Constructing Data Cube as an Object Oriented Class

Antoaneta I. Ivanova[1]

Abstract **– The goal of this article is to describe Object Oriented Conceptual Model Data Cube using different kind of data – cancer register, experiments in toeing tank. Problems of data analyzing, saving and summarizing at various levels of detail and on various combinations of attributes are explained. Future tasks to develop appropriate methods to analyze and save consequence data are outlined.**

Keywords **– data warehouse, data cube, OLAP, object oriented classes**

I. Introduction

OLAP databases often need to summarize data at various levels of detail and on various combinations of attributes [3], [4]. Data warehouse, multidimensional database (MDB), and OLAP applications emphasize multidimensional modelling, which offers two benefits:

- closely parallels how data analyzers think and therefore, helps users understand data;
- helps to predict what the final users want to do, thereby facilitating performance improvements.

The objective of this work is:

- to present a method of structuring and description of different kind of data, providing a faster and efficient data analysis by user defined aggregation functions, as well as consequent storage of the processed data;
- to analyze outlined models of classes and demonstrate that offered model of dynamic object oriented class is universal and useful for completely different data.

II. Model of Data Cube

A. Related works

Trujillo, Palomar and Gomez [5] propose an OO approach to accomplish the conceptual modeling of data warehouses, MDB, and OLAP applications. They use UML to design Data Warehouses because it considers, at conceptual level, the structural and dynamic properties of the information system more naturally than do the classic approaches such as the Entity-Relationship model. Further, UML provides powerful mechanisms—such as the Object Constraint Language and the Object Query Language—for embedding Data Warehouse constraints and initial user requirements in the conceptual model. This approach to modeling a Data Warehouse system yields simple yet powerful extended UML class diagrams that represent main data warehouse properties at the conceptual level.

[1]Antoaneta I. Ivanova is with Department of Computer Science and Engineering, Technical University of Varna, Studentska 1 Str., 9010 Varna, Bulgaria, E-mail: antoaneta_ii@yahoo.com

Some recent papers [1] and [2] as a previous work, describe examples for use of a similar class without emphasis however, on the algorithms for construction of a data cube, the OLAP operations and methods about cancer register and toeing tests – completely different data.

B. Attributes of DATA CUBE

A data cube is constructed from a subset of attributes in the database. Certain attributes are chosen to be **measure attributes**, i.e., the attributes whose values are of interest. Other attributes are selected as dimensions or **functional attributes**. The measure attributes are aggregated according to the dimensions.

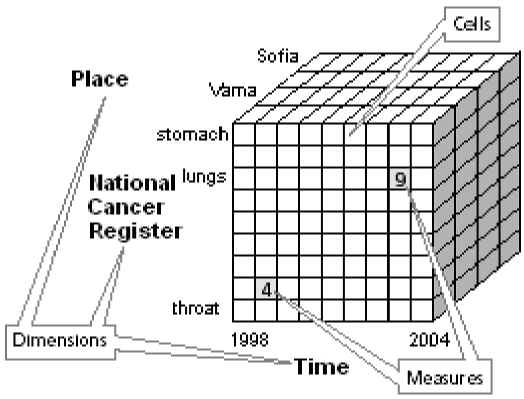

Fig. 1. Model data cube of Cancer Register

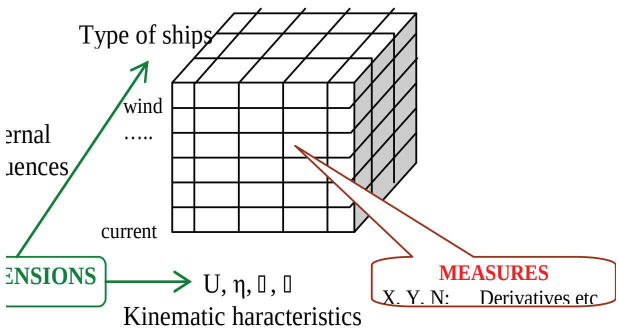

Fig. 2. Model dta cube of PMM ship tests

The measure is additive along a dimension. The User Defined aggregation operator is used to aggregate attribute values along all hierarchies defined on that dimension. The aggregation of some fact attributes—called roll-up in OLAP terminology— might not however, be semantically meaningful for all measures along all dimensions.

By selecting cells, planes, or subcubes from the base cuboid, we can compare and analyze the feature attributes which are stored in the data cube during data processing. In total, a d-dimensional base cube is associated with 2d cuboids.

B. Classification hierarchies

The definition of the **classification hierarchies** of certain dimension attributes is crucial since these classification hierarchies provide the basis for the subsequent data analysis. As a dimension attribute can be aggregated also to more than one other attribute, multiple classification hierarchies and alternative path hierarchies are also relevant.

Defining the classification hierarchies of certain dimension attributes is crucial because these classification hierarchies provide the basis for the subsequent data analysis. Because a dimension attribute can also be aggregated to more than one other attribute, multiple classification hierarchies and alternative path hierarchies are also relevant. For this reason, directed acyclic graphs provide a common way of representing and analyzing dimensions with their classification hierarchies.

The main of first model on Cancer Register is outlined on Fig.3 – National Cancer Register and Patient.

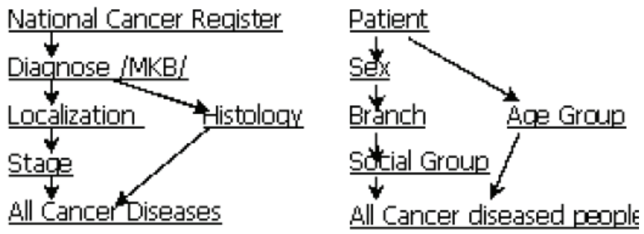

Fig. 3. The classification hierarchies with alternative path - National Cancer Register, Patient

In this model there are other classification hierarchies -***Time*** dimension- 1) Time–quarter–year–5 years 2) Time–season.

Place dimension - classification hierarchy without an alternative path have been defined:

Place–Town/Village–Municipality–Region–Country.

For the second example Fig 4 shows the different classification hierarchies defined for the type of ship dimension. On this dimension, a multiple classification hierarchy is defined so that data values can be aggregated by user defined function along different hierarchy paths.

Ship dimension

1)Ship–Types Depending on navigation region (see, river, etc.)

2) Ship–Hull

3) Ship – Control means – Number of control means – Type

4) Ship – Propulsion means - Number of propulsion means - Type

The data cube may store data, witch are derived from other tests. An alternative path classification hierarchy has been defined with two different paths that converge into the same hierarchy level, for the ***Type of Test*** dimension

1) Type of tests–Free running maneuvers

2) Type of tests–PMM

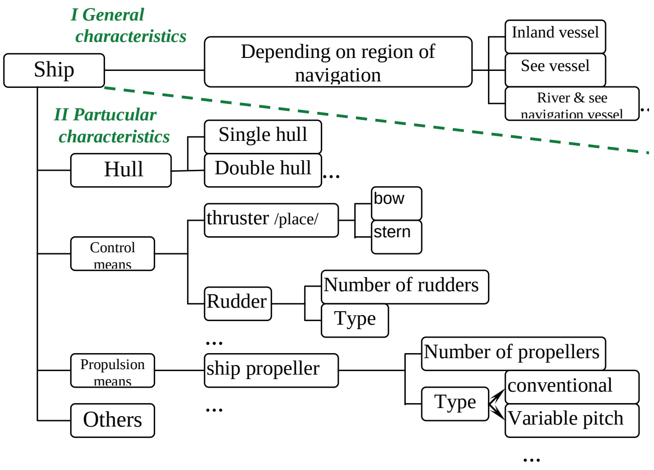

Fig. 4. *Type of ship* - the classification hierarchy display the dimension that defines the cube

In most cases, however, classification hierarchies are not so simple. The concepts of strictness and completeness are important for both conceptual purposes and for further multidimensional modeling design.

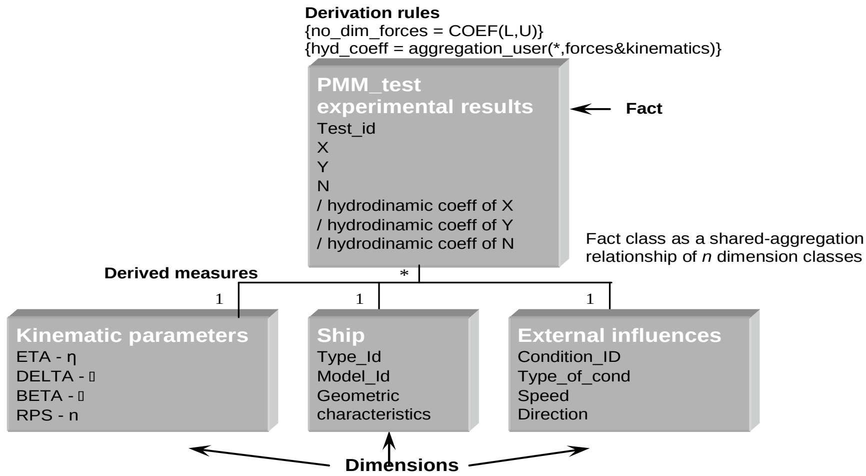

Fig. 5 The PMM_test class consists derivation rules, derived measures and has shared – aggregation relationship with the Kinematics parameters, Ship, Conditions

III. Object Oriented Conceptual Model - Class Definition

OO approach can elegantly represent multidimensional properties at both levels.

A. Structural level

This OO approach is not restricted to flat UML class diagrams in order to model large, complex data warehouse systems. UML's package grouping mechanism groups classes into higher-level units, creating different levels of abstraction and simplifying the final model. In this way, a UML class diagram improves and simplifies the system specifications, created with classic semantic data models such as the Entity-Relationship model. This approach clearly separates the structure of a multidimensional model specified with a UML class diagram into facts and dimensions. Fig 5 shows the example on PMM tests.

A. Dynamic Level

These cube classes are used to represent initial user requirements as the starting point for the subsequent data-analysis phase.

The basic components of the cube classes include the:

- head area, which contains the cube class's name;
- measures area, which contains the measures to be analyzed;
- slice area, which contains the constraints to be satisfied;
- dice area, which contains the dimensions and their grouping conditions to address the analysis; and
- cube operations, which cover the OLAP operations for a further data-analysis phase.

B. Examples

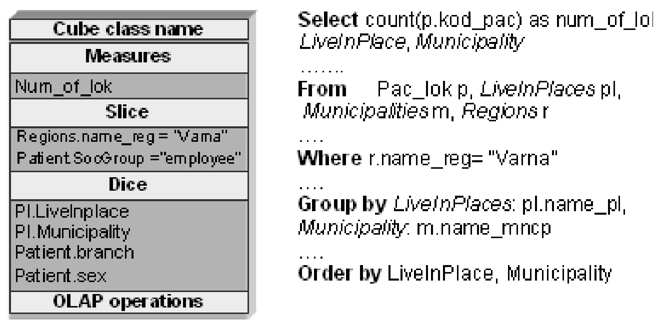

Fig. 6. Cube class example with parameters specified in the measures, slice, dice, and operations areas, and the class's corresponding Object-Query Language specification on model of Cancer Register

In the first model about Cancer Register can add additional methods which calculate for example: 1) disease by localization, location (place), Age group 2) , 2) death-rate on year or 5 year, for different kind of social group.

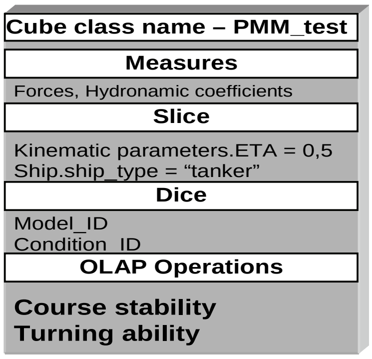

Fig. 7 Cube class definition on PMM tests

Into the second model is represented an extension of created class about OLAP operations – calculate course stability and turning ability on different models.

On the basis of aggregation function over data we can calculate and analyze new data which save on Database.

IV. CONCLUSION

OO Conceptual models of data cube are a very interesting direction for constructing and using efficiently information extracted from data cube.

OLAP tools implement a multidimensional model from two different levels:

• Structural—the structures that form the database schema and the underlying multidimensional model—also known as the metadata—that provides the model’s key semantics (facts, measures, dimensions).

• Dynamic—refers to the definition of final user requirements—also known as method and OLAP operations for further analyzing data.

The offered model of dynamic object oriented class is universal and useful for completely different data.

For a further work will be to develop additional properties or methods in class definition for analyzing data and saving results.

ACKNOWLEDGEMENT

The work presented in this paper was supported within the project BG 051PO001-3.3.04/13 of the HR Development OP of the European Social Fund 2007-2013.

REFERENCES

[1] Ivanova A., V.Chotukova, “A model of Object Oriented Class for Extraction and Use of Data Cube Unit Information”, Proceedings of CompSysTech’06, Rousse (Bulgaria), 2006.

[2] Ivanova A., “Constructing Object Oriented Class for extracting and using data from data cube”, Proceedings of CompSysTech’05, Varna (Bulgaria), 2005.

[3] Lewis P., “Database and Transaction Processing: An Application – Oriented Approach”, Addison Wesley, New York, 2003

[4] Molina.H, J. Ullman, J. Windom, “DataBase Systems: The Complete Book”, Prentice Hall, 2002, 1119p.

[5] Trujillo J., M. Palomar, J. Gomez, Y. Song, “Designing Data Warehouses with OO Conceptual Models”, Computer, 66-75, December 2001, IEEE