Nitin S.Kharat / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 3, Issue 3, May-Jun 2013, pp.119-122

Deciding the Correct Usage of Database Queries, Data Mining and OLAP in an Applications

Nitin S.Kharat.
B.E. (I.T.) M.E.(Computer Engg.)
Department of Computer Engineering, University of Pune, Pune,Maharashtra,India.

ABSTRACT

In the recent years, numbers of the studies have been done on different techniques of information retrieval. These retrieval techniques includes Database Queries, Data Mining and Online Analytical Processing (OLAP).The retrieved information is used for various purposes according to the different requirements. The retrieved information might be used for the purpose of Analysis, for the purpose of various users behavior prediction or for the purpose of Decision Support System (DSS).Now the confusion is that when to use Database Queries, when to use Data Mining and when to use Online Analytical Processing (OLAP).This paper elaborates the usage of Database queries, Data Mining, OLAP according to the user's purpose, requirements at the particular instant. The paper also emphasized on performance of these techniques with appropriate examples. The goal of paper is to give the better clue to the user about the usage of techniques such as Database Queries, Data Mining and OLAP in an application to get the information in an easy way with efficient performance.

Key Terms: Database Queries, Data Mining, Decision Support System (DSS), Online Analytical Processing (OLAP)

1. INTRODUCTION

In today's world because of the lack of time number of the people avoids to go through the large volume of database. As a result the numbers of various information retrieval techniques have been developed. These information retrieval techniques allows to retrieve the required ,interested information from the large volume of database within compact time and in an easy way that's why the number of people chooses these technique as a source of information retrieval. Such a techniques includes the Database Queries, Data Mining and Online Analytical Processing (OLAP) .As we all known these techniques plays an vital role in business including Sales,Marketing,Finanace etc. for the purpose of Sales forecasting, Marketing Research and Finance Analysis. But the crucial skill is that utilizing appropriate technique
at the appropriate time according to purpose, requirement, and performance is an important scenario.

This paper elaborates the usage of these techniques according to the requirement, purpose, and performance with appropriate examples in an efficient manner.

The remaining of this paper is organized as follows. Paper briefly review the related work in section 2, in section 3 described the usage of each technique according to purpose, requirement and performance, the future work and conclusion in section 4.

2. RELATED WORK

In the previous research several studies [1, 2, 3, 4] focused on the possibilities and combining of OLAP and Data Mining.
And also H.Zua [5] proposed and developed association rule mining with interesting pattern, this approach is called as Online Analytical Mining of Association Rule.
In paper [6] developed an approach that operates on two factors support and confidence of users demanded items. This technique identifies item sets which are frequently accessed through the web pages on web site by particular user. The output contains list of customer associated with that product.Dzeroski et al.[7] combines OLAP and output which proved that CUBE File outperforms as compared to the hierarchical clustering.
The paper [8] also proved that the efficiency and capability of integrating data mining into OLAP and OLAM systems .OLAM is an way for mining the required data from large volume of multidimensional database.
The paper [9] also gives an efficient way to convert the XML schema to ROLAP which helps to speed up the execution time which results in improved performance of OLAP as compared to previous approaches.
In the paper [10] targeted to extract the previously unidentified information from large volume of database and supported the speed up process by automatic OLAP schema generation which ultimately reduces the manual work and speed up the performance.

3. USAGE OF DIFFERENT TECHNIQUES

3.1 Database queries:

Basically Database Queries Composes of:

.Database table which is representation of mathematical relation that is set of items that share the certain attributes.

.Each table column represents an attribute of relation

.In relational databases the tables are usually named after kind of they represent.

.Database queries allows the user to efficiently and effectively to manipulate a database.

3.1.1 When to Use:

.When the user needs to retrieve the data from database table such a that the retrieved data must satisfies certain criteria defined by user.

.Suitable to use for the application which gives query response as a small volume of data.

.When the User wants to retrieve the set of data items strictly fulfilling defined constraint.

.When the user needs a specific information or data.

3.1.2 Example:

Consider you have a following database:

Id	Name	Gender	Age	City	Income	Vehicle	Total
4	Nitin	M	23	Pune	3 LK	Skoda	2 LK
.....	.....	.....	.....	.....	.....	.....	.....
24	Mahesh	M	21	Mumbai	8 LK	BMW	4 LK

Table 1: Customer Database

.If the user wants to retrieve the names whose income is greater than 3 LK and city is Pune then he should fire

Database Queries as follows:

. Select Name from customer where Income>3 LK and City=”Pune”.

3.1.3 Performance:

.Gives better performance when there are several numbers of the fulfilling criteria.

.Gives better performance when the retrieved sets of items are small in volume.

.It gives better performance against the specific retrieval of information as per user need.

.Gives better performance when user fires ad hoc queries to virtually any instance in a database.

3.2 Data Mining

Basically Data Mining Composes of:

.Different models for the analysis purpose

.Models can be viewed as high level summarizing of underlying data

.Includes the different forms such as:

-Decision Tree

-Rule for classification task

-Association Rule e.g. Market Basket Analysis

-Clustering for Market Segmentation etc.

3.2.1 When to Use:

.It can be used when user wants to generate high-level, actionable summaries of data residing in database tables.

.When the user wants to builds the Association Rules.

.When the database is too large for analysis.

.When user need to analyze the database in visual form.

.When the User wants to analyze dataset using more simple form i.e. Models such as:

-Decision Tree

-Graphs

-Histogram

-Pie Chart etc.

3.2.2 Example:

Consider the same database:

Id	Name	Gender	Age	City	Income	Vehicle	Total
4	Nitin	M	23	Pune	3 LK	Skoda	2 LK
.....	.....	.....	.....	.....	.....	.....	.....
24	Mahesh	M	21	Mumbai	8 LK	BMW	4 LK

Table 2: Customer Database

.Now here you can use Data Mining technique for building the Association Rules for the Decision Support System (DSS).

.The Association Rule for Income can be built as follows:

If Age>21 and City=”Mumbai” then Income> 3 LK

OR

If Vehicle=”BMW” and Total>3 LK then Income>3 LK

.These Association Rules are very simple to understand for any user because it summarizes number of records from Customer database but if we use Database Queries then it will become a very complicated to the user to understand the above scenario.

3.2.3 Performance:

.Gives the better performance than the Database Queries over the large number of database for analysis.

.Easy to understand as there are no more complex queries

. Association Rules gives analysis at a once glance.

.No need to analyze each and every query.

.Visual representation increases more effectiveness for analysis.

3.3 OLAP

Basically OLAP (Online Analytical Processing) Composes of:

.Summarizing the data before it is possible to execute the queries.
.Summarization can be represented as cubes and subcubes .
.Cube is precalculated and preaggregated data
.It allows reporting data, visualizing data and interaction with views of data.
.It uses the star schema, snowflake schema which consists of fact table and dimension table.

3.3.1 When to Use:

.Used against the normalized set of database tables to get the set of items that fulfills the constraints on its attribute value.
.When the user needs to be data collected from different tables i.e. Complex Joins.
.To speeds up the execution time because it uses ROLAP.
.When user needs summarized data from detailed data i.e. roll up.
.When the needs detailed data from summarized data i.e. Drill Down.
.When the user wants methodology for organization of databases along the dimensions of a business making the database more comprehensible.
.When the user needs to compare and contrast measures along the business dimension in real time.

3.3.2 Example:

Consider the same database:
OLAP constructs the Star Schema(see Fig 1) for above database to speed up the execution time as follows:
.The dimension tables give rise to the dimensions in the pre-aggregated data cubes..The fact table relates the dimensions to each other and specifies the measures which are to be aggregated.
.Here the measures are "dollar_total", "sales_tax", and shipping_charge".

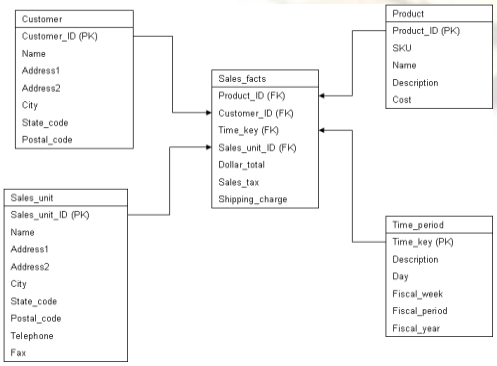

Fig 1: Star Schema of the Database

.Figure 2 shows a three-dimensional data cube pre-aggregated from the star schema
.Data cubes can be seen as a compact representation of pre-computed query results.
.Essentially, each segment in a data cube represents a pre- computed query result to a particular query within a given star schema.
.The efficiency of cube querying allows the user to interactively move from one segment in the cube to another enabling the inspection of query results in real time.
.Cube querying also allows the user to group and ungroup segments, as well as project segments onto given dimensions.
.This corresponds to such OLAP operations as roll-ups, drill-downs, and slice-and-dice, respectively.

.Cube is represented as follows:

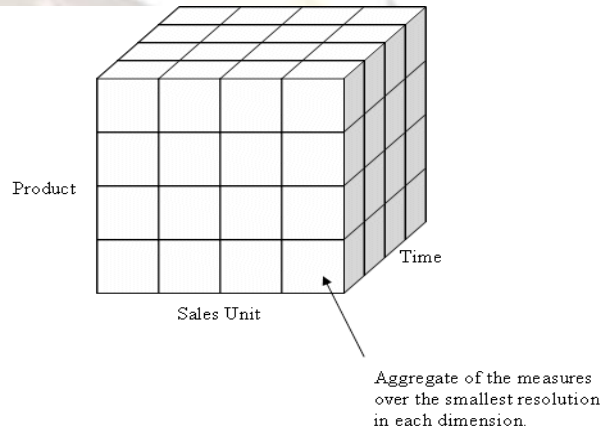

Fig 2: 3-Dimensional Cube

3.3.3 Performance:

.Gives the better performance than Database queries and Data mining when the data need to be collected from different table's i.e. over Complex Joins.
.Cube gives the better performance since segments inside the cubes are precalculated and preaggregated.

4. CONCLUSION

This paper covers the correct usage of information retrieval techniques such as Database Queries, Data Mining and OLAP according to users' purpose, requirement and performance along with examples. The paper also concluded that in Database Queries ad hoc queries are fired by user to access the specific information whereas in Data Mining generally modeling is used to represent the huge retrieved data while in OLAP, it allows the user for real time gain to pre-aggregated measures along with dimension for more effective analysis.

The future work can be enhanced in such a way that one can integrate these different techniques according to requirement, purpose and performance which is absolutely right for the application. If he is able to do so then it will lead to huge benefits to their application as well as organization.

REFERENCES

[1]. S. Goil and A. Choudhary, "High performance OLAP and data mining on parallel computers," *Data Mining and Knowledge Discovery*, vol. 1, no. 4, pp. 391-417, Dec. 1997.

[2]. S. Asghar, D. Alahakoon and A. Hsu, "Enhancing OLAP functionality using selforganizing neural networks," *Neural, Parallel and Scientific Computations*, vol. 12, no. 1, pp. 1-20, March 2004.

[3]. R. B. Messaoud, O. Boussaid and S. Rabaseda, "A new OLAP aggregation based on the AHC technique," in *Proc. of the 7th ACM Int'l Workshop on Data Warehousing and OLAP* (DOLAP), ACM New York, 2004, pp. 65-72.

[4]. J. Han, "Towards online analytical mining in large databases," *ACM SIGMOD Record*, vol. 27, no. 1, pp. 97-107, March 1998

[5]. H. Zhu, "Online analytical mining of association rules," Master Thesis, Simon Fraser University, 1998, pp. 1-117

[6]. J. Fong, H. K. Wong and A. Fong, "Online analytical mining Webpages tick sequences," *J. of Data Warehousing*, vol. 5, no. 4, pp. 59-67, 2000

[7]. S. Dzeroski, D. Hristovski and B. Peterlin, "Using data mining and OLAP to discover patterns in a database of patients with Y chromosome deletions," in *Proc. AMIA Symp*., 2000, pp. 215–219.

[8]. J. Seo, et. al., "Interactive color mosaic and dendrogram displays for signal/noise optimization in microarray data analysis," in *Proc. of the Intl. Conf. on Multimedia and Expo-Volume 3*, 2003, pp. 461-464.

[9] Sarbani Dasgupta, Soumya Sen, Nabendu Chaki; "A Framework To Convert XML Schema to ROLAP"; *Proc. of 2nd Intl. Conf. on Emerging Applications of Information Technology, 2011.*

[10] Muhammad Usman, Sohail Asghar, Simon Fong"Data Mining and Automatic OLAP Schema Generation".

[11] Soumya Sen, Ranak Ghosh, Debanjali Paul, NabenduChaki; "Integrating Related XML Data into MultipleData Warehouse Schemas"; *Proc. of the First International Conference on Information Technology Convergence and Services (ITCS 2012), Bangalore, India.*