

## Abstract

Data warehouse is organized as collection of multidimensional cube, which represents data in the form of data values, called measures, associated with multiple dimensions and their multiple classification levels. In multidimensional cube, object oriented approach may be used for structuring the data. The focus of the paper is on operators defined for the object oriented model, required for formatting the query and results. In this paper object oriented multidimensional data model is defined for description of data. Seven operators for the object oriented model are introduced: intersection, difference, symmetric difference, restriction, union, join, projection. This paper extends the work of [7] which includes the four operators: restriction, union, join and projection. For every operator input, output, symbol and mathematical representation of the operators are defined.

*Keywords*: Data Warehousing, OODBMS, OLAP.

## 1. Introduction

Multidimensional (MD) modeling is the foundation of data warehouses (DW), MD databases, and On-Line Analytical Processing (OLAP) applications. These systems provide companies with many years of historical information for the decision making process. The analysis of this historical information is done by various types of tools which are based on OLAP technology. OLTP systems are not used because of imposing different requirement than OLAP system and therefore different data models are required for each type of system.

However, there is no formal multidimensional model that is commonly accepted for OLAP applications. Lots of work regarding multidimensional modeling for a data warehouse has been proposed in last few years. However most of them, focus either on an OLAP queries language such as [1], [2], and [3] or on some key issues such as aggregation, the additivity on fact attributes multiple classifications etc.

On the other hand, W.Lehner in [4] introduces an object oriented approximation for the design of multidimensional database in the proposed nested multidimensional data. This approach does not consider the key issues such as, the additivity on fact attributes along dimension, derived measures.

\*USIT, Guru Gobind Singh Indraprastha University, New Delhi, INDIA.

\*\*Maharaja Surajmal Institute of Technology, New Delhi, INDIA.

Juan Trujillo [5] introduces the GOLD model: An object oriented multidimensional data model for multidimensional databases, for the design of multidimensional database which discuss the issues such as derived measures, derived dimension attributes, the additivity on fact attributes along dimensions and the encapsulation of data with its operations in classes. In [6] author discuss various key issues such as derived measures, many-to-many relationship, additivity on fact attribute along dimension, classification hierarchies strictness and completeness. On the other hand Alexander [7] introduces, an object oriented model for a data warehouse. This model introduces the concept of multidimensional object to define the multidimensional cube on which a group of four operations are defined, in order to permit subsequent data analysis. This approach does not consider the whole database schema and as a consequence the derived measures and operations such as intersection, difference, union, and projection are not considered.

In our paper we discuss an object oriented multidimensional data model for a data warehouse over which we are defining group of operations like intersection, union, difference, projection to obtain a more expressive model for description of the data.

The remainder of this paper is structured as follows: Section 2 introduces the object-oriented multidimensional model for DW. Section 3 describes operations over this model. Section 4 contains paper summary.

### 2. Object Oriented Multidimensional Model

In object oriented multidimensional modeling, data is structured into facts and dimensions. A fact is an item of interest for any enterprise and is described through a set of attributes called measures or fact attributes which are contained in cells in the data cube. This set of measures is based on a set of dimensions that determine the granularity adopted for representing facts. On the other side, dimensions provide the context in which facts are to be analyzed. Moreover, dimensions are also characterized by attributes which are usually called dimension attributes.

Mathematically, object oriented multidimensional model denoted by F (D1, D2, D3.....Dn) is made of fact name and list of dimensions. Each dimension Di (A1, A2, A3.....An) is made up of dimension name D and a list of attributes A1.....An. Each attribute Ai is the name of role played by some domain D in the object schema. D is called the domain of Ai and is denoted by dom (Ai).

Let us introduces object oriented multidimensional model by an example of sale system of an organization .shown in Fig [1]. In this diagram the fact sales of products is analyzed along three dimensions (product, time and market).

Facts are considered as classes in shared aggregation relation (a special case of association in which the association relationship between the classes is considered as a "whole-part" and the parts may be parts in any whole) of n dimension class.

With respect to dimensions, each level of classification hierarchy is considered as a class. These classes must form a Directed

Acyclic Graph starting from each dimension class. A class2 (level2) of a hierarchy is considered as an association of a class1 (level1). Directed Acyclic Graph allow us to consider alternative path and multiple classification hierarchies shown in fig[1].The classification hierarchies such as strictness and completeness are also considered.

Dimensions can be categorized by Generalization/Specialization hierarchies. This we can observe in the diagram by the product dimension.

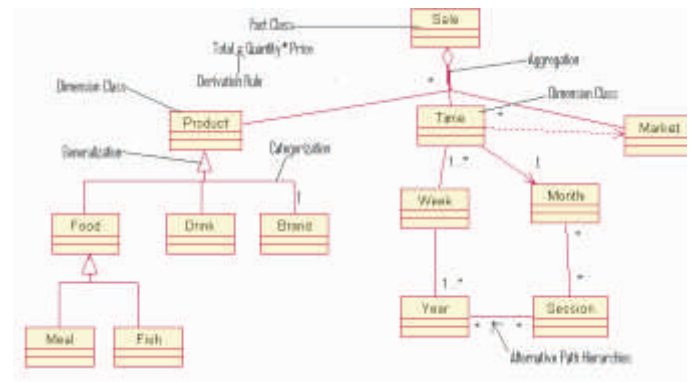


Fig 1 : OO Model of the Sale System

### 3. Operators

The operators are atomic operations on object oriented multidimensional model for a data warehouse. We can perform operations one after another to make more complex operator. Here in this paper we define seven basic operators necessary to make a query and format result: intersection, difference, exclusive OR, restriction, union, join and projection.

#### Intersection

The intersect operator takes set of object with one of class which are common with other classes. Note that this operator realizes common instances of a cube in multidimensional database. Intersection operator is denoted by the symbol '∩'. The intersection of class A and class B is denoted by "A ∩ B". Formally:

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

Or

$$A \cap B = \{x \in A \wedge x \in B\}$$

If the intersection of two classes A and B is empty, that means they have no object in common, then they are said to be disjoint.

#### Intersection of more than Two Classes

If A1, A2,..... An denote classes, then the intersection of these classes denoted by  $\bigcap_{i=1}^n A_i$  is defined as :  $\bigcap_{i=1}^n A_i = \{x : x \in A_i : \text{for every } i (i = 1, 2, \dots, n)\} = \{x : x \in A_i : \forall i, i = 1, 2, \dots, n\}$

**Input:** Set D1 of class c1 D1 = {O | O is instance of c1} and set D2 of class c2 D2 = {O | O is instance of c2}

**Output:** Set Dans contains the objects of new class that is intersection of c1 and c2 and data is common of objects. If no objects of D are common then Dans is an empty set.

**Mathematically:** Intersection  $(D1, D2, D3, \dots, Dn) = \text{Dans}$   
 Class of Dans is  $c1 \cap c2 \cap c3 \dots \cap cn$ . If  $\forall 01, 02, 03, \dots, On \setminus 01 \in D1, 02 \in D2, 03 \in D3, \dots, On \in Dn \Rightarrow (O1, O2, O3, \dots, On) \in \text{Dans}$ .  
**Symbol:**  $\cap$

#### Difference or Complement

If A and B are the class, then relative complement of B in A, is the set of all object in A which are not in B. This operator is denoted by "A-B".

$$\therefore A - B = \{x: x \in A \text{ and } x \notin B\}$$

**Input:** Set D1 of class c1  $D1 = \{O \mid O \text{ is instance of } c1\}$  and set D2 of class c2  $D2 = \{O \mid O \text{ is instance of } c2\}$

**Output:** Set Dans contains the objects of new class that is difference of c1 and c2 and data is difference of objects.

**Mathematically:** Difference  $(D1, D2) = \text{Dans}$   
 Class of Dans is  $c1 - c2$ . If  $\forall O1, O2 \mid O1 \in D1, O2 \in D2$  and difference  $(O1, O2) \Rightarrow (O1 - O2) \in \text{Dans}$ .

**Symbol:** '-'

#### Symmetric difference or exclusive or:

The symmetric difference of class A and class B is the set containing those objects which are in either A or B, but not in both A and B. In other words the symmetric difference of the class A and B is the difference of  $A \cup B$  in  $A \cap B$ .

The symmetric difference of A and B is denoted by  $A \oplus B$ .  
 Symbolically:

$$A \oplus B = \{x: x \in A \cup B, \text{ and } x \notin A \cap B\}$$

This operator we can derive from the other operator's i.e. union, intersection and complement operator by applying the following formula:

$$c1 \oplus c2 = (c1 \cup c2) - (c1 \cap c2)$$

**Input:** Set D1 of class c1  $D1 = \{O \mid O \text{ is instance of } c1\}$  and set D2 of class c2  $D2 = \{O \mid O \text{ is instance of } c2\}$

**Output:** Set Dans contains the objects of new class that is Symmetric Difference of c1 and c2 and data is  $\oplus$  of objects.

**Mathematically:** Symmetric Difference  $(D1, D2) = \text{Dans}$   
 Class of Dans is  $c1 \oplus c2$ . If  $\forall O1, O2 \mid O1 \in D1, O2 \in D2$  and symmetric difference  $(O1, O2) \Rightarrow (O1 \oplus O2) \in \text{Dans}$ .

**Symbol:**  $\oplus$

#### Union:

Let A and B be two class. The union of class A and B is the set that contain those object that are either in A or B or in both.

The union of the class A and B is denoted by  $A \cup B$  and is read as "A union B". Symbolically:  $A \cup B = \{x: x \in A \text{ or } x \in B\}$

Or

$$A \cup B = \{x: x \in A \vee x \in B\}$$

**Symbol:** ' $\cup$ '

Union of more than two classes: If  $A1, A2, \dots, An_n$  denote classes, then the union of these classes denoted by  $\bigcup_{i=1}^n A_i$  is defined as:  $\bigcup_{i=1}^n A_i = \{x: x \in A_i \text{ for at least one } i\}$ . For detail study of union operator please refer [7].

#### Restriction

The restrict operator operate with one of classes and remove the objects of this class which do not satisfy a state condition. Note that this operator realizes slicing/dicing of a cube in multidimensional database. In general, the select operation is denoted by  $\sigma_{\langle \text{condition} \rangle (\langle \text{class name} \rangle)}$ . In [7] Alexander Konovalov has elaborated input, output and mathematical representation of the restriction operation.

**Symbol:**  $\sigma$

#### Join:

The join operator takes set of objects with one class and makes join with another set of Objects. For input, output and mathematical representation of the Join operator, refer [7].

#### Projection:

The projection operator operates with one class. It returns values of selected attributes. If we are interested in only certain attributes of a class, we use the PROJECT operator to "project" the relation over these attributes. For detailing refer [7].

## 4. Conclusion and Future Scope

In this paper we have presented the object oriented multidimensional model for a data warehouse that allows us to accomplish major relevant MD properties at the conceptual level. Adding features like hierarchies, aggregation, generalization, association etc. Operations over the model: union, intersection, projection, restriction, joins, symmetric difference, differences are proposed.

We have based our approach in UML as it is widely accepted object-oriented modeling language, which saves developers from learning a new model and its corresponding notations for specific MD modeling. Secondly UML can be easily extended using stereotype, tagged value and constraint.

As a further future work, we will consider the proposing of group of features and group of more operations as a means to describe good OO multidimensional model for a data warehouse.

## 5. References

1. C. Li and X. Wang, "A Data Model for Supporting On-Line Analytical Processing". *Proc. in Intl. Conf. on Information and Knowledge Management, (CIKM'96), Rockville (Maryland) USA Nov. 1996, pp. 81-88.*
2. M. Gyssens and L. Lakshmanan "A Foundation for Multi-Dimensional Databases". *Proc. in the 33rd Intl. Conf. On Very Large Database Conference (VLDB'97), Athens, Greece, August, 1997, pp.106-115.*
3. R. Agrawal, A. Gupta and S. Sarawagi, "Modeling Multidimensional Databases". *Proc. 13th Intl. Conf. On Data Engineering, (ICDE'97), Birmingham, U.K., April 1997, pp. 232-243.*
4. W. Lehner "Modeling Large Scale OLAP Scenarios". *Lecture Notes in Computer Science, number 1377 in proc. of the 6th Intl. Conf. On Extending Database Technology, (EDBT'98), Valencia, Spain. March 1998, pp. 153-167.*

5. Juan Trujillo, "the GOLD definition language: An object oriented formal specification Language for multidimensional databases", 2000, pp346-350.
6. Juan Trujillo, Manuel Palomar, Jaime Gomez, "Designing data warehouses with OO conceptual models" IEEE 2001.
7. Alexander Konovalov, "Object-Oriented data model for data warehouse" Springer 2002.
8. W.H.Inmon, "Building the data warehouse" Third Edition, 2002.
9. Dori, D.Feldman, R.Sturm, "Transforming an operational system model to a data warehouse model: a survey of techniques" 2005 47-56
10. Sergio Lujan-Mora, J Trujillo, Yeol Song, A UML profile for multidimensional modeling in data warehouses, *Data and knowledge engineering Elsevier* 2006, 725-769
11. Alberto Abello, Jose Samos, Felix Saltor, "YAM2: a multidimensional conceptual model extending UML", *Information system, Elsevier* 2006, vol31, pp.541-567.
12. Deepti Mishra, Ali Yazici, Beril Pinar Basaran, "A case study of data models in data Warehousing, 2008, pp314-319
13. D. Kuznetsov, Yu. A. Kudryavtsev, "A mathematical model of the OLAP cubes" *journal of programming and software*, 2009 vol. 35, no-5, pp257-265