Data Cube Representation for Vehicle Insurance Policy System

Narander Kumar
Department of Computer Science
B. B. Ambedkar University
Lucknow (U.P.), 226025,INDIA

Vishal Verma
Department of Computer Science,
B. B. Ambedkar University
Lucknow (U.P.), 226025,INDIA

Vipin Saxena
Department of Computer Science,
B. B. Ambedkar University
Lucknow (U.P.), 226025,INDIA

ABSTRACT

On-Line Analytical Processing (OLAP) systems have a strong focus on the interactive analysis of data and typically provide extensive capabilities for visualizing the data and generating summary statistics. Most of the data sets can be represented as a table, where each row is an object and each column is an attribute. Data cube represents the multidimensional data with all possible aggregates. The three dimensional data cubes represent the different attributes entirely controlled with the help of objects. In general, a data cube is generalization of statistical terminology as a cross-tabulation. In the present work, authors have designed a framework of OLAP data cube to analyze the Vehicle Insurance Policy (VIP) system to identify the entity, which is highly preferred by the customer. The study describes a methodology with OLAP data cube and pivot table as well as a correlation technique which represents strong relationship among the data attributes. Tables and graphs are designed for the sample database of the Vehicle Insurance Policy System

Keywords

OLAP, OOMD, Data Cube, Pivot Table, Correlation Coefficient.

1. INTRODUCTION

In the current scenario, storage of large amount of data is a big challenge due to rapid increase in the size of files related to the audio and video files. Small software applications include these files for the interactive representation of applications, therefore, data ware houses is a technique for storage of large amount of databases. Since, structured technology based softwares are shifting towards the object-oriented softwares which include the object-oriented databases also. Recent softwares support the object-orientation and from the literature it is observed that very little amount of research is available on the object-oriented databases [1]. On the other hand, data cubes support the object-orientation and one can say that the data ware houses and data cubes designing for OLAP are essential components of the decision support system [2]. Let us first describe some of the important references related to the data ware houses. Recent advancements and various kinds of research problem are well described by [3]. The techniques for maintaining data cubes are represented by [4]. Codd's rules are famous for the structured database designs but in the year 1993, authors [5] formulated the concept of OLAP which described an enabling architecture for OLAP and applied on the counter Galley available product to evaluate the performance as per rules of OLAP. In [5] presented a formal model of dimension updates in a multidimensional view through an efficient algorithm. Since multidimensional databases support logical approach, Therefore authors [6] he discussed a logical model for OLAP system and presented a design methodology to obtain an multidimensional databases scheme from an operational database, implemented by describing translations into relational tables and multidimensional arrays. Others related references on multidimensional databases are [7-9]. From the literature .it is observed that object-orientation is also proposed for the OLAP by the various authors from time to time. This is because of increasing interacted in multidimensional databases and on-line analytical processing [10]. it has described object-oriented multidimensional model concerning of classes fact classes and presented the cube classes as basic structure to allow the analysis of the different data's stored in the system. Unified modelling language for the presented of object oriented database is also used by [11] and authors introduce the different discussion and data cubes with their operators. .A framework for the object-oriented on- analytical processing is given by [12]. The books on the above aspects are [13-14]

In the present work, three dimensional databases for Insurance Policy System are considered and the selected database is presented by the three dimensional data cube representation. Several queries for the OLAP cube are performed and a correlation technique of data mining is described on the database for checking the strong relationship among the data. A pivot table is also designed for the OLAP data cube

2. METHODOLOGY

2.1 OLAP Data Cube and Pivot Table

OLAP cube is a designed database that is optimized for the reporting purpose. While most databases for On-Line Transaction Processing (OLTP) such as those use claims processing designed for efficiency in data storage. The design of OLAP cubes is used to retrieve data and check the efficiency. The meaning is the data is stored in such a way as to make it easy and efficient for reporting. OLAP cubes categorize into "dimension" and "measures". Measures represent item that are counted, aggregated, such as costs of units service. Before OLAP technology was well developed, data is to be extracted from databases using "queries"

A pivot table is a Microsoft excel feature that allows access to data is organised into Dimensions and measures. It acts as an OLAP client to allow user to interact with data stored in OLAP cubes using the familiar Excel spreadsheet format.

In this table describe the data of Vehicle Insurance Policy System; POLICY-ID and DATE- OF-PURCHASE are organized into Dimensions and Measured. Each POLICY-ID is created in particular DATE-OF-PURCHASE is denoted by 1(one).Grant total is equal both sides. This is described in the pivot table.

Table 1: A Pivot table of OLAP Data Cube

POLICY_ID	DATE_OF_PURCHASED		Grand Total
33468522	1		1
33468523	1		1
33468548	1		1
33471849	1		1
33471851	1		1
33471852	1		1
33471853	1		1
33529624		1	1
33529625		1	1
33529627		1	1
33529628		1	1
Grand Total	7	4	11

2.2 Correlation Technique

To characterize a correlation between two random variables x and y, which is the data set

$$X = \{x1,\ldots\ldots xn\}$$

and

$$Y = \{y1\ldots\ldots\ldots..yn\}$$

Usually, the linear correlation coefficient r is defined as

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\ \sqrt{\sum_i (y_i - \bar{y})^2}} \quad \ldots\ldots\ldots\ldots(1)$$

Where $\bar{x}$ and $\bar{y}$ are the sample means for values of x and y respectively. When the correlation coefficient is known to be significant, the coefficient r is one in conventional way to summarize its strength, However there is no standard way to compute the r distribution in the case of the null hypothesis, where the variables x and y are not correlated. In other words r is a weak statistic to decide whether a correlation is statistically significant or whether one observed correlation is significantly stronger than another if the data is not Gaussian.

To solve this problem, often a rank statistic is used. It has a precisely known pdf function (probability distribution function), that allows calculation of the significance of observed correlation. In this paper the rank correlation is not discussed in details because the rank test is based on sorting algorithms that are computationally intensive for the large data sets. Instead, a robust correlation test (sign test) is considered, that is much simpler to use and easy to implement.

Let us consider the following sample table as per the 3D, OLAP Cube representation for vehicle insurance policy system. The sample data is shown in table 2. The data samples are POPICY_ID, DATE_OF_PURCHASE and BIKE_MODEL ,and can be connected for the N policy.

Table2. Sample Data from Vehicle insurance Policy System

POLICY_ID	DATE_OF_PURCHASED	BIKE_MODEL
33468522	18-Mar-10	DISCOVER
33468523	18-Mar-10	PLATINA
33468548	19-Mar-10	DISCOVER
33529624	20-Mar-10	DISCOVER
33529625	21-Mar-10	PULSAR 135 DTSI
33529627	22-Mar-10	PULSAR 135 DTSI
33529628	22-Mar-10	PULSAR 135 DTSI

3. IMPLEMENTATION

3.1 Data Cube Design

The OLAP Data cube formed from data sample which is a 3-dimentional representation and the cells are POLICY_ID, DATE_OF _PURCHASE and BIKE_MODEL of the cube representation with combination of attributes as shown in Table 2.

A sample Data cube for the front view of the data cube attributes is shown in fig.1. Data Cube can be used to retrieve information from the sample data of the vehicle insurance policy system. The main aim of this design is to retrieve the decision support information from the data cube in a very easiest and faster way. Queries have performed on the given sample data cube to retrieve necessary information.

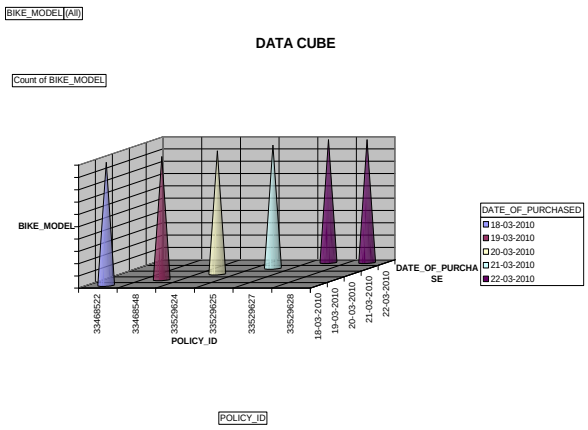

Fig. 1 View of Sample OLAP Data Cube

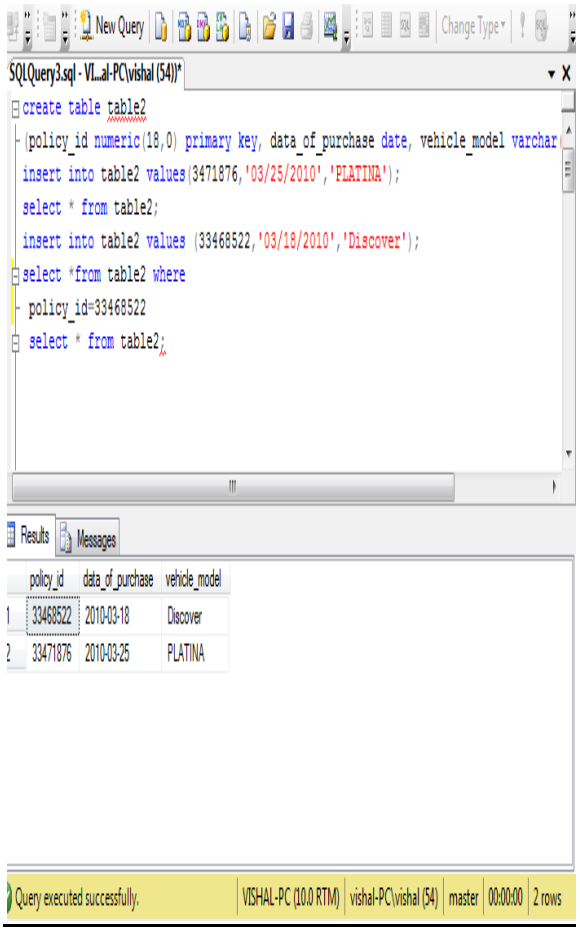

Fig. 2 Data cube Queries

3.2 Correlation values

The sample data is shown below in the following table

Table 3: Sample Table from Bajaj Allianz Vehicle Insurance

POLICY_ID	BIKE_MODEL	X (Policy_No.)	Y (Model_no.)	xy	X^2	Y^2
33468522	DISCOVER	1	1	1	1	1
33468523	PLATINA	2	2	4	4	4
33468548	DISCOVER	3	1	3	9	1
33529624	DISCOVER	4	1	4	16	1
33529625	PULSAR 135 DTSI	5	3	15	25	9
33529627	PULSAR 135 DTSI	6	3	18	36	9
33529628	PULSAR 135 DTSI	7	3	21	49	9
33471849	DISCOVER	8	1	8	64	1
33471851	PLATINA	9	2	18	81	4
33471852	DISCOVER	10	1	10	100	1
33471853	PLATINA	11	1	11	121	1
33471854	PLATINA 125 DTS-I	12	4	48	144	16
33471859	PULSAR 150 DTS-I	13	5	65	169	25
334718560	PLATINA	14	2	28	196	4
33551306	PULSAR 135 DTS-I	15	3	45	225	9
33551307	DISCOVER	16	1	16	256	1
33551308	DISCOVER	17	1	17	289	1
33551309	PULSAR 150 DTS-I	18	5	90	324	25
33551312	PLATINA	19	3	57	361	9
33551313	PLATINA	20	2	40	400	4
33551314	DISCOVER	21	1	21	441	1
33551315	DISCOVER	22	1	22	484	1
33551316	PULSAR 150 DTSI	23	5	115	529	25
33551317	PULSAR 150 DTSI	24	5	120	576	25
33551319	PULSAR 150 DTSI	25	5	125	625	25
1139138211		325	62	922	5525	212

To find correlation values of sample data of vehicle insurance policy system we put the values of x and y in the above equation (1) and find the value of r (correlation coefficient).

Correlation coefficient (r) = 0.94895431.3810

4. CONCLUSIONS

From the above work, it is observed that the OLAP cube is a versatile technique to store large amount of databases and one can apply search for several number of records in a fraction of seconds. The correlation computed in the above work shows that the data binding is excellent. The present work is confined only for the numerical data, however, the work can be extended for text storage of data and also for the audio and video files. For the complex system, the presented technique is suitable for long database as well as for the complex queries which may be written in the form of object-oriented databases.

5. REFERENCES

[1] James Rumbaugh, ER Is UML, Journal of Information Systems Education, Vol. 17(1), 2006.

[2] Chaudhuri, S. and Dayal, U., An Overview of Data Warehousing and OLAP Technology, SIGMOD Record Volume 26, Number 1, September 1997.

[3] Samtani, S., Mohania, M.K., Kumar, V. and Kambayashi, Y., Recent Advances and Research Problems in Data Warehousing, ER Workshops 1998.

[4] Hurtado, C., Mendelzon, A. and Vaisman, A., Maintaining Data Cubes under Dimension Updates, Proc IEEE/ICDE' 99.

[5] Codd, E. F., Codd, S.B. and Salley, C. T., Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate, Technical report, 1993.

[6] Connolly, T. and Begg, C., Database System: A Practical Approach to Design, Implementation, and Management, Addison-Wesley Longman, Inc., 1999.

[7] Buzydlowski, J. W., Song, II-Y. and Hassell, L., A Framework for Object-Oriented On-Line Analytic Processing, DOLAP 1998.

[8] Cabibbo, L. and Torlone, R., A Logical Approach to Multidimensional Databases, EDBT 1998.

[9] Kimball, R., The Data Warehouse Lifecycle Toolkit, John Wiley & Sons, Inc., 1998.

[10] Cabibbo, L. and Torlone, R., A Logical Approach to Multidimensional Databases, Lecture Notes in Computer Science, number 1377 in proc. of the 6th Int. Conf. On Extending Database Technology, (EDBT'98), pages 183-197, Valencia, March, 1998.

[11] Nguyen. T. B., Tjoa, A M. and Wagner, R. R. Conceptual Object Oriented Multidimensional Data Model for OLAP, Technical Report, IFS, Cited by 51, Vienna 1999.

[12] Trujillo J. and Palomar, M., An Object Oriented Approach to Multidimensional Database, Conceptual Modeling (OOMD), DOLAP, 1998.

[13] Codd, E. F., Codd, S.B. and Salley, C.T., Providing OLAP to User-Aanalysts: An IT Mandate". Technical report, 1993.

[14] Kimball, R., The Data Warehouse Lifecycle Toolkit, John Wiley & Sons, Inc., 1998.

[15] Gyssens. M. and Lakshmanan. L., A Foundation for Multi-Dimensional Databases, In the 33rd Intl. Conf. On Very Large Database Conference (VLDB'97), pp 106-115, 1997.

[16] Hurtado C., Mendelzon A. and Vaisman A., Maintaining Data Cubes under Dimension Updates, proceedings in the fifteenth international conference on Data Engineering, Washington DC (pp.346-355).

[17] Vassiliadis, P., Modeling Multidimensional Data Bases, Cubes and Operations, In proc. 10th Scientific and Statistical Database Management Conference (SSDBM '98), Capri, Italy, June 1998.

[18] Mangisengi O., Tjoa A. M. and Wagner R. R., Multidimensional Modeling Approaches for OLAP, Proceedings of the Ninth International Database Conference "Heterogeneous and Internet Databases", 1999, ISBN 962-937-046-8.Ed.J. Fong, Hong Kong, 1.

[19] Chaudhuri S and Dayal U., "An Overview of Data Warehousing and OLAP technology", ACM Sigmod Record vol. 26 (1), March 1997.

6. AUTHOR'S PROFILE

Dr. Narander Kumar received his Post Graduate Degree and Ph. D. in CS & IT, from the Department of Computer Science and Information Technology, Faculty of Engineering and Technology, M. J. P. Rohilkhand University, Bareilly, Uttar Pradesh, INDIA in 2002 and 2009, respectively. His current research interest includes Quality of Service (QoS), Software Engineering, Computer Networks, Resource Management Mechanism, in the networks for Multimedia Applications, Performance Evaluation. Presently he is working as Assistant Professor, in the Department of Computer Science, Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow, INDIA.

Vishal Verma is a research scholar in Department of Comupter Science, Babashaheb BhimRao Ambedkar University, and Lucknow, India. Earlier he got his Master of Computer Application (MCA) from the above University and presently he is working on Data Mining Applications through UML.

Vipin Saxena is a Professor and Head, Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India. He got his M.Phil. Degree in Computer Application in 1992 & Ph.D. Degree work on Scientific Computing from University of Roorkee (renamed as Indian Institute of Technology, Roorkee, India) in 1997. He has more than 17 years of teaching experience and 20 years of research experience in the field of Scientific Computing & Software Engineering. He has published more than hundred International and National research papers and authored four books in the Computer Science field. Dr. Saxena is a life time member of Indian Science Congress.