The performance evaluation of relational and object oriented database in data warehouse and enhancing its performance using query optimization

Kavisha Duggal[1]
Computer Science and Engineering
Lovely Professional University
Pharwara, India
kavisha_duggal@yahoo.com

Mr. Kewal Krishan[2]
Computer Science and Engineering
Lovely Professional University
Pharwara, India
kewal.krishan@lpu.co.in

Abstract:. **The last few years have been a growing recognition of information as a key business tool. Those who perfectly gather, examine, recognize and act upon the information are real winners in this information era. My area of research will concentrate on the performance evaluation of data warehouse using relational and object oriented database and enhancing its performance using query optimization. Data warehouse acts as a huge repository for storing large amount of valuable data.**

Keywords: Extract Transform Load (ETL); Object Oriented Data Base Management System (ODBMS); Online Analytical Processing (OLAP); Object Relational Data Base Management System (ORDBMS).

I. INTRODUCTION

The data warehouse is an environment which affords an architectural construct of IS (Information Systems) that provides user with the current and historical decision support information that is hard to access or present in traditional operational data store.

A number of new upcoming technologies have emerged that are focused on improving the information content of the data to empower the knowledge workers of today and tomorrow. Among these technologies are data warehousing, Metadata repositories, online analytical processing (olap) and data mining. In some ways these technologies are the appearance of the maturity of the client/server computing model and its applicability to a wide variety of business problem [1]. Metadata is defined as a data about data [2]. The basic function of Meta data are to build, maintain and mange the data ware house. One of the major concerns of tools in data ware housing is OLAP tool which is defined as online analytical processing. This tool is basically based on the concept of

multidimensional data bases and helps to analyze the data by elaborating it having different views.

As we categorize data ware house as subject oriented, it is due to the presence of data marts.

Presence of various data marts leads to implementation of data ware house. Data mart is defined as de normalized and summarized form of data. Data marts can be dependent or independent. We normally prefer to make dependent data marts to have an efficient query redirection and also dangerous as they include several issues with multiple operational systems which leads to multiple non-integrated data marts which are normally overlapping in content of data, scheduling of job, its connectivity and managing the whole source in the proper manner.

Demand of Data ware house according to the business prospective includes accurate and fast decision should be made according to the available data, the whole product will be used by business domain experts, not the professionals of the computer, increase in data, after every 6 months therefore the availability of the data doubles and also there is an increase in the demand of the data ware house as organization distributes control away from the middle management layer to the screened business information[7].

II. PROBLEM DESCRIPTION

The performance evaluation of data warehouse using relational and object oriented database and enhancing its performance using query optimization. As we all are working on relational databases we always give less consideration to object-oriented database. We generally ignore the utility of time while working with small database but when we have to handle large data in terms of data ware house proper consideration has to be given to the speed with which query is answered and its accuracy.

We are always fascinated by easier functionality but when we have to handle large databases several issues can be faced so my research will provide a deep insight which could be the better approach to be used in data warehouse which is far more the extend of databases. Our research will explore the real difference between the concepts of ORDBMS and OODBMS in data warehouse. Usage of real data set will be there in implementation so that a proper defined result could be generated.

III.   SOLUTION DESCRIPTION

Normally it is very difficult to develop a database/data warehouse which is optimized for particular set of queries. As the database moves from infancy to maturity there are many check points where it needs optimization for particular set of queries for specific objectives.

In this research work there is scope for performing analysis, effective comparison and enhancing the performance with query optimization. The concept of analysis will result in better, deeper insights of the query optimization techniques.

- Develop database design for analysis of educational parameters of educational institution (lpu).
- Integrate Meta information and other forms of data in the database design.
- Integrate the concept of knowledge base, structure, density with respect to timeline.
- Identify the parameters for queering the database.
- Develop a workout plan using Microsoft analytical services for each query being fired in an ecosystem.
- Identify highly significant database metrics for query optimization.
- Apply a suitable and most relevant optimization technique that enhances the performance of query.
- Derive conclusion, inferences based statistical analysis of the data set developed in previous step.

IV.   PROPOSED FRAMEWORK

Firstly ETL will be performed with the help of business intelligence tool (PENTAHO) so that a clean data should be provided to Microsoft visual studio (team version). Basically the first step here is to extract data from various sources may be operational systems, flat files etc.

Then next step to be followed will be transformation, when data is available on one platform then following operations like filtering and sorting of data can take place .[1] Then the final step is loading the data into the data marts. Data warehouse will be constructed with the help of Microsoft visual studio with its respective available languages [2]. In this the object oriented approach will be used as we have classes formed in VB. Net or ASP. Net. Then, we have two defined parameters

For ORDBMS there will be usage of SQL server2008 for backend. As here the data will be stored in the form of tables and various relations will be formed.

For OODBMS DB4o with Microsoft visual studio will be used. Db4o is a purely object oriented. It is the world's leading open source object database for Java and .NET. It is the lightweight embedded data-base for scenarios that range from industrial automation to mobile devices . In this database objects retain their properties throughout the database[5]. The contents, nature and as well as the relationship of objects are preserved even if class is too complex.
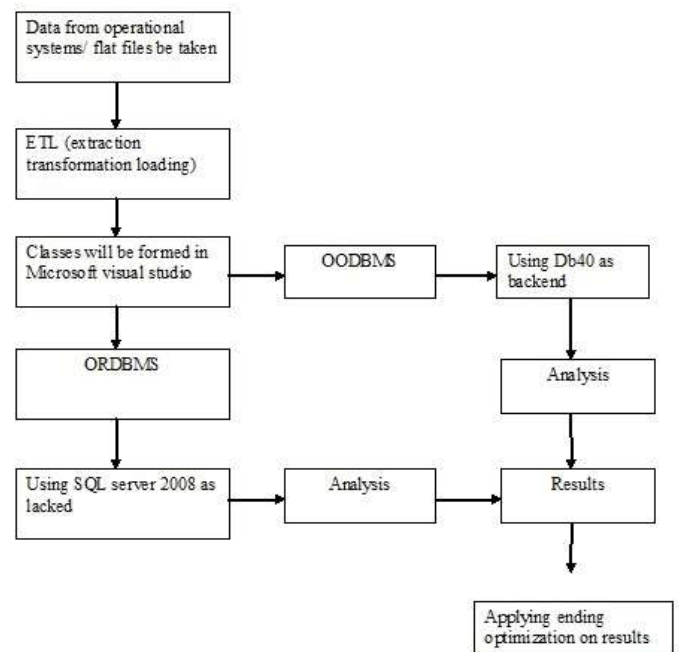

Figure 1 design

Analysis result will be obtained from both OODBMS and ORDBMS. OLAP tools act as a great advantage for the analysis as cube like structure is formed which provides us a three dimensional view. Here the three major components for the cube can be 1) country 2) marks percentage 3) name. Then with the following results I will accordingly apply the query optimization technology to enhance the performance of queries. Performance of queries can be judged on the basis of time to be taken to execute the query on the specified data set.

For query optimization we shall be using some data base metrics to identify how good it is performing. The basic parameter that will be taken into the consideration will be the speed for query optimization. How the optimization enhances the speed factor will be the major

concern. Even the increase of speed in milliseconds will lead to the success full research.

We shall be using SQL server analyzer and profiler to identify optimization parameter for query.

I shall be using work out plans and analysis to also identify optimization for each query.

As in my research work, I will be evaluating the performance of ORDBMS and OODBMS on several parameters and enhancing the performance using query optimization.

Analysis parameters-: I will be taking data set of students of various colleges mainly of lovely professional university the data will be yearly and it will be based on calculating the average performance of students which are coming from abroad.

V. RESUTS

In OODBMS Integrating the database db40 (which is completely object oriented) with visual studio. In this we have tried to insert data from excel sheet into the db40 database and fetch the data on console application from db40. A yap file is created which actually deals with OODBMS. The data stored in the database is highly secured and stored in the form of encrypted data. Here we don't have any tables various objects are formed and data is saved. After saving the data in the db40 data base the query will be applied to fetch the whole data. Proper time analysis will be done to fetch the same amount of data that was stored.
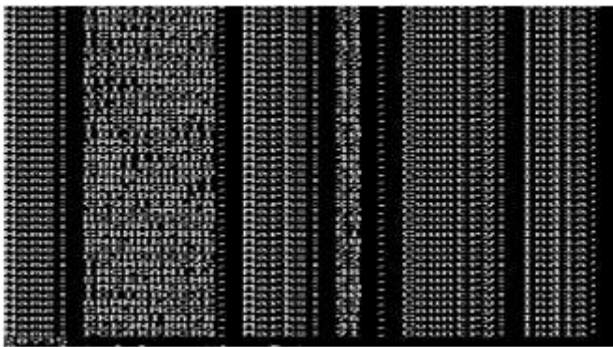

Figure 2 For fetching data in db40

Same query was run 10 times and the range was obtained which is shown by the graph.
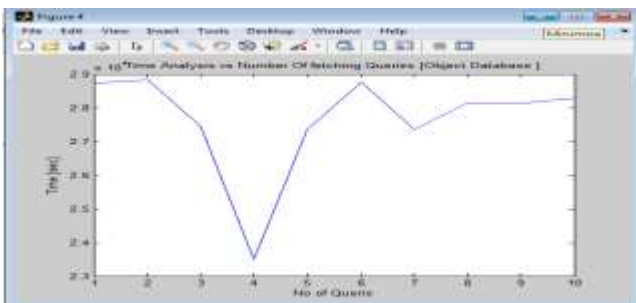

Figure 3 Graph for fetching

In ORDBMS, integration of the database SQL server2008 with visual studio takes place. In this data is inserted from excel sheet into the database and fetch the data by running the application on the server. Here the data is fetched from the tables. The unique id is provided to each entry which acts as a primary key.

Loading all data :48220ms

Studentid	Name	Country	Marks
422811	Awaludin'	Indonesia'	51
422812	Sapruddin'	Indonesia'	56
422813	Muda'	Indonesia'	65
422814	Muhammad'	Indonesia'	56
422815	Dabir'	Bangladesh'	70
422816	Ahmed'	Bangladesh'	40

Figure 4 Fetching data on server

Same query was run 10 times and the range was obtained which is shown by the graph.
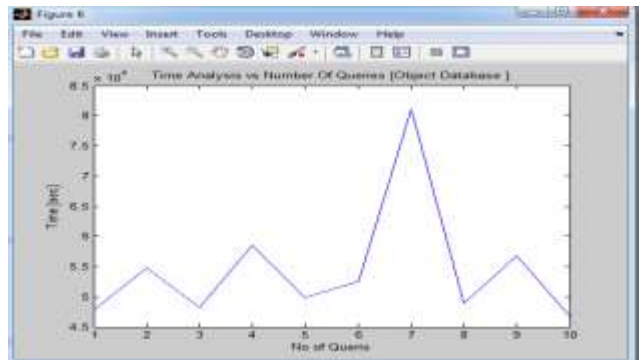

Figure 5 Graph for fetching data on server

The next query which is applied is for updating. In this the time taken to update the particular data is calculated. For updating the data two things are required. Firstly define the value that we want to replace. Secondly define the value with which the replacement will take place. Graph for updating a data here we can have a perfect visualization of minimum and maximum value in OODBMS. The range varies from 2434 to maximum time taken 2527.
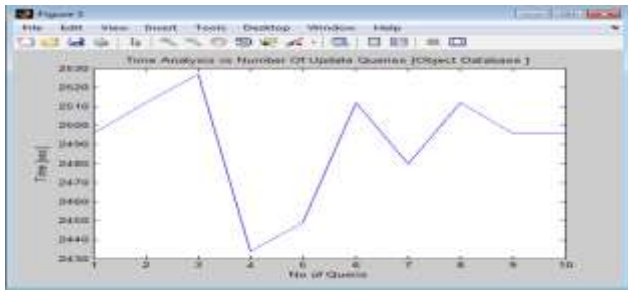
346

Figure 6 Graph for updating

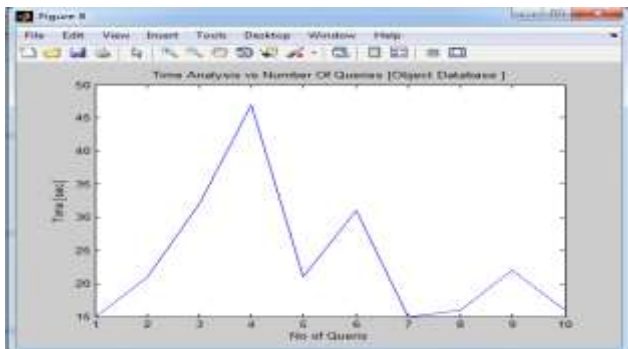Graph generated for ORDBMS for the update query.

Figure 7 Graph for updating in ORDBMS

The query is fired for deletion and analysis is done by the time taken for deleting the particular entry in data base.
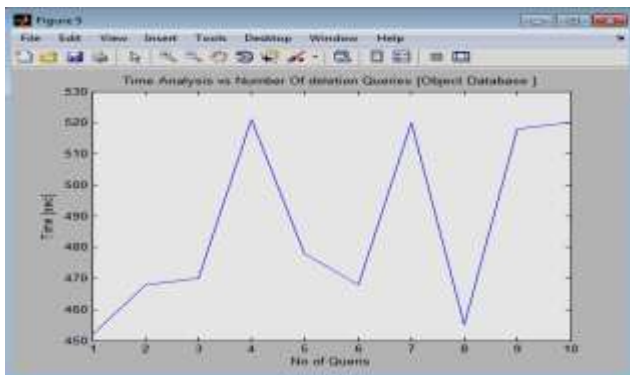

Figure 8 Graph for deleting in OODBMS

The range defined for deleting the value in ORDBMS varies from minimum value 16 to maximum value 281.

For optimization the linq is used. Basically linq is known as .Net language integrated query. It directs the query in an integrated feature of the programming language of the developer. Its main advantages are it permits a general purpose single declarative query to be facilitated and applied to all in memory information, it provides compile time syntax checking embedded with static typing and intelligence, one of the main advantage

is it allows query expressions to benefit from rich metadata, its integration acts as one of strong typing over relation data, it actually maintain the full power of expression and without losing their actual meaning and it acts a consistent model for working with the various data sources and formats and there is

VI. CONCLUSION

After conducting a thorough research which is based on analytical frame work, we conclude that the object-oriented data base queries perform very well when the data set size is small, the performance starts deteriorating when the data is large

While evaluating the performance of OODBMS and ORDBMS, the performance of loading the whole data from database is shown much better in OODBMS. The time range varies from minimum value 23498 to maximum value 23498.Whereas fetching the 37000 values in ORDBMS using Microsoft visual studio platform. Time range varies from minimum value 46732 to maximum value 81136.

While the other queries like updating and deletion were performed the time taken by ORDBMS was less but there was time lag which delays the performance in ORDBMS

REFERENCES

[1] Alex Berson, Data Warehousing, Data Mining and OLAP, Tata Mcgraw hill, 2010.

[2] Joy Mundy, Microsoft Data Warehouse Toolkit, Wiley Second Edition, 2011.

[3] Aernizar Abu Ali et al. "Data Warehouse Critical Success Factors", European Journal of Scientific Research. ISSN 1450-216X Vol.42, 2010.

[4] Dr. Kawaljeet Singh et.al, " A Descriptive Classification of Causes of Data qality Problems in Data Warehousing", IJCSI International Journal of Computer Science Issues, Vol. 7,May 2, 2010.

[5] Mrs. Pushpa Suri, Mrs. Meenakshi, " Comparative Study Between the Relational And Object Oriented Database in Data Warehouse" International Journal of Database, 2011.

[6] Esther Borowski, "Workflow to improve data quality in oracle warehouse builder", Journal of applied quantitative methods, December 5, 2008.

[7] Fatihah Ramli, Rosita Mohamed, " Developing Conceptual View of Schema Mapping on Date Format for Healthcare Data Warehousing", World Academy of Science and technology, February 8, 2009.

[8] G. Satyanarayana Reddy et. al., "Data warehousing, data mining, olap and oltp technologies are essential to support decision-making process in industries", (IJCSE) International Journal on Computer Science and Engineering, September 02, (2010.

[9] Jeff lawyer, Shamsul Chowdhury, "Best practices in data warehousing to support business initiatives and needs" 37th hawai internal conference, 2004.

[10] MansafAlam, Krishan Wasan, " Migration from relational to object oriented database", Journal of Computer Science, october10, 2006.

[11] Richard J, Robert H, "Flexibility of your data warehouse" communication of the ACM, October, 2007.

[12] Yang Bao Lu Jing Zhang, "Decision Support System Based on Data Warehouse", World Academy of Science and technology, November 30, 2010.