OLAP CUBE REPRESENTATION FOR OBJECT-ORIENTED DATABASE

Vipin Saxena[1] and Ajay Pratap[2]

Department of Computer Science
Babasaheb Bhimrao Ambedkar University
(A Central University) Lucknow (U.P.), India

E-Mail: [1]vsax1@rediffmail.com,[2]pratap_aj@yahoo.co.in

ABSTRACT

In the current scenario, the size of database related to any organization is rapidly increasing and due to evolution of the object-oriented approach, many of the Software Industries are converting the old structured approach based softwares into the object-oriented based softwares. Therefore, for the large amount of database, it is necessary to study the faster retrieval system as On-Line Analytical Processing (OLAP) which was introduced by E.Codd in 1993. The present paper is an attempt in this direction and object-oriented approach with the help of Unified Modeling Language (UML) is used to create OLAP cube which supports three dimensional data. A real case study of Indian Postal Services is considered to store large amount of data in the OLAP cube and can be easily used for the analysis purpose. UML class model, sequence diagram and star schema of object-oriented database are designed along with the OLAP cube. For testing purpose, some sample queries are also performed on the object-oriented database.

KEYWORDS

OLAP, Cube, UML, Class Diagram, Sequence Diagram and Star Schema.

1. INTRODUCTION

Information Technology (IT) has revolutionized the past working of the Indian Postal Services to do more business and profits. By the use of IT infrastructure, Indian Postal Offices have started various kinds of services like Electronic Money Transfer, On-line payment of premium for various schemes, etc. In India, more than ten thousand post offices are functioning, and currently 8263 post offices have already been computerized but still lots of computerization work is needed in the remaining Indian Post Offices. The post office savings bank scheme is the largest savings bank in the country in terms of network, and having more than 164.3 million accounts. In this research paper, a technique is proposed to store the non-postal data in the form of OLAP cubes. An OLAP solution allows the users to have quick access to summarized data, which is generated from a large amount of detailed information.

OLAP cube performs a multi-dimensional analysis of business data. These data cubes also provide the capability of complex calculations, patterns and trends also. It provides intelligent solution that may be related to forecasting, financial reporting, budgeting, and knowledge discovery. OLAP databases are also known as multi-dimensional database (MDDB). OLAP

DOI : 10.5121/ijsea.2012.3209

technology has two major categories, which are: Relational OLAP (ROLAP) and Multi-Dimensional OLAP (MOLAP). Their combined approach is a Hybrid OLAP.

The non-postal services are Public Provident Fund (PPF), National Savings Certificate (NSC), Kisan Vikas Patra (KVP), Savings Bank Account (SBA) and Recurring Deposit Account (RDA) [1]. Data warehousing and on-line analytical processing (OLAP) technologies are one of the most important decision support systems in the recent years [2]. OLAP technology is based on a Multidimensional Data model, called as MD, whose main parts are a dimension and a fact table. Cabibbo and Torlone presented a logical approach for the multidimensional database. Lehner et al. [4] proposed the various normal forms for multidimensional database with key concepts in OLAP which are measures, facts and data cubes. The measures are the numerical attributes analyzed against the different dimensions, which can built-in functions like sum, count and average, as well as complex, user-defined formulas. Dimensions provide the context for the measure and together they constitute facts. A set of measures aggregated according to a set of dimensions comprise a data cube. In this context, various kinds of OLAP solutions are proposed by Thomsen et al. [5]. Several data cubes can be built from the same database for different analysis needs. On the other hand, in relational DBMS, the multidimensional model of the data warehouse is mapped in most cases through star schemes as explained by Kimball [6] consisting of a set of dimension tables and a central fact table. Dimension tables are strongly de-normalized and then they are used to select the facts of interest based on the user queries. The fact table stores fact attributes; its key is defined by importing the keys of the dimension tables. Different versions of these base schemes have been proposed by Barquin and Edelstein [7] in order to improve the overall performances. In case of queries performed on very large databases, response time should be small and query optimization is also a critical task. The user views the data in the form of multi-dimensional data cube. Materialization of cells is the most powerful query optimization technique. Harinarayan et al. [8] presented the applications of greedy algorithms to materialize the cells of data cubes. Mining of stream data, social networks data, spatial and multimedia data came in the category of new mining topics. Mining techniques for these complex data and algorithms are discussed by Han and Kamber [9]. Saxena et al. [10] has presented a security model for credit/debit card system and suggested to store the data in the form of data cubes with practical electronic protocol suggested for safe and secure transaction. Data warehouses have large amount of data but they are supposed to give quick answers of queries. Input output efficient technique based upon multi resolution wavelet decomposition is used for the representation of data cube by Vitter et al. [11]. The presented technique is better than the approximation technique. Outlier analysis is used in the field of fraud detection, network robustness analysis etc. These data contain hundred of dimensions. Many recent algorithms use concepts of proximity in order to find the outliers, which are based on their relationship to the rest of the data. New techniques for outlier detection which find the outliers by studying the behavior of projections from the data set are discussed by Aggarwal and Yu [12]. Rumbaugh et al. [13] discussed the UML concepts and better representations such as two-color diagrams, coverage of both semantics and notation and better explanation of concepts through diagrams. Sequence and state diagrams are used for the validation and performance evaluation of the system but the UML lacks a formal semantics. Bernardi et al. [14] have proposed a technique which translates the state charts and sequence diagrams into generalized stochastic petri nets and it helps for the validation and performance evaluation. Martin [15] described about the efficient object-oriented software design using the latest version of the Industry-standard for UML 2.0 version. Chaurasia and Saxena [16] presented a mobile based electricity bill deposit system through UML with a real case study of electricity department.

2. BASIC CONCEPTS AND TERMS OF OLAP

E. Codd introduced the term OLAP (Online Analytical Processing) in 1993. The objective of the OLAP is to facilitate solving data analysis problems and accurate decision-making. Codd formulated 12 features of the OLAP data, and the majority of modern OLAP meet these features. Twelve features of the OLAP are further transformed into four key definitions formulated by N. Pendse. According to these four key definitions, an OLAP system should have the following properties:

1. It should be fast, which can provide almost instantaneous response to the majority of queries.
2. It should be a multi-user system, which can control access to data and simultaneous works done by many users.
3. It should be multidimensional.
4. It must provide information, which means that data should be complete from the point of view of an analyst.

Measures and dimensions define an OLAP cube. Measures are based on one or more fact table and the dimensions are based on one or more dimension tables. Dimensions are the attributes, which are mapped to a column in the dimension tables.

3. A CONCEPTUAL MODELING OF OLAP FOR INDIAN POSTAL SERVICES

The first step is to design the UML class model which shows the static behavior of the system and in the next step a sequence diagram is designed to represent the dynamic behavior of the system. After that a snowflake schema is designed for the object-oriented databases. These are described below in brief:

3.1. UML Class Diagram

The UML class diagram shows the static structural behavior of the system, in which attributes and operations are designed for the complete system. The classes can be related to each other in number of ways, like they can be associated, dependent, specialized or packaged. A system can have a number of class diagrams because not all classes participate in a single class diagram. Figure 1 presents the class diagram of the post office system. The class diagram has eight persistent classes, which are Person, Address, Customer, Postal_wrkr, Postal_Agent, Schm_Invnt, Scheme, Sales_Report and one transient class Control. These classes are connected to each other by various relationships with their multiplicities as shown in the diagram and the Control class is used to manage all persistent classes. Customer, Postal_wrkr and Postal_Agent are inherited from non-abstract class Person.

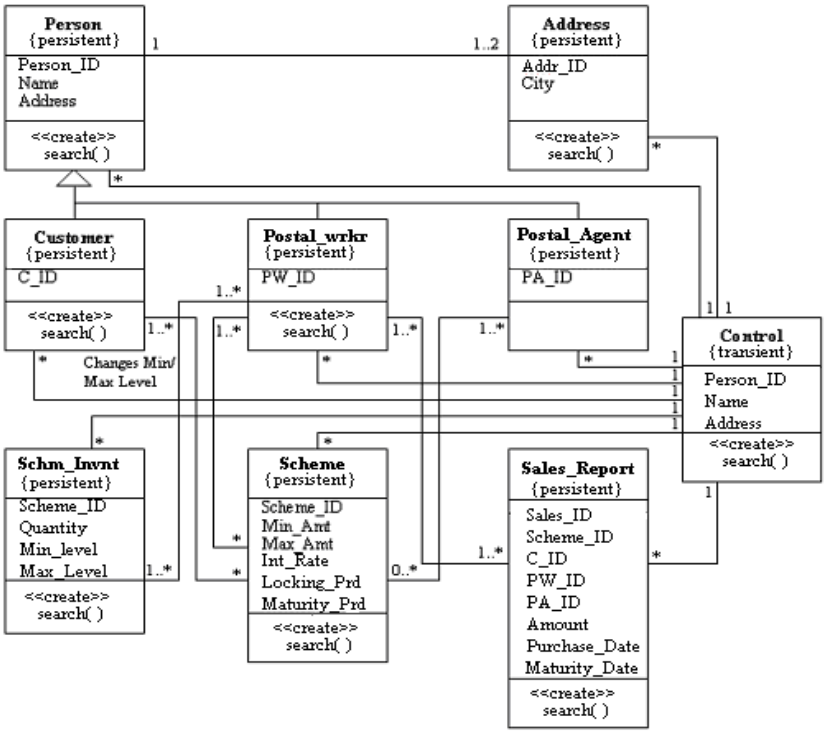

Figure 1. Post Office's UML Class Diagram

3.2. UML Sequence Diagram

Sequence diagram shows how objects interact by sending messages to each other. The vertical axis shows time and lifeline of the object and the horizontal axis shows the messages, which are being sent from the sender to the receiver.

Sequence diagram shows five objects Customer, Postal_Agent, Postal_wrkr, Schm_Invnt and Sales_Report. The diagram shows the interaction among these objects. In this diagram, the customer sends a request to the postal agent or to a postal worker directly. After that the postal worker checks the availability of the inventory, validation and integrity. If the request is valid then payments is made by the customer and after updating the scheme inventory status, a sales report is generated by the postal worker.

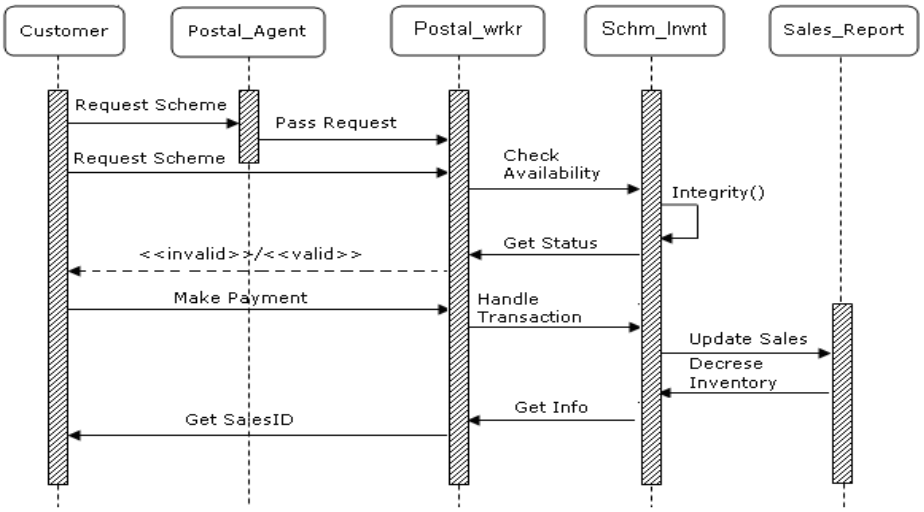

Figure 2. Post Office's UML Sequence Diagram

3.2. Schema Designing for OLAP Cube

To construct an OLAP cube, one needs to think about the measures and the dimensions. In this work, authors created OLAP cube for the post office data warehouses which store the non-postal data. This is useful for the top management and also for the users who can perform query from the cube.

In the designing of cube, five dimensions are considered which are Scheme, Customer, Location, Time and Scheme_Invnt. Dimension Scheme, Customer, Location, Time and Scheme_Invnt are related to dimensional table "Scheme", "Customer", "Location", "Time" and "Scheme_Invnt", Their description are given below:

Scheme (Scheme_ID, Min_Amt, Max_Amt, Int_Rate, Locking _Prd, Maturity_Prd);

Customer (Cust_ID, Cust_Name, Cust_mob_No, Cust_Address);

Location (Location_ID, Locality, City, State, Country);

Time (Time_ID, Day, Month, Quarter, Year);

Scheme_Invnt (Scheme_ID, Quantity, Min_Level, Max_Level);

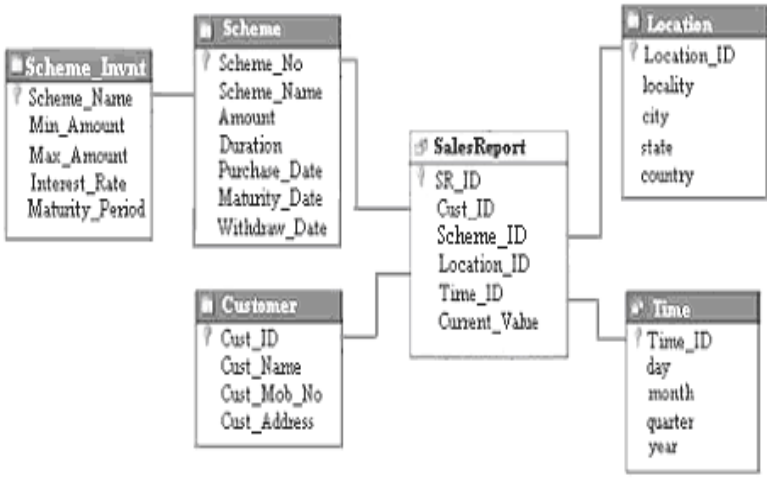

Figure 3. Snowflake Schema for Conceptual Designing

The above figure shows snow-flaking schema and a fact table is created which contains all the key attributes of the dimension table.

On the basis of records available in the dimension table, the description of fact table "Sales Report" is given below:

Sales_Report (SR_ID, Cust_ID, Scheme_ID, Location_ID, Current_Value);

Data warehousing needs a subject-oriented schema, which can facilitate online analysis of data. Multi-dimensional model is also used for this purpose, which can be in the form of star schema, snowflake schema or fact constellation schema. In the present work, authors are using snowflake Schema where scheme dimensional table is normalized and further divided into Scheme_Invnt table.

Data mining query language is used to specify the data mining task. Cube definition and dimension definition are given below as represented in Figure 4.

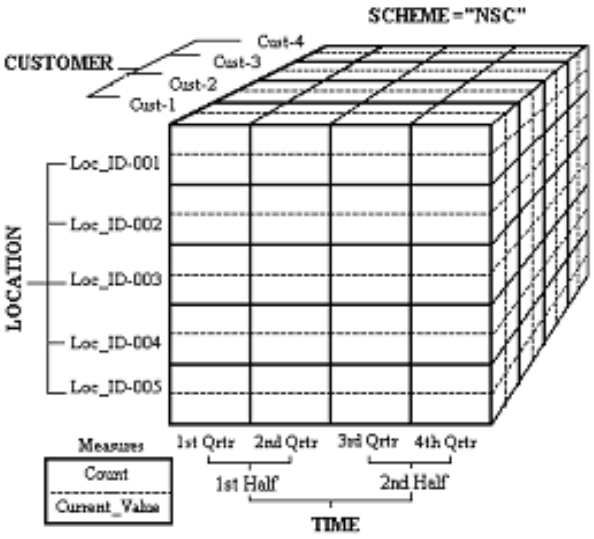

Figure 4. OLAP Cube for the given Schema

Defining a Snowflake Schema in DMQL:

define cube sales [Time, Customer, Scheme, Location]:

Current_Value = Schemes_Sold = Count(*)

define dimension time as (Time_ID, Day, Month, Quarter, Year)

define dimension customer as (Cust_ID, Cust_Name, Cust_Mob_No, Cust_Address)

define dimension location as (Location_ID, Locality, City, State, Country)

define dimension scheme as (Scheme_ID, Scheme_Name, Min_Amt, Max_Amt, Int_Rate, Locking_Prd, Maturity_Prd (Quantity, Min_Level, Max_Level))

For testing purpose, some of the sample queries are performed on the database of Indian Postal Services and corresponding results are given below:

Sample Query I

This query gives total number of schemes, which are sold in particular location and the group by clause is generalized by using the cube construct.

Select Scheme_ID, Location_ID, Count (*)

From SalesReport

Group by cube (Location_ID)

Sample Query II

This query searches name–wise detail, those has purchased any particular scheme in any particular city. Two dimensional tables Customer & Location and fact table SalesReport is used.

```
Select Cust_Name, Scheme_ID, City

From  SalesReport, Customer, location

Group by cube (Location_ID)
```

Sample Query III

The given query gives the total number of schemes sold in any particular month, in particular city and interest rate given on that scheme. Rollup operator is used to generate aggregates at multiple level of a hierarchy on a column.

```
Select Scheme_ID, Int_Rate, Month, City, Count (*)

From  SalesReport,  location, Scheme

Where SalesReport. Scheme_ID = Scheme. Scheme_ID

Group by rollup (Scheme_ID)
```

5. CONCLUSIONS

In India, many of the post offices are functioning manually while the database related to the various kinds of postal services is very long, therefore, present work is an attempt to create OLAP cube for storing a large database and one can perform various queries and get result within a few seconds. Implementation of the present work requires the computerization of all the branches of Indian postal services. Now days the system is using relational database management system but using the proposed UML model, one can also use the Object Oriented database stored in the form of OLAP cube, which is more powerful in complex situation and queries.

REFERENCES

[1] "Indian Post Offices to go Hi-Tech," 25 November 2010. www.indiapost.gov.in (Accessed on 12th August 2011).

[2] S. Chaudhuri and U. Dayal, "An Overview of Data Warehousing and OLAP Technology," In SIGMOD Record, Vol. 26, Issue 1, pp. 65-74, March 1997.

[3] L. Cabibbo and R. Torlone, "A Logical Approach to Multidimensional Databases," 6th International Conference on Extending Database Technology (EDBT'98, Valencia, Spain), G. Alonso (eds.), LNCS, Vol. 1377, Springer, 1998.

[4] W. Lehner, J. Albrecht and H. Wedekind, "Normal Forms for Multidimensional Databases," In Proceedings of the 10th International Conference on Scientific and Statistical Data Management (SSDBM'98), Capri, Italy, 1998.

[5] E. Thomsen, G. Spofford and D. Chase, Microsoft OLAP Solutions, John Wiley and Sons, 1999.

[6] R. Kimball, The Data Warehouse Toolkit, 2nd ed., John Wiley & Sons, 1998, pp 63-70.

[7] R. Barquin and S. Edelstein, Planning and Designing the Data Warehouse, Prentice Hall, 1997.

[8] V. Harinarayan, A. Rajaraman, and J. Ulman, "Efficient Implementation of Data Cube via Materialized Views," In Proceedings of ACM SIGMOD Conference, Vol. 25(2), Montreal, Canada, 1996.
[9] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd ed., Morgan Kaufmann Publishers, , March 2006.
[10] V. Saxena, N.M.P. Verma and A. Pratap, "A Data Mining Technique for a Secure Electronic Payment Transaction," International Journal of Economics and Finance, Vol. 2(4), pp. 12-17, 2010.
[11] J. S. Vitter, M. Wang, and B. Iyer. "Data Cube Approximation and Histograms via Wavelets," In Proceedings of the 7th International Conference on Information and Knowledge Management, pages 96–104, November 1998.
[12] C. Aggarwal and P. Yu, "Outlier Detection for High Dimensional Data," Proceedings of the 2001 ACM SIGMOD international conference on Management of data (SIGMOD '01), Volume 30 Issue 2, June 2001.
[13] J. Rumbaugh, I. Jacobson, and G. Booch., The Unified Modeling Language Reference Manual, 2nd ed. Addison-Wesley, Boston, 2005.
[14] S. Bernardi, S. Donatelli, and J. Merseguer, "From UML Sequence Diagrams and Statecharts to Analysable Petri Net Models," In Proceedings of the Third International Workshop on Software and Performance, pages 35–45, New York, USA, ACM Press, 2002.
[15] F. Martin., UML Distilled: A Brief Guide to the Standard Object Modeling Language, 3rd Edition, Addison- Wesley, Boston, 2000.
[16] P.K. Chaurasia and V. Saxena, "Mobile Based Electricity Bill Deposit System Through UML," Journal of Software Engineering and Applications, Vol. 4(3), 187-190, 2011.

Authors

Dr. Vipin Saxena is a Professor & Head, Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India. He got his M.Phil. Degree in Computer Application in 1991 & 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 16 years teaching experience and 19 years research experience in the field of Scientific Computing & Software Engineering. Currently he is proposing various software designs by the use of Unified Modeling Language for the research problems related to the Software Domains & Advanced Computer Architecture. He has published more than 91 International and National research papers in various refereed Journals and authored four books covering Software Engineering, E-Learning and Operating System. Dr. Saxena is a life time member of Indian Congress.
Mobile: 0091-9452372550
E-mail: vsax1@rediffmail.com

Ajay Pratap is a research student in Department of Computer Science, Babashaheb Bhimrao Ambedkar University, Lucknow, India. Earlier, he got his M.Phil. Degree in Computer Science in 2008 and Master of Computer Application from the said University & presently he is working on performance evaluation of object oriented database through UML.
Mobile: 0091-9936-189152
E-mail: pratap_aj@yahoo.co.in