

# Development of multidimensional model of data for information and analytical decision-making support system

**R Uskenbayeva, B Kurmangaliyeva, N Mukazhanov\***

*Department of Computer Science and Software Engineering, International University of Information Technologies, 34 «A»/8 «A» Manas Str./Zhandosov Str., Almaty, Kazakhstan*

*Received 1 February 2014, www.tsi.lv*

---

## Abstract

Today almost in all spheres of industry it is observed the urgent need of information technologies development and increase of using opportunities. This paper is devoted to creation of multidimensional model of data for information and analytical system of decision-making support for management. The information and analytical system is developed by means of OLAP technology (On-Line Analytical Processing) which in turn analyses collected materials at a real time based on strategic level of solution. In information systems this technology displays data in the form of a multidimensional cube, and necessary data for users will be expressed in the form of cube slices. The multidimensional model of data is developed for information system intended for decision-making support of higher education institutions management.

*Keywords:* Multidimensional model of data, data warehouse, hypercube, dimension, higher education

---

## 1 Introduction

Today for higher education institutes there is a very small amount of decision-making support systems. Recently introduced education systems and forms require a new mode of management. Higher education quality management and control is one of the main goals of HEIs. In many cases HEIs management make decisions on base of old management systems, their own practice or international practice. The information and analytical system helps university management in decision making on base of real data.

Education quality control is identified by means of quantitative and qualitative indication measures. What does the indication of education quality mean? How it can be controlled? Among academic community there are various kinds of opinions on this issue. For instance, at educational institutes quality of education is evaluated on base of students' academic progress of their quantitative and qualitative indicators, while quality of graduates' skills can be assessed by labour market demand. Employers evaluate quality of higher education institutes on base of graduate's efficiency and their contribution into a company development.

Quantitative and qualitative indicators of HEIs are as follows: faculty staff, research activity, international cooperation, alumni and their achievements, financing, national and international rankings and others.

In order to control at a real time period of higher education quality on base of modern informational technologies it is necessary to gather the previously mentioned indicators and to organize work on this big

data; to prepare and to process data for analysis; to make new methods of data modelling and analysis; big data processing with constant updating and information improving for its multidimensional analysis; on base of obtained analysis it is necessary to set up the information and analytical system for HEIs development and control of decision making data.

There are a variety of informational systems used at HEIs of our country. These systems are intended for students registering and control of educational process. Where such components like HEIs' activities data analysis, prognostic analysis and process control at a real time are not considered at nowadays existing analytical systems. Management of the decision-making support systems is based on analytical data handling and data model is aimed to use the collected data as an educational resource. The explicit measure indicators, analysis, methods and conducted monitoring used in the system must contain the whole process (working procedure) of the educational organization, and derived results must be continually analysed and compared with the targeted objectives approved by the educational institute. On its basis should be conducted error corrections and can be defined in advance HEIs' future activities. Use of new informational technologies and methods in the system conducting monitoring for the management of institutions of higher education contributes to the effective management. The OLAP technology is used to implement the system. This technology is designed for big data of indicators analysis. The main advantage of it is its capability of data analysis at a real time and quick execution of multidimensional inquiries. The OLAP

---

\* Corresponding author - e-mail: [mukazhanovn@gmail.com](mailto:mukazhanovn@gmail.com)

technology is used in many information and analytical systems and is embed as particular service in extensively used management systems of database. As the OLAP technology works with multidimensional data warehouse, appears a task of multidimensional model set up [1, 2].

**2 The multidimensional data model**

The theory of acquisition is used in the multidimensional data model and presentation of aggregation indicators. On the basis of OLAP technology that presents the information in a type of cube lays the idea of the multidimensional data model. Logic of human is many-sided. When a person asks a question he sets a certain limitation, so he changes one question in many dimensions, that's why multidimensional mode of analytical process is very similar to human mind's versatility. We can say that the multidimensional data model discussed in this research work is appeared from versatility of human logic. Creation of the multidimensional data model intended for analytical system presented in this paper is composed of the following steps:

- To analyse the object sphere;
- To determine measures;
- To determine the quantitative indicators;
- To create logical model of multidimensional database. In this step we get relational schema of data warehouse in type of "snowflake" by connecting the table of facts and measure tables through logical connective [2].

Measure dimensions in the multidimensional data model are given as cube edge axis. Multidimensional cube is a multidimensional structure; multidimensional structure is a database that is closely connected with each other. Concerned multidimensional data model includes such basic concepts as hypercube, dimensions, measure indicators (members), cells, quantitative measures (indicators).

Cube consists of several dimensions, and every dimension is composed of measure indicators that form on edge of cube (figure 2 and figure 3). Cells can be composed of quantitative indicators or can be empty. Inspection of data saved in cube is conducted as a slice set composed of selected measures and quantitative measures (indicators).

Dimensions – a set of one-type or a polytypic object that defines a content of quantitative indicators. It can be simple one-dimensional or multileveled hierarchic structural by organization structure.

Measure indicators (members) – a table row or a table column consisted of indicators summary. The summary of inner objects is measure components (members). For instance, the components of an academic degree: baccalaureate, magistrate, Ph.D. The measure components can be various typed as: qualitative indicators of students, indicators of research works, time indicators: academic year, term, and week. The

components of every measure must be organized structural correctly and conveniently for expert. If measure indicators are organized by hierarchic structure, one typed objects will be placed on the same level. It is accepted to indicate measure indicators in a type of multidimensional cube's edge (axis).

Quantitative measure (fact) – a size in a quantitative type and the main component of analysis. There can be one or several quantitative indicators in an OLAP cube. For example, numerical strength of students and professional and teaching staff, finance and the quantitative measures of research works.

Cells – a cube structure that conforms to exact value of measures. During demonstration, cells are placed inside of cube and it is accepted to indicate quantitative measure value according to these cells.

We take the following marks through mentioned detection in accord with the theory of acquisition:

$D$  – a set of dimensions,  $d_i$  - element of a set  $D$  that is  $D = \{d_1, d_2, \dots, d_n\}$ .

$n$  – a number of dimensions;

$M_{d_i} = \{m_{1_i}, m_{2_i}, \dots, m_{k_i}\}$  – a set of measure indicators (dimensions' elements) or subset of a dimensions,  $k$  – a number of derived inner elements set,  $i = 1 \dots n$  an amount of marks in dimension.

$M$  – an amount of cube measure indicators (dimensions' elements):

$$M = M_{d_1} \cup M_{d_2} \cup \dots \cup M_{d_n} \quad [4, 5, 6]$$

*Definition 1:* A set  $M$  is a subset of a set  $D$  if every element of  $M$  is also an element of  $D$ :

$$M \subset D := m_{1_i} \in M \Rightarrow m_{1_i} \in D, \quad (1)$$

In this case,  $M$  is a subset of  $D$ ,  $D$ - $M$  superset. If  $M \subset D$  and  $M \neq D$ , we call  $M$  – a proper subset of  $D$  (figure 1). Any other subset of  $D$  contains at least one element of  $D$ , but not all of its elements. Note that the empty set is a subset of every set.  $\emptyset \subset M$  for every set  $M$  [16].

*Definition 2:* Two sets are identical if and only if they have exactly the same members. So  $M = D$  if for every  $m_{1_i}, m_{1_i} \in M \Leftrightarrow m_{1_i} \in D$ .

$$M = D := M \subseteq M \ \& \ D \subseteq M, \quad (2)$$

In any sense of having a problem are usually considered a subset of "most" of the set  $U$ , which is called the universal set. In our problems, the universal set is a set of cubes. For example, Figure 1 shows a universal set  $U$  and two subsets - of  $M$  and  $D$ ,  $M \subset D$ . Figure 1 is called the Euler-Venn's diagram [16].

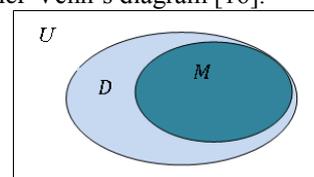


FIGURE 1 Euler-Venn's diagram

If we consider a cube as a set of cells, then hypercube –  $H(D, M)$ ,  $D$  and  $M$  – accordingly they are inner summaries of cube. If we mark cube cell as  $h$ , then  $h \in H$ , and measure indicators also will be given as cells  $M_h$ , hereby:  $M_h \in M$  [5].

As mentioned above, user can see the data as a cut set composed of several measures. If we mark a summary of measures selected for cut set as  $D$ , if we mark a summary of measure indicators relating to selected measures as  $M'$ , then a cut set of cube will be indicated as  $H'(D, M')$ . Dimensions of cut set –  $d_i \in D'$  and measures  $m_{k_i} \in M'$ . Slice set dimensions are formed by consolidation of measure summary  $M_{d_1} \cup M_{d_2} = M'$ , and quantitative measures are formed by intersection of measure indicators' summary (an aggregative value given in an intersection axis of dimensions):  $c_i = \{c_{1,1}\{m_{1_1}, m_{1_2}\}, c_{1,2}\{m_{1_1}, m_{2_2}\}, \dots, c_{k_1,1}\{m_{k_1}, m_{1_2}\}, c_{k_1,2}\{m_{k_1}, m_{2_2}\}, \dots, c_{k_1-1,k_2-1}\{m_{k_1-1}, m_{k_2-1}\}, c_{k_1,k_2}\{m_{k_1}, m_{k_2}\}\}$

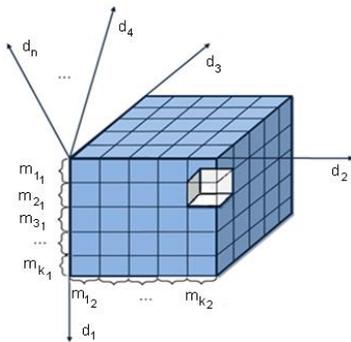


FIGURE 2 Multidimensional data model given as a cube

If a cube consists of  $n$  - dimensions, then it will be able to determine a quantity of all possible two-dimensional slice sets by the following formula:

$$N_s = \frac{n^2 - n}{2}, \tag{3}$$

Here,  $N_s$  - a quantity of two-dimensional slice sets outgoing from cube;  
 $n$  - a number of dimensions.

A set of quantitative measures is marked as  $x_{k_i}$ , quantitative measures of dimensions  $x_{k_i}$  ( $i=1..n, k=1..|d_i|$ ) - values of measures are given as aggregative summand. Usually aggregative summand is determined by horizontal and vertical calculating of quantitative measures situated in intersection axis of dimension inner summaries (measures).

Bringing higher educational institutions' measures through cube dimensions to the marks derived by the theory of acquisition:

Set of dimensions:  $D = \{\text{academic years, levels of educational system, specialties, education forms,}$

$\text{indicators of students, international students, student clubs, graduates, professional and teaching staff, academic degree, academic title, research activities, scientific infrastructure, international relations, selection committee, prizes for undergraduate students, financial and educational programs, quality evaluation, accreditation, international rating, accreditation organizations, infrastructure, sport buildings, institutions, departments}\}$ ;

Set of dimensions' inner measures:  $M_{d_i} = \{\{2009-2010, 2010-2011, 2011-2012, 2012-2013, 2013-2014, \dots\}, \{\text{Total amount, amount of governmental grants, ..., amount of international students}\}, \{\text{Computer science, Information systems, Software engineering, ...}\}, \dots\}$

Set of dimensions inner measures:  
 $m_{1_1, \dots, k_1} = \{2009-2010, 2010-2011, 2011-2012, 2012-2013, 2013-2014, \dots\}$ ;

$m_{1_2, \dots, k_2} = \{\text{Total amount, amount of governmental grants, ..., amount of international students}\}$ ;

$m_{1_3, \dots, k_3} = \{\text{Computer science, Information systems, Software engineering, ...}\}$

...  
 $m_{1_4, \dots, k_4} = \{\dots, \dots, \dots\}$

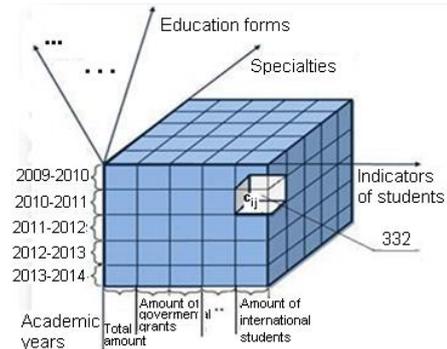


FIGURE 3 Higher educational institutions' dimension measures are given in the shape of a multidimensional cube

Hierarchical structure. Hierarchic structure is embedding to data model. It allows passing to hierarchic level at any time.  $L$  – if we will consider  $d_1$  dimension composed of hierarchic levels, e.g.  $d_1$  – time dimension, an academic year – term – a week. In the hierarchic structure, initial values are situated in lower level (FIGURE 4).

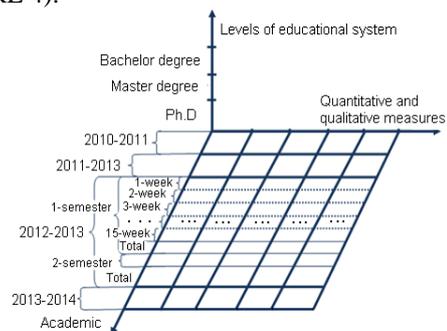


FIGURE 4 Hierarchic structure of an academic year

In this structure quantitative measures are taken from the lowest level and its' value is passed on higher levels. In order to conduct quickly the process of passing on hierarchic levels, calculating values of various levels must be saved as multidimensional model.

**3 Data warehousing logical design**

There is a proper order of creation and connection tables of data warehouse. There two types of tables: fact table and dimensions table. All the information of informational sphere components in resolving this task must be included in a logical data model creation process. They are intended to create entity-relationship logic database model, also they provide with presentation of data structure in defining the objects. The main motive of

selection of this notation is that it is independent from database physical creation projection and used informational platform. In the paper takes place information about information and analytical system data warehouse model in type of "snowflake" schema created by the project "Higher education quality control and corporative informational system of management" (model of data warehouse is presented in the figure 5). In computing, a snowflake schema is a logical arrangement of tables in a multidimensional database such that the entity relationship diagram resembles a snowflake in shape. The snowflake schema is represented by centralized fact tables, which are connected to multiple dimensions. "Snow flaking" is a method of normalizing the dimension tables in a star schema [15].

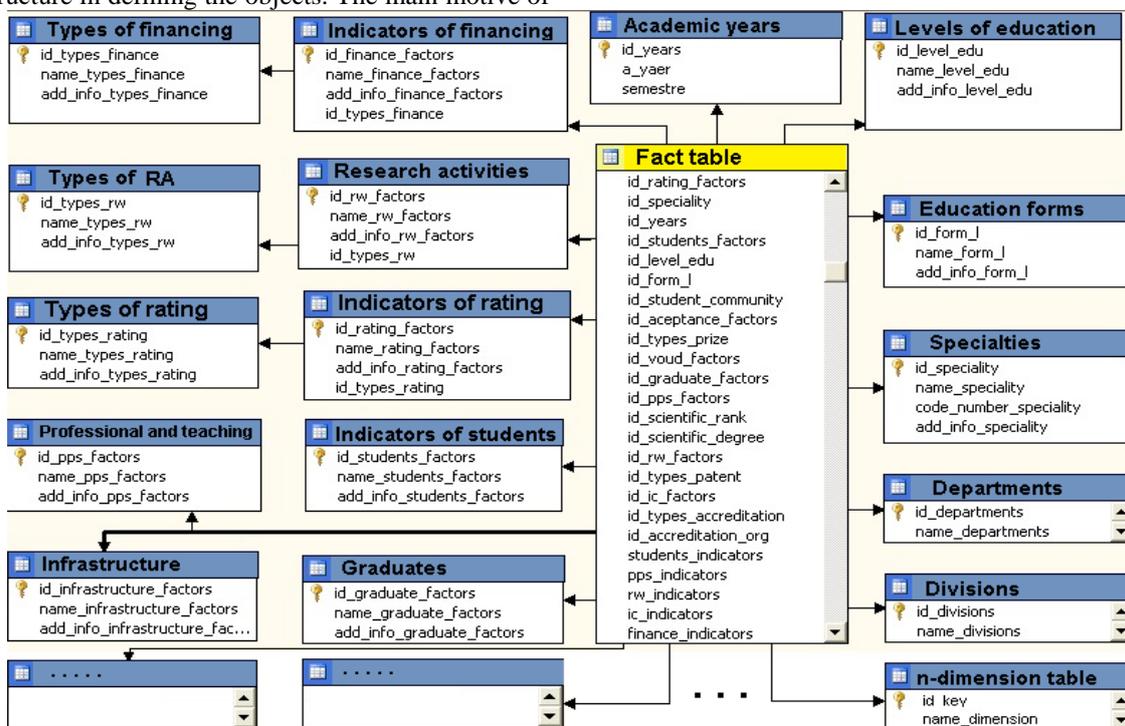


FIGURE 5 Relational schema of data warehouse in type of "snowflake"

When it is completely normalized along all the dimension tables, the resultant structure resembles a snowflake with the fact table in the middle. The principle behind snow flaking is normalization of the dimension tables by removing low cardinality attributes and forming separate tables.

Multidimensionality of an information and analytical system must be in the following three levels:

1. Multidimensional data presentation - the means of the end user providing multidimensional visualization and a manipulation with data;
2. Multidimensional processing – means of a formulation of multidimensional inquiries and the processor, able to process and execute such inquiry;

3. Multidimensional storage – the means of a physical data structure providing effective implementation of multidimensional inquiries.

The first two levels without fail are present in the system means. The third level, though is widespread, is not obligatory as data for multidimensional representation can be taken and from usual relational structures.

**4 Conclusions**

This article provided with the following information:

- The necessity of information and analytical system in decision making support in higher educational institutions;
- Basic definitions of multidimensional data model of one of the most important components of

information and analytical systems – hypercube structure are given;

- The theory of acquisition was used in defining the hypercube structure;
- Information and analytical system's data model meant for higher educational institution's management was prepared on the basis of hypercube.

The analytical system for higher educational institutions' management was created due to the possibilities of OLAP technologies. OLAP technologies are means intended to data handling in a certain time. The modern OLAP servers have the possibilities of quick handling of solid information taken from various sources; presenting cut sets, different diagrams and other forms of information to users. We think that for these valuable reasons the OLAP technologies are very useful for heads of higher educational institutions in decision making on basis of gathered data.

## References

- [1] Uskenbayeva R K, Mukazhanov N K 2013 *VESTNIK of KazNTU*, 4(98) Almaty, Kazakhstan 106
- [2] Ermolkevich A A, Nezhurina M I 2010 Postroenie modeli informatsionno-analiticheskoi sistemy monitoring obrazovatel'nogo protsessa prinalichiy biznes-zakazchika *Biznes-informatika* 2(12) (in Russian)
- [3] Kulagin V P, Matchin V T 2010 Matematicheskoe modelirovanie OLAP-kuba v konexite agregirovaniya prostykh Ierarkhicheskikh izmereniy *Izvestiya Tomskogo politekhnicheskogo universiteta* 5 72 (in Russian)
- [4] Kuznetsov S D, Kudryavtsev Yu A 2009 A Mathematical Model of the OLAP Cubes *Programming and Computer Software* ISSN 0361-7688 35(5) 257–265
- [5] Zabolotnev M S Metody predstavleniya informatsii v razrezhnnykh gipercubakh dannih <http://www.olap.ru/basic/theory.asp> 18 Feb. 2014 (in Russian)
- [6] Hrustalev E M 2014 Agregatsiya dannih v OLAP-cubah <http://www.olap.ru/home/mut.asp> 17 Feb. 2014 (in Russian)
- [7] Nandi A, Cong Yu, Bohannon P, Ramakrishnan R January 2012 Data Cube Materialization and Mining over MapReduce *Transactions on knowledge and data engineering* 6(1)
- [8] Mironov V V, Makarova E S 2012 Agregatsiya pokazateley v OLAP-cubepri svedeniy po zavisimym izmereniyam *VESTNIK-UGATU* T 16 3(48) 180–186 (in Russian)
- [9] Lucian B 2010 Optimized Data Indexing Algorithms for OLAP Systems *Database Systems Journal* 1(2)
- [10] Golfarelli M, Rizzi S 2009 Data Warehouse Design. McGraw-Hill *Modern Principles and Methodologies*
- [11] You J, Xi J, Zhang C 2008 A High Performance Large Scale Data Warehouse *IEEE, In International Multi-Symposium of Computer and Computational Sciences*. 200–202
- [12] Zhang D, Zhai C, Han J 2009 Topic cube: Topic modeling for olap on multidimensional text databases *Proceedings of the Ninth SIAM International Conference on Data Mining, SIAM* 1124–1135
- [13] Nedjar S, Casali A, Cicchetti R, Lakhal L 2007 Emerging Cubes for Trends Analysis in Olap databases *Lecture Notes in Computer Science*. 4654 135–144
- [14] Cuzzocrea A, Song I-Y, Davis K C 2011 Analytics over large-scale multidimensional data: the big data revolution! *In Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP, DOLAP*. New York, NY, USA, ACM. 101 – 104
- [15] Berger A B, Gorbach I V, Melomed E L, Sherbinin V A, Stepanenko V P 2007 *Microsoft SQL Server 2005 Analysis Services. OLAP i mnogomerniy analiz dannih* SPb: BHV-Peterburg 928 (in Russian)
- [16] Borschev V, Partee B September 6, 2001 Lecture 1. Basic Concepts of Set Theory *Mathematical Linguistics* p. 1

| Authors   |  |
|---|--|
|  | <p><b>Raissa Uskenbayeva, born in September 19, 1953, Kazakhstan</b></p> <p><b>Current position, grades:</b> Vice-rector on the academic affairs, professor department of CSSE<br/>University studies: International University of Information Technologies</p> <p><b>Scientific interest:</b><br/>Macro- and micro-economics, finance and banking, Industrial Automation and Control Theory, Marketing, Management and logistics, Information technology and software engineering, <b>Informatics problems</b></p> <p><b>Publications:</b> more than 100 scientific articles, <b>monographs</b> on: <i>Theory of control and automation industry, Information Technology and Systems, Reliability of mathematical and software IP</i></p> <p><b>Experience:</b></p> <ul style="list-style-type: none"> <li>• 2012- at present Almaty: International University of Information Technologies, Vice-rector on the academic affairs</li> <li>• 2003-2012 Almaty: Kazakh National Technical University after K. I. Satpaev, The head of the Depart. «Software of systems and nets» of the Institute Information Technologies, The Doctor of Science, professor</li> <li>• 1999-2003 Almaty: Doctorate at Kazakh National Technical University after K. I. Satpaev</li> <li>• 1987–1999 Almaty: Kazakh National Technical University after K. I. Satpaev, The Candidate of Technical Science, Docent</li> <li>• 1981–1987 Almaty: Kazakh National Technical University after K. I. Satpaev, tutor</li> <li>• 1978-1981 Almaty: Research Institute of the State Planning Committee of the KazSSR, The Junior Research Fellow</li> <li>• 1975–1978 Almaty: Almaty Special. Design Bureau of The Ministry of Telecommunications Industry USSR, Engineer-mathematician</li> </ul> |
|  | <p><b>Bikesh Kurmangaliyeva, born in November 03, 1963, Kazakhstan</b></p> <p><b>Current position, grades:</b> PhD studies<br/>University studies: International University of Information Technologies</p> <p><b>Scientific interest:</b> IT management, Database, Distributed data processing, Decision support systems, Artificial intelligence</p> <p><b>Publications:</b> more than 10 scientific articles (IT management, Database, Distributed data processing, Decision support systems)</p> <p><b>Experience:</b></p> <ul style="list-style-type: none"> <li>• 2012- at present Almaty, International University of Information Technologies, Ph.D student</li> <li>• 2008–2010 Almaty: Kazakh National Technical University after K. I. Satpaev, studied in magistracy</li> </ul>  |
|  | <p><b>Nurzhan Mukazhanov, born in November 26, 1986, Kazakhstan</b></p> <p><b>Current position, grades:</b> Ph.D studies<br/>University studies: International University of Information Technologies</p> <p><b>Scientific interest:</b> Database, Information and analytical systems, Distributed data processing, Decision support systems, Artificial intelligence</p> <p><b>Publications:</b> more than 10 scientific articles (Distributed data processing, Information and analytical systems)</p> <p><b>Experience:</b></p> <ul style="list-style-type: none"> <li>• 2012- at present Almaty, International University of Information Technologies, Ph.D student</li> <li>• 2010–2012 Almaty: Kazakh National Technical University after K. I. Satpaev, tutor</li> <li>• 2008–2010 Almaty: Kazakh National Technical University after K. I. Satpaev, studied in magistracy</li> </ul>   |