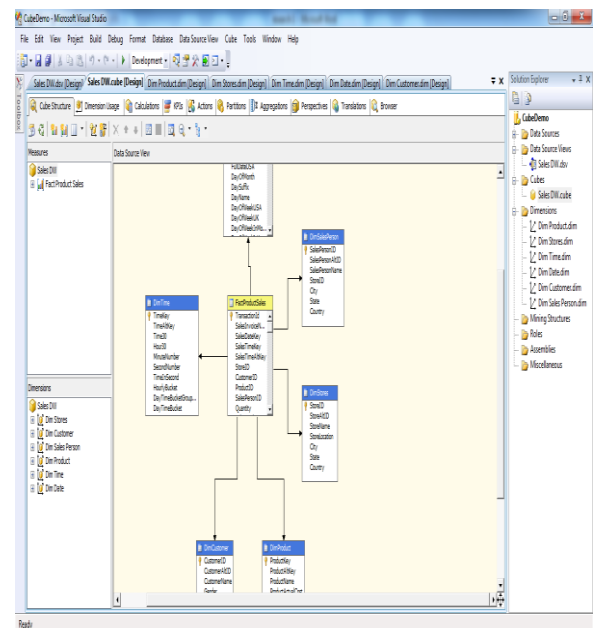

Figure1: Creation of a cube

Figure1 shows that creation of cube in the Microsoft BIDS development environment. There are basically two types of data warehouse schema named as- star schema and snowflake schema. The star schema is the simplest data warehouse schema. Because the design of the star schema is resembles to a star so, it is called star schema. The center of star design contains one or more fact tables and points of star contain dimension tables. The fact tables contain the primary information in data warehouse and the dimension tables contain detailed information about entries of attribute in fact tables. The dimension tables in the star schema are not joined to each other but these tables are joined to fact table by using primary key to foreign key. Other data warehouse schema is snowflake schema. The snowflake schema is more complex than the star schema. Because the design of this schema is resembles to a snowflake it is called snowflake schema. As compared to star schema in snowflake schema the dimension tables data are grouped into multiple tables instead of in one large table. For OLAP systems data is organized in star schemas. In this research work the star schema was used. A cube created by using star schema and in figure1 shows a star schema of Sales_DW database which contains a fact table in the center and many dimension tables that connected to this center fact table by the primary key to foreign key join.
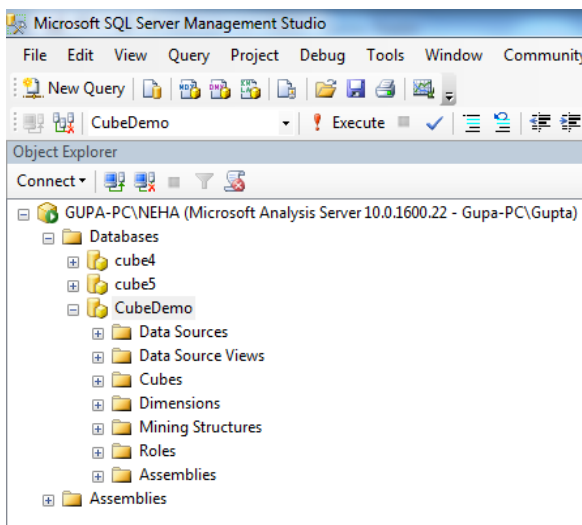
Figure2: Connecting analysis project in Microsoft SQL server management studio (SSMS).

After creating a cube in analysis project deploy this in Microsoft SQL server management studio (SSMS) for writing the MDX (multidimensional expression) queries on cube. Figure2 shows the analysis project connected in the SSMS.
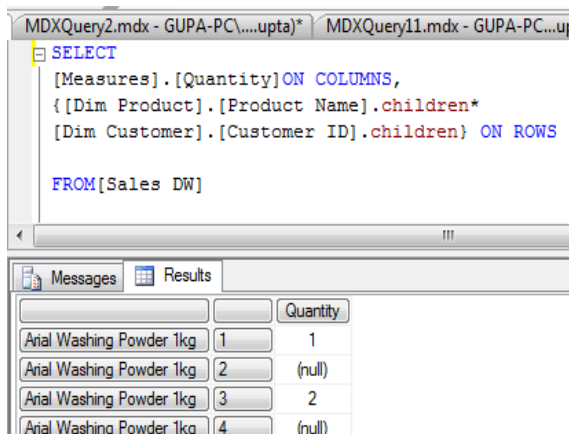
Figure3: MDX query for querying OLAP cube data.

After deploying analysis project in SSMS tool a user can write MDX (multidimensional expression) queries for cube in the query window to retrieve the knowledge from raw data. The MDX queries are different from the SQL (Structured query language) queries. Figure3 show an MDX query fired on a cube and it was observed that query took 26 seconds to retrieve results.
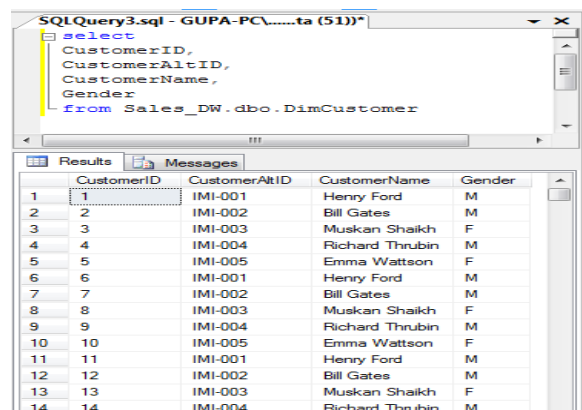
Figure4: SQL query for querying table data.

Figure4 shows an SQL query written for a table to retrieve data from table. After executing query the time 31 seconds has been observed as its execution time.
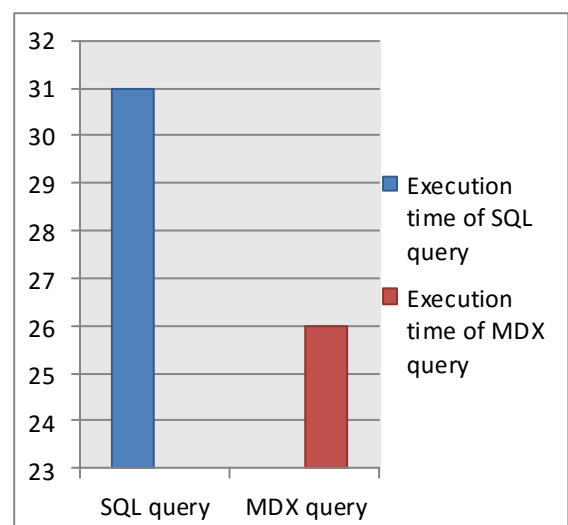
6.1 Comparative analysis

Now, on the basis of query execution time a comparative analysis has been done between the SQL queries and the MDX queries.

Table 1: Comparison table of execution time between SQL query & MDX query.

	Execution time of SQL query (sec)	Execution time of MDX query (sec)
SQL query	31	
MDX query		26

Graphical Representation

Figure5: Graphical representation of execution time between SQL queries & MDX queries.

By doing the comparative analysis between the SQL queries for relational database and MDX queries for OLAP cube it was observed that the time taken by the MDX queries is less than the time taken by SQL queries.

7. CONCLUSION

Among various techniques analyzed in many papers, the OLAP is found to be very effective and efficient technique to be work with. The main objective of this paper is to show that how user can retrieve efficient information using multidimensional OLAP cube and the comparative analysis between SQL queries for OLTP system and MDX queries for OLAP cube. In this paper, Microsoft SQL server management studio has been used to manage database and Microsoft business intelligence development studio has been used to create a cube and from the graph which shown here it can be conclude that the MDX queries perform better than the SQL queries and takes less query execution time than the SQL queries. OLAP can summarize through dimensions to extremely improve query execution time over relational database. Although, an OLAP cube provides efficient information but still there is an issue of time efficiency.

8. REFRENCES

[1] Adrienne H. Slaughter, "OLAP".

[2] C.sepia, M.Blaschka, and G.Hofling, "An overview of multidimensional data models for OLAP," FR-1999-001, February 1999.

[3] Aparajita Suman, "Data warehousing and OLAP technology for knowledge discovery", 2nd international CALIBER-2004, 11-13 February, 2004, INFLIBNET Centre Ahmedabad, pp.542-549.

[4] Bora Beran, Catharine Van Ingen, Ilya Zaslavsky, David Valentine, "OLAP cube visualization of environmental data catalogs", Microsoft technical report MSR-TR-2008-70, July 14, 2008.

[5] Sellappan Palaniappan & Chua Sook Ling, "Clinical decision support using OLAP with Data mining," IJCSNS international journal of computer science and network security, vol.8, No. 9, pp.290-296, September 2008.

[6] Constanta Zoie Radulescu, Marius Radulescu, and Adrian Turek Rahoveanu, "A multidimensional data model and OLAP analysis for agricultural production," 10th WSEAS Int. conference on Mathematics and Computers in Business and Economics 2009, ISSN:1790-5109, pp.243-248.

[7] Joseph M. Firestone, "Dimensional modelling and E-R modelling in the data warehouse", white paper no. eight, June 22, 1998.

[8] Srimani P.K. and Rajasekharaiah K.M, "The advantage of multidimensional Data model- A case study," International journal of current research vol. 3, issue.11, pp. 110-115, October, 2011.

[9] Vipin Saxena and Pratap, "OLAP cube representation for object oriented database," international journal of software engineering & applications (IJSEA), vol.3, No.2, pp.109-117, March 2012.

[10] S. Badiozamany, "Microsoft SQL server OLAP solution-A survey," examensarbete 15 hp, pp. 3-13, 2010.

[11] Web-link: https://www.codeproject.com/Articles/658912/Create-First-OLAP-Cube-in-SQL-Server-Analysis-Serv