Object Oriented Data Modeling for Data Warehousing (An Extension of UML approach to study Hajj pilgrim's private tour as a Case Study)

Kamal Alaskar[1] and Akhtar Shaikh[2]
[1]Bharati Vidyapeeth University, Institute of Management Kolhapur, India
[2]The New College, Kolhapur, India

Abstract- *We are all familiar with the growth rate of the public Web. Regardless of the metric used to measure its growth attached networks, servers, users or pages the growth rate continues to exhibit an exponential pattern. In the same vein, the adoption rate of intranet and extranet data warehouses (i.e., Web warehouses) has exhibited a similar pattern, although the pattern has lagged public adoption. While data warehouse and business intelligence vendors have offered Web-enabled versions. All over world there are hundreds of private tour operators (PTOs) managing tour for hajj pilgrims. According to the visiting Indian Ministry 138,000 Indian pilgrims performed hajj during 2007.Out of above mentioned figure 110,000 through the Hajj committee and 38,000 through PTOs. In India there are about 397 registered PTOs up to 2007. We introduce the UML Profile for Modeling DWH Usage for modeling the different kinds of DWH usage on a conceptual level. It uses features of UML intended for the purpose of creating abstract, general models. The profile distinguishes four perspectives of usage, and allows to model details of the users. The UML Profile is applied to example illustrating Hajj pilgrims private tour.*

Keywords- *Business intelligence, data modeling, private tour operator, unified modeling language, web enabled warehouse*

Received December2, 2008; Accepted March 17 2009

1. Introduction

The end of the 20th Century has seen the rapid development of new technologies such as web-based, communication, and decision support technologies. Companies face new economical challenges such as e-commerce or mobile computing, and are drastically changing their information systems design and management methods. They are developing various technologies to manage their data and their knowledge. These technologies constitute what is called business intelligence. These new means allow them to improve their productivity and to support competitive information monitoring. In this context, the web is now the main farming source for companies, whose challenge is to build web-based information and decision support systems. Our work lies in this field. We present in this paper an approach to build a Decision Support Database (DSDB) whose main data source is the web.

The Unified Modeling Language (UML) [1], [2], [12] has become a standard for object modeling during analysis and design steps of software system development. If we use it as a modeling technique for development of an information system, which will use a RDBMS or ORDBMS to store persistent objects, it is necessary to add the ability to model not only objects, but also relational tables and database structures of the object extensions. Such modeling will be called object-relational in this paper.

A Data Warehouse (DWH) system is more than just a big database. Defined as "a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management's decision-making process" [3], DWH systems represent a single source of information for analyzing the status, the development and the results of an organization [4]. Analysts and decision makers take measures such as the number of transactions per customer or the increase of sales during a promotion and use them to recognize trends or warning signs and to decide on future investments.

In our study decision makers are the Government bodies dealing with Hajj section. In order that such data is put to an effective use in facilitating decision-making, a data base can be constructed over the historical data. It permits several types of queries requiring complex analysis on data to be addressed by decision makers. Pilgrims facilities are the core objective of decision maker body. For Pilgrims of different state, it is important to have the right information about the facility provided by the private tour operators (PTOs). Such a data base can be beneficial to both the decision makers and pilgrims as well be described in the following subsections.

1.1 Utility to the Decision Makers

They can perform extensive analysis of stored data to provide Gradation to the PTOs and answers to the exhaustive queries to the Pilgrims performing through PTOs. This helps them to formulate more effective strategies and policies for Pilgrims facilitation.

1.2 Utility to the pilgrims.

They are the ultimate beneficiaries of the new policies formulated by the decision makers. They can view frequently asked queries whose results will already be there in the database and will be immediately exhibited to the user saving the time required for processing. They can see the gradation of PTOs and accordingly registered their names based on performance from any where.

2. Research Objectives

This paper aims to answer the following research questions:

- How can the relationship between the Data Warehouse and the structure, behavior, and goals of the organization be formally described?
- How can the relationship between the Data Warehouse and the structure, behavior, and goals of the organization support the interpretation of data?

3. Problem Area and Business Intelligence

The Government of Saudi Arabia has notified that only registered Private Tour Operators (PTO) involved in Hajj Pilgrimage with Government of India will be eligible for grant of (Group) Hajj visas subject to fulfillment of other conditions that may be laid down by the concerned Saudi Authorities. Non-registered PTOs will not be eligible for applying for the Group Hajj Visas.

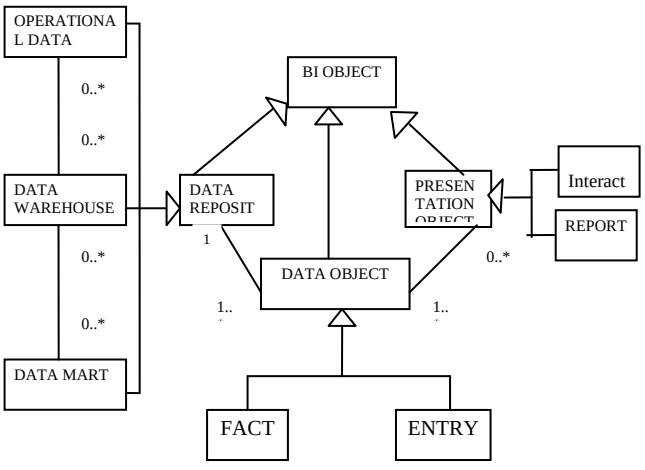

Figure1. Meta model for business intelligence

"All India Hajj Umrah Tour Organizers Association (AIHUTOA)" was established in the year 1997 at Mumbai. During last Eight years many tour organizers have become the members of this association. This association is registered and recognized by the Government of India's Ministry of External Affairs since the last four years. Similarly this association is duly recognized by all the concerned Hajj offices and Departments in Saudi Arabia.

We have identified three main categories of Business Intelligence objects: Data Repositories (representing the elements of the DWH architecture), Data Objects (representing the data model of a certain repository), and Presentation Objects (representing the means of presentation, either a static report or an interactive analysis). The relationships between the Business Intelligence objects are shown in Fig. 1.

Business Intelligence objects chosen for a model depend on the target audience and the level of detail of the model. In an overview business process model suited for DWH managers, one might show the DWH or individual data marts as a whole. In a more detailed model for developers, sub-processes can be described as accessing individual entities and facts. Additionally, decision makers often receive relevant data in form of reports, for instance a report on pilgrim's data for the past five year, which may also be relevant for business process modeling.

4. Modeling Data Warehouse

Our approach applies UML to the Data Warehousing domain. It is aimed at encompassing all the different ways that users may use a DWH. Our goal is to provide an overview over all aspects of DWH usage, not only focusing on the data model. Nevertheless, due to the special characteristics of DWH data, it is necessary to take the data model especially into account.

DWH paradigm, which allows data accessing a way that comes more natural applications involve complex queries on large amounts of data, which are difficult to manage for human analysts. In Data Warehousing, data is often organized according to the multidimensional to human analysts. The data is located in n-dimensional space, with the dimensions representing the different ways the data can be viewed and sorted (e.g. according to time, store, customer, product, etc.).

A multidimensional model, also called star schema or fact schema, is basically a relational model in the shape of a star (see Fig. 2 for an example). At the center of the star there is the fact table. It contains data on the subject of analysis (e.g. sales, transactions, repairs, admissions, expenses, etc.). The attributes of the fact table (e.g. cost, revenue, amount, duration, etc.) are called measures or fact attributes. The spokes/points of the star represent the dimensions according to which the data will be analyzed

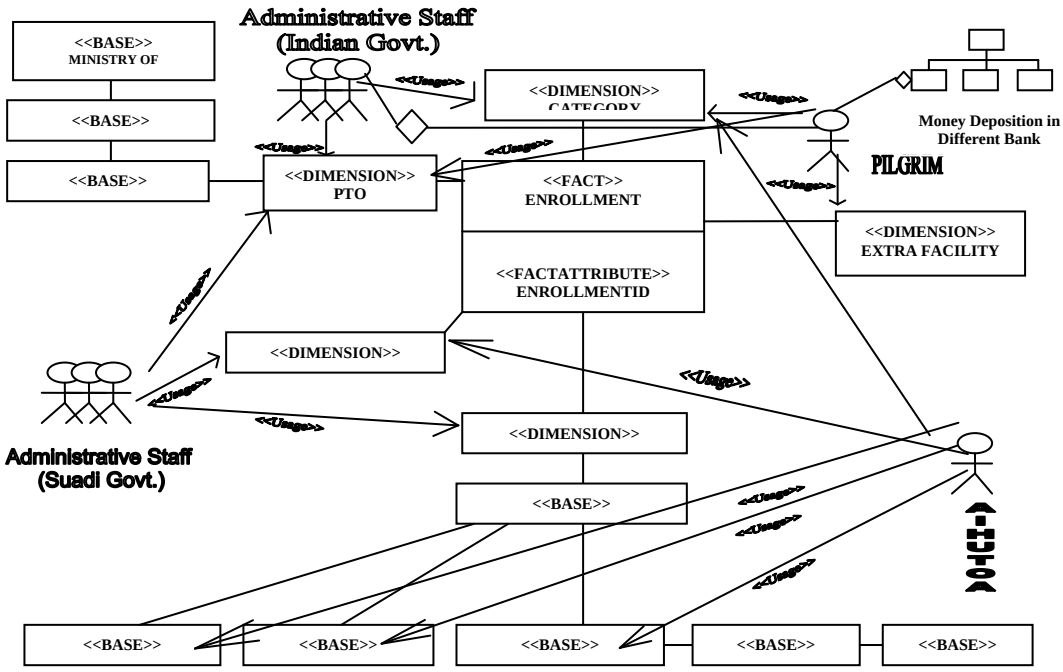

Figure 3. Hajj pilgrims enrollment

(sorted/aggregated by data, by store). The dimensions can be organized in hierarchies that are useful for aggregating data (e.g. store, city, region, country). Stars can share dimensions, thus creating a web of interconnected schemas that makes drill-across operations possible.

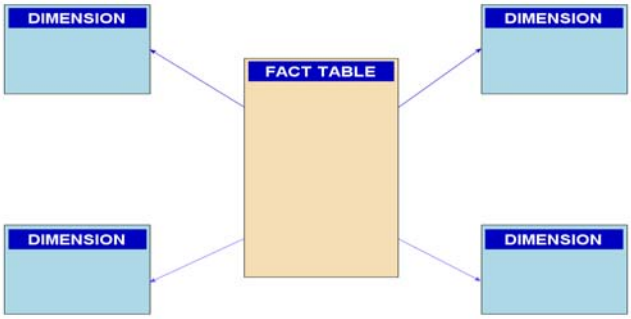

Figure 2. Star schema

5. Conceptual Model and UML Profile for Modeling Data Warehouse

We first designed a conceptual UML model for a complex object representing a super class of all the types of multiform data we consider [5]. Note that our objective is not only to store data, but also to truly prepare them for analysis, which is more complex than a mere ETL (Extracting, Transforming, and Loading) task. Then, we translated our UML conceptual model into an XML schema definition that represents our logical model. Eventually, this logical model has been instantiated into a physical model that is an XML document. In order to provide models of DWH usage that is useful to different application scenarios.

Our goal is to achieve a broad view of usage, while maintaining concise models. The objective of this paper is to introduce the way in which object-relational target environment can be modeled in the UML. But we focused on only fundamental features of the approach. Rational's white papers [9] concerning relational data modeling in the UML may provide a useful information too.

There are many approaches to modeling the multidimensional data structures of data warehouses, some of which are object-oriented models or based on the Unified Modeling Language (UML) [13, 11, 14]. For modeling multidimensional data, we choose to use the UML Profile of Luj´n-Mora et al. as described in [11]. This Profile allows to model not only the core features of multidimensional models (facts, measures, and dimensions), but also many advanced features such as degenerate dimensions or non strict and complete dimensional hierarchies, and also provides three levels of detail.

In this section we illustrate the use of the UML Profile for Modeling Data Warehouse Usage with an example (Fig.3) It also useful in a later stage of the DWH design process may serve as rough input. Diagrams like this make it possible to identify preliminary groups of users, based on their data needs.

6. Data Collection From Different Sources (Locations)

The City of Mumbai has a very long association with Hajj. Pilgrims have been proceeding for Hajj through

Sea Route during British Rule and earlier, starting their holy journey from the Sea-Port of Mumbai. After introduction of air travel, it was no more necessary for all the Hajj Pilgrims to reach Mumbai. Gradually other Embarkation Points were introduced. The number of Embarkation Points went on increasing year after year as number pilgrims increased. During Hajj 1428 (H) – 2008 (A.D.), there were Sixteen (16)

These are the points from where the pilgrims fly to Jeddah and from there to Makkah for performing Hajj and returning back after performing holy Hajj. Primary Data collection can be taken place from these points manually. The growth rate of hajj pilgrims is shown in the Table 1. Since last four years Saudi Government and Government of India giving permission to group visas around 40000 pilgrims to perform hajj through PTO's.

Primary Data collection can be taken place from above mentioned locations manually but it would be very interesting to mine them automatically with text mining [6], image mining [7], or XML mining [8] techniques.

Table 1. Previous record showing increase of Pilgrims

HAJ	PILGRIMS PROCEEDED		
YEAR	SHIP	AIR	TOTAL
1990	4625	19602	24227
1991	4528	18790	23318
1992	4723	19494	24217
1993	4562	20643	25205
1994	4650	21035	25685
1995	CLOSED	30503	30503
1996		50346	50346
1997		53770	53770
1998		63592	63592
1999		62100	62100
2000		71909	71909
2001		71133	71133
2002		70276	70276
2003		69795	69795
2004		71711	71711
2005		80772	80772
2006		99660	99660
2006-II		1,08,816	1,08,816
2007		1,10,415	1,10,415

*Source: http://hajcommittee.com/index.php?value=records

7. Conclusion

In this paper, we have introduced the UML Profile for Modeling DWH Usage for modeling the different kinds of DWH usage on a conceptual level. It allows to model details of the users, such as functional grouping or organizational affiliation. We base "usage" on UML information flows, which are intended for a general representation of information exchanges. Also the aim of this paper is to address the issue of web data integration into a database. This constitutes the first phase in building a multiform data warehouse. We propose a modeling process to achieve this goal.

This paper has targeted the relationship between data warehouses and the structure, behavior and goals of the organization. It consists of model of data warehouse usage, which includes modeling the users, users groups, and user skill levels, the intensity with which they use the data warehouse infrastructure.

References

[1] UML Modeling Language Specification. Version 1.3, June 1999. Available at http://www.rational.com/uml/resources/documentation/ (March 2009).

[2] Booch G., Rumbaugh J., and Jacobson I., *The Unified Modelling Language User Guide, Addison-Wesley Longman*, p.482, 1999.

[3] Inmon W., Hackathorn, R., "Using the data warehouse," *Wiley-QED Publishing, Somerset, NJ, USA* ,1994.

[4] Kimbal R., Reeves L., Ross M., Thornthwaite, W., "*The Data Warehouse Life cycle Toolki,"* John Wiley & Sons, Inc.",1998.

[5] Miniaoui S. et al.; "Web data modeling for integration in data warehouses," *First International workshop on Multimedia Data And Document Engineering (MDDE 01)*, France, 88-97,2001.

[6] Tan A., "Text Mining: The state of the art and the challenges," *Workshop on Knowledge discovery from Advanced Databases (KDAD 99)*, China, 71-76; 1999.

[7] Zhang J. et al. "An Information-Driven Framework for Image Mining," *12th International Conference on Database and Expert Systems Applications (DEXA 2001)*, Germany; LNCS 2113, 232-242; 2001.

[8] Edmonds A., "A General Background to Supervised Learning in Combination with XML Technical paper, Scientio Inc.," http://www.metadatamining.com.

[9] [*Mapping Objects to Data Models with the UML. Rational Software White Paper. Rational Software Corporation, 2000. p. 12*". Available at http://www.rational.com/products/whitepapers/101576.jsp (March 2009).

[10] Object Management Group, Inc. :UML2.0 Superstructure. http://www.omg.org/cgi-bin/apps/doc ?formal /05-07-04.pdf (2005)

[11] Luj´n-Mora S., Trujillo J., and Song, I., "A UML profile for multidimensional modeling in data warehouses," *Data Knowl. Eng*. 59(3) 725–769, 2006.

[12] Vassiliadis P., and Sellis, T., "A Survey of Logical Models for OLAP" *Databases. SIGMOD Record* 28(4), 64–69, 1999.

[13] Abello A., Samos J., and Saltor F., "YAM2 (Yet Another Multidimensional Model):An Extension of UML," *IEEE Computer Society*, 172–181 2002.

[14] Nguyen T., Tjoa A., and Wagner R., *An Object Oriented Multidimensional Data Model for OLAP*, Springer-Verlag 69–82; 2000.

Kamal Alaskar has completed M.Sc(Electronics)(1991),MCM(1993), M.Phill(2006) MCA(2009) and submitted Ph.D.(Dec 2008). He has worked as a Lecture for 10 years in the faculty of Computer Applications and System Studies. Currently he is working as a Assistant Professor and Head of Computer Applications. Recently Government of India Ministry of External Affairs deputed him as a Professor for two years in Royal University of Butan's Geaddu College of Business studies. He has 6 books on his credit on "Information Technology" for XIth and XIIth Std. which is approved by Board of Higher Education Pune Governement of Maharashtra state India.He worked as a member of Board of studies and Board of Appointment for Bharati Vidyapeeth University Pune Maharashtra India and Women's Karanata University Bijapur Karnataka India.

Akhtar Shaikh has completed M.Sc.(1988), Ph.D.(1999). He has 20 years of teaching experience and 12 years of research experience. Currently he is working on a minor project funded by University Grants Commission (UGC) India. He wrote six books on "Information Technology" and having 10 research papers on his credit. Currently he is working as a Co-coordinator of BCA course. He worked as a referee of three research journals including International journal of Engineering and Material Science.