A model of Object Oriented Class for Extraction and Use of Data Cube Unit Information

Antoaneta Ivanova, Valya Chotukova

Abstract: *An Object Oriented Conceptual Model Data Cube is described, on the basis of an example for storage, extraction and analysis of data derived from tests of ship models, conducted in the Bulgarian Ship Hydrodynamics Center - Varna. The main storage parameters as well as the structure of the system are presented. The dynamical level in an object oriented class definition of data cube is described. Future problems of OLAP operations and data analysis methods, included in the UML class definition are outlined.*

Key words: *data warehouse, data cube, MOLAP, object oriented classes, model ship experiments in toeing tank.*

INTRODUCTION

Figure 1 presents the main data storage levels: object data, experimental values, results from primary and secondary processing of the experimental data. The ultimate objective of the experiment is derivation of the hydrodynamic coefficients (derivatives) taking part in the equations of motion of the ship. The main purpose of the system, which stores data in each stage, is to provide analysis and comparison of the characteristics depending on the kinematic features and conditions of the experiment. Since the variety of the ship models is large and the multitude of respective characteristics and data depends on the results of previous levels, the Data Warehouse Concept can be applied. In accordance with DataBase Theory [7] such a system can be classified as OLAP (On-Line Analytical Processing) database.

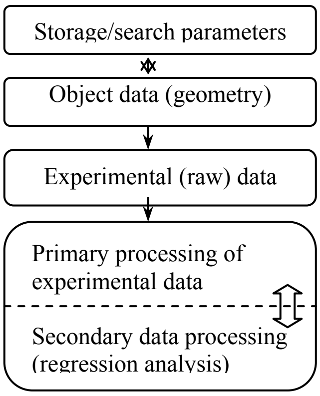

Figure 1. General structure of the data storage system

Aggregation is a predominant operation in decision support database systems. OLAP databases often need to summarize data at various levels of detail and on various combinations of attributes.

Most developers agree that data warehouse, multidimensional database (MDB), and OLAP applications emphasize multidimensional modeling, 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.

Developers have proposed various approaches for the conceptual design of multidimensional systems. These proposals try to represent the main multidimensional properties at the conceptual level with special emphasis on data structures – for example as Object Oriented (OO) structures.

The objective of this work is to present a method of structuring and description of experimental data, providing a faster and efficient data analysis by user defined aggregation functions, as well as consequent storage of the processed data.

RELATED WORKS

Trujillo, Palomar and Gomez [8] propose an OO approach to accomplish the conceptual modeling of data warehouses, MDB, and OLAP applications. This approach introduces a set of minimal constraints and extensions to UML for representing multidimensional modeling properties for these applications. They base these extensions on the standard mechanisms that UML provides for adapting itself to a specific method or model, such as constraints and tagged values.

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.

Some recent papers [1] 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.

SYSTEM STRUCTURE

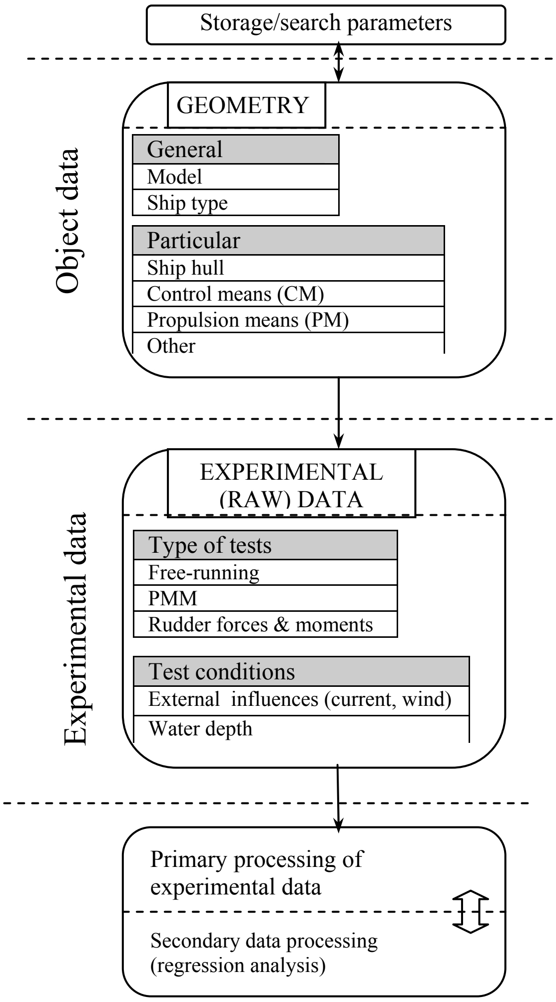

Figure 2.

The main data stored in the data base are divided in three levels (Figure 2):

- ✓ Object data – describe the geometric characteristics of the object – ship hull, rudder, bow thruster, engine, shafts, propellers, etc.;
- ✓ Experimental raw data, intermediary results of primary and secondary data processing;
- ✓ Hydrodynamic coefficients (derivatives) of the mathematical model of the ship, obtained as a final result of the data processing.

Various simulation analyses of ship performance (e.g. course stability, turning ability, etc.).

The main ship model tests are: Free-running (Circle, Zig-Zag, Pull-Out, etc.), PMM tests, Rudder Forces & Moments and others.

The values and characteristics related to the PMM tests represent a small part of the whole data base [9].

Experimental Data

The main objective (Figure 3) is to derive the forces in longitudinal **X** and transverse **Y** directions, as well as the moment around the model center of gravity **N**, on the basis of measured forces X, Y for different values of the kinematic characteristics – speed **U**, drift angle **β**, rudder angle **δ**, ship propeller revolutions **n.**

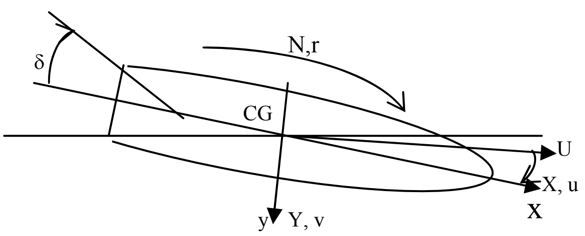

Figure 3.

$$\eta = \frac{u.n_c}{U_0.n}, \qquad .$$

where **u** – ship speed, $\mathbf{U_0}$ – initial ship speed, $\mathbf{n_c}$ **–** steady propeller revolutions, **n** – varying propeller revolutions

$$X, Y, N = f\left(\beta, \delta, \eta\right)$$

This actually covers a large part of the "Static" PMM tests.

The various cycles (stages) of the experiment are arranged depending on the variables used:

- ✓ *Calibration Cycle* – an initial but very important cycle, which data are used to determine the calibration coefficients for X and Y . These coefficients are applied in the primary processing to convert the measured electrical quantities [mV] into physical values – forces [kg];
- ✓ Cycle A – *X,Y,N=f(η);*
- ✓ Cycle B – *X,Y,N=f(β, η);*
- ✓ Cycle C – *X,Y,N=f(δ, η);*
- ✓ Cycle D – *X,Y,N=f(δ , β, η)*

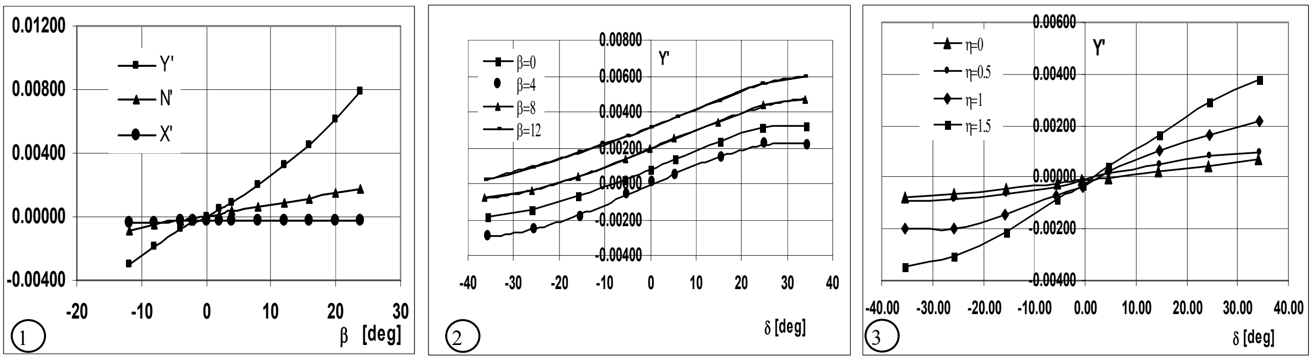

Figure 4 – 1) Static test with varying drift angle **β**; **2)** Static test with varing rudder angle **δ** and drift angle **β**
3) Static test with varing rudder angle **δ** and propulsion ratio **η**

Primary Data Processing

The processing comprises the following stages:

- ✓ Derivation of *X, Y, N* forces as physical values [kg], with the use of the calibration coefficients.
- ✓ Reduction of X, Y, N in a dimensionless form by the use of *COEF* function, reflecting ship geometry (L- ship length) and ship kinematic particulars (U – ship speed). $COEF = f(L, U)$.

  This operation provides dimensionless characteristics, suitable for comparison with various ships.

Secondary Data Processing

- ✓ Approximation of the obtained relations *X,Y,N* in the form $F = a_i p_1^n \quad n \le 4$
- ✓ Regression analysis or a secondary approximation of the derived coefficients.
- ✓ Derivation of the hydrodynamic derivatives of the mathematical model of the ship participating of the equation of motion:

$$\left| \begin{array}{l} X' = f(\text{Hydrodynamic coefficients in equation X}, \eta, \text{u}, \text{v}, \beta, \delta, \text{r}) \\ Y' = f(\text{Hydrodynamic coefficients in equation Y}, \text{u}, \text{v}, \beta, \delta, \text{r}) \\ N' = f(\text{Hydrodynamic coefficients in equation N}, \text{u}, \text{v}, \beta, \delta, \text{r}) \end{array} \right.$$

MULTIDIMENSIONAL MODELLING PROPERTIES

A *data cube* [2],[6],[7] 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.

Figure 5 below depicts a practical data cube example for PMM ship tests and considers a hypothetical database. This particular data cube has three feature attributes – *Kinematic characteristics , Type of ship and ship geometry* and *External influences* and a multiple measure attributes –– *Forces and moments, Nondimentional data and Hydrodynamic coefficients (derivatives).*

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 2^d cuboids. Each cuboid represents a unique view of the data at a given level of granularity. Not all of these cuboids need actually to be present, since any cuboid can be computed by aggregating across one or more dimensions in the base cuboid. Figure 6 shows the different classification hierarchies defined for the *Type of ships* dimension.

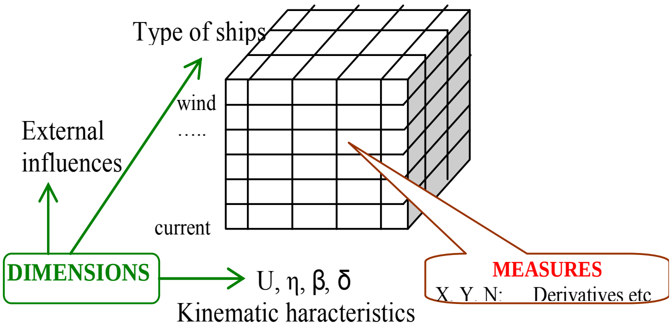

Figure 5. A multidimensional model data cube: - The cube is composed of cells that define fact attributes Kinematic characteristics , Type of ship and ship geometry, and External influences

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.

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. For this reason, directed acyclic graphs provide a common way of representing and analyzing dimensions with their classification hierarchies.

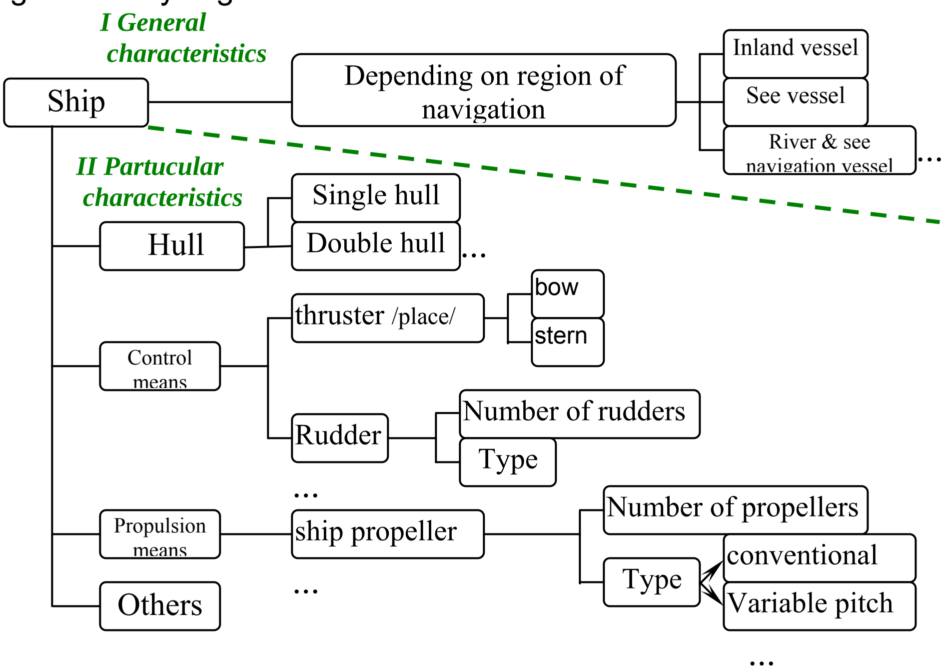

Figure 6. *Type of ship* - the classification hierarchy display the dimension that defines the cube

Figure 6 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–Types Depending on navigation region (see, river, etc.)*
- ✓ *Ship–Hull*
- ✓ *Ship – Control means – Number of contol means - Type*

- ✓ *Ship – Propultion means - Number of propultion 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:

- ✓ *Type of tests–Free running manoeuvers*
- ✓ *Type of tests–PMM*

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

Once developers define the multidimensional model structure, users can define a set of initial requirements such as a starting point for the subsequent data-analysis phase, etc. From these initial requirements, users can apply a set of OLAP operations to the multidimensional view of data for further data analysis. These OLAP operations usually include the following:

- *roll-up,* which increases the level of aggregation along one or more classification hierarchies;
- *drill-down,* which decreases the level of aggregation along one or more classification hierarchies;
- *slice-dice,* which selects and projects the data;
- *pivoting,* which reorients the multidimensional data view to allow exchanging dimensions symmetrically for facts.

OO CONCEPTUAL MODEL APPROACH

OO approach can elegantly represent multidimensional properties at both levels:

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.

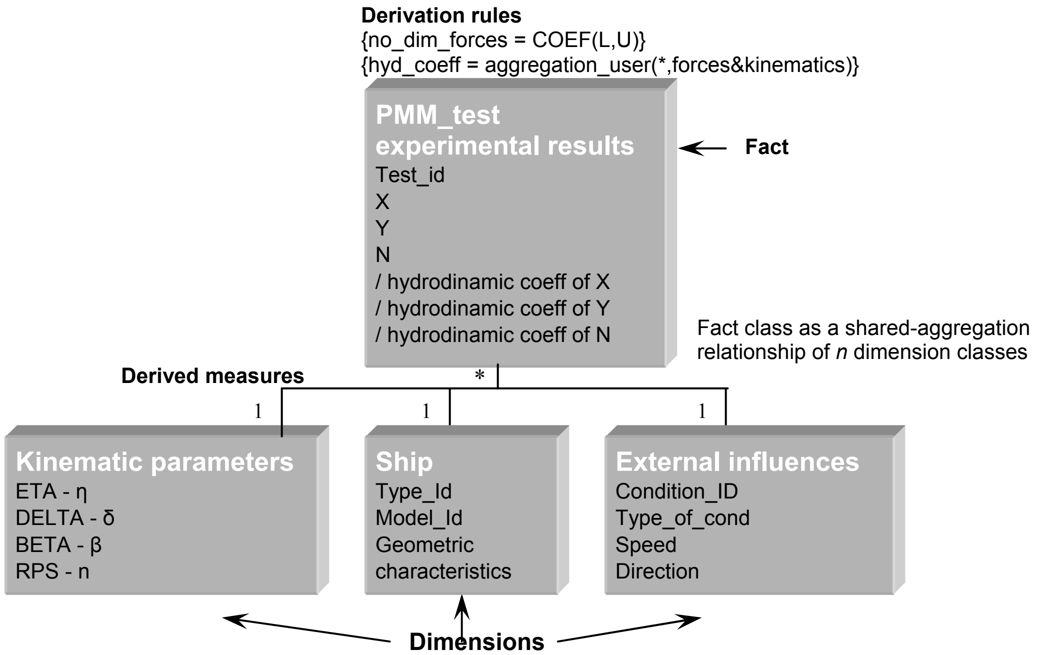

Figure 7. The *PMM_test* class consists *derivation rules*, *derived measures* and has shared – aggregation relationship with the *Kinematics parameters*, *Ship*, *Conditions*

Facts and dimensions - Fact classes represent facts and measures – X,Y,N dimensional or nondimensional, defined as attributes within these classes. Dimension classes represent dimensions.

Figure 7 shows the *PMM_test* fact class — consisting of experimental data — and the dimension classes *Kinematic parameters*, *Ship* and *External Influences*. The fact class is thus specified as a shared-aggregation relationship between all dimension classes.

Derived measures are considered by placing the constraint /next to a measure in the fact class/, for example - *no_dim_forces, hyd_coeff.* Derivation rules are presented between braces in Figure 7. This are user defined functions or custom aggregation methods.

Classification hierarchies – A base class represents each classification hierarchy level, for the dimensions. An association of classes specifies the relationships between two levels of a classification hierarchy. The structure can represent both *alternative path* and *multiple classification hierarchies*.

Dynamic level

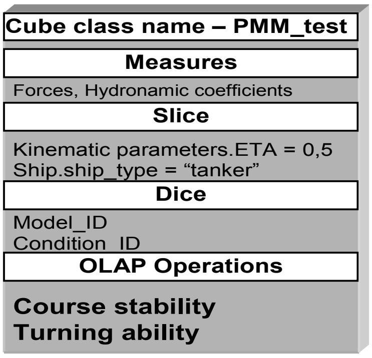

Figure 8. Cube class definition

These *cube classes* are used to represent initial user requirements as the starting point for the subsequent data-analysis phase. (Figure 8)

The basic components of the cube classes include:

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

CONCLUSIONS AND FUTURE WORK

OO Conceptual models of data cube are a very interesting direction for constructing and efficient use of information, extracted from data cube.

OLAP tools implement a multidimensional model from two different levels:

• *Structural*—the structures that form the database scheme and the underlying multidimensional model - facts, measures, dimensions.

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

Further work:

- to examine iwhether this class is appropriate for other type of data – for example data describing properties of geographical map and motion of the ship on the map;

- to add additional properties or methods in a class definition for further analyzing data on the basis of experimental data.

REFERENCES

[1] Ivanova A. Constructing Object Oriented Class for extracting and using data from data cube, *In Proceedings of CompSysTech'05***,** Varna (Bulgaria), 2005.

[2] Agarwal S., R. Agrawal, P. Deshpande, On the Computation of Multidimensional Aggregates, *In Proceedings of the .22nd international Conference on Very Large Databases,* 506-521**,** Mumbai (Bombay), 1996.

[3] Geffner S., D. Agrawal, A. El Abbadi, T. Smith. Relative Prefix Sums: An Efficient Approach for Querying Dynamic OLAP Data Cubes, *In Proc. of the 15th International Conference on Data Engineering*, Sydney, Australia, March 1999.

[4] Gray J.,A. Bosworth, A. Layman, H. Pirahesh. Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-total, *In Proceedings of the 12Th. IEEE International Conference on Data Engineering,* 152–159, New Orleans,LA, February - March 1996

[5] Kotidis Y., N. Roussopoulos, An alternative storage organization for ROLAP aggregate views based on cubetrees, *ACM SIGMOD Record* 1998*,* 27(2): 248-258

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

[7] Molina G., J. Ullman, J. Windom, DataBase Systems: The Complete Book, Prentice Hall, 2002, 1119p.

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

[9] Goodman A., M. Gertler, R. Kohl, Experimental Techniques and Methods of Analysis used at hydronautics for surface-ship maneuvering predictions, *Eleventh ONR Symposium on Naval Nydrodynamics*, London, June 1976

ABOUT THE AUTHOR

Antoaneta Ivanova, PhD Student, Department of Computer Science and Engineering, TU-Varna, E-mail:antoaneta_ii@yahoo.com

Valya Chotukova, M.Sc., Research Scientist, Bulgarian Ship Hydrodynamics Centre, E-mail:v.chotukova@bshc.bg