

The Role of Multidimensional Databases in Modern Organizations

Ana Tănăsescu

Faculty of Economic Sciences, Petroleum-Gas University of Ploiești, Bd. București 39, 100680, Ploiești, Romania

e-mail: atanasescuro@yahoo.com

Abstract

The relational databases represent an ineffective data storage method when organizations manipulate a large amount of heterogeneous data and must realize complex analyses on these data. Thus, a new concept is required, the multidimensional database concept. In this paper the multidimensional database concept will be presented, its characteristics will be described, a comparative analysis between multidimensional databases and relational databases will be realized and a multidimensional database will be presented. This database was designed using Palo for Excel tool and can be used in order to analyse the academic staff research activity from a faculty.

Keywords: *multidimensional databases; multidimensional modelling; relational databases*

JEL Classification: *C63; C81*

Introduction

The large amount of heterogeneous data accumulated by organizations represents one of the reasons why organizations need a technology that allows them to quickly achieve complex analyses on these data. Because the relational databases are ineffective from this point of view, a new concept is required, the multidimensional database concept.

Annually, a large volume of information regarding the teaching and research activity of the academic staff from faculty departments is accumulated. The departments' managers as well as the teaching process and research activity responsible persons must, periodically, elaborate reports, synthesizing the information accumulated in different time periods. A multidimensional database where this information will be stored is required to be designed in order to streamline their activity.

In the specialty literature there are also presented other methods of organizing and analysing information from a modern organization^{1,2}.

¹ Bucur, C., Opinion Mining Platform for Intelligence in Business, *Economic Insights – Trends and Challenges*, Vol. III (LXVI), No. 3, 2014, pp. 99 – 108.

² Cambria, E. et al, New Avenues in Opinion Mining and Sentiment Analysis, *IEEE Intelligent Systems*, Vol. 28, No. 2, pp. 15-21, March-April 2013.

Further, the multidimensional database concept and its characteristics will be presented, a comparative analysis between multidimensional databases and relational databases (RDB) will be achieved and a multidimensional database designed using Palo for Excel tool will be described. This database will be used to manage the information regarding the articles publishing activity of the academic staff from a faculty.

Concept and Characteristics

A *multidimensional database* (MDB) represents a database type where the data are stored in **cells** and each cell position is defined by a number of variables named *dimensions*. Each cell signifies a business event and the dimensions values indicate the moment and the location where the event happens³.

In the specialty literature the following synonyms for the MDB concept are used: cube, hypercube, OLAP cube, OLAP database, multidimensional data store.

Although, logically a MDB resembles a cube, physically it is stored as a compressed *multidimensional array*, with offset positioning (Fig. 1 and Fig. 2).

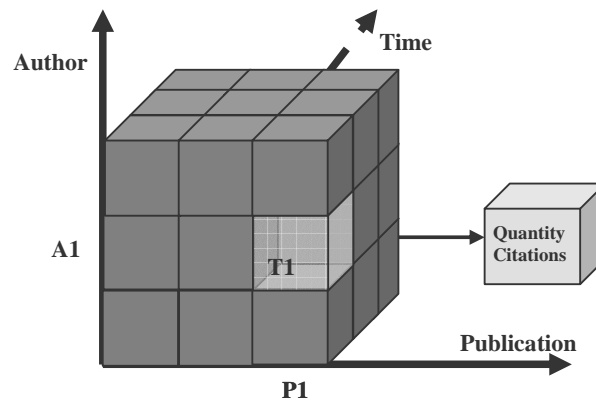


Fig. 1. The data cube for articles publishing activity management.

Source: made by the author.

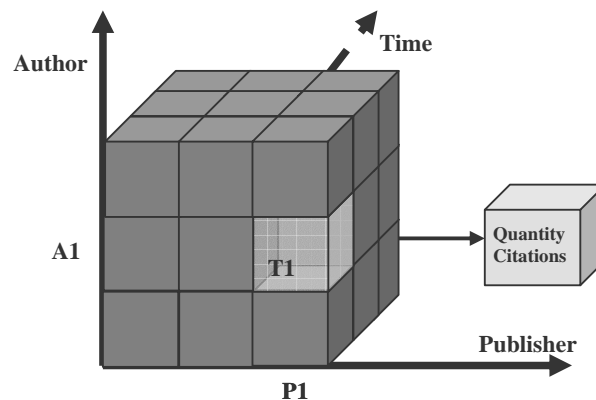


Fig. 2. The data cube for books publishing activity management.

Source: made by the author.

³ Rainardi, V., *Building a Data Warehouse With Examples in SQL Server*, Apress, 2007.

The data cube from Figure 1 has three dimensions: *Author*, *Publication* and *Time* and the cell contains two measures: *the quantity of articles from a publication* and *the citations number*. This data cube manages the information regarding the articles published by the academic staff.

As in Figure 1, the data cube from Figure 2 is intended for the management of information about the books publishing activity.

The data model corresponding to a multidimensional database design is the multidimensional data model that contains fact tables with measure type attributes and dimension tables with descriptive attributes. The members of dimension tables can be organized in hierarchies^{4,5,6,7,8,9,10,11}.

In multidimensional modelling can be used more multidimensional modelling schemas: star schema, snow flake schema etc.

The *star schema* corresponding to the data cube from Figure 1 is an improved version of a previously created schema¹² and it is presented in Figure 3. The objects of this schema represented in a star form are: the fact table (*Articles_Analysis*) that is placed in its centre and the dimension tables (*Author*, *Date*, *Publication*) that are connected to the fact table.

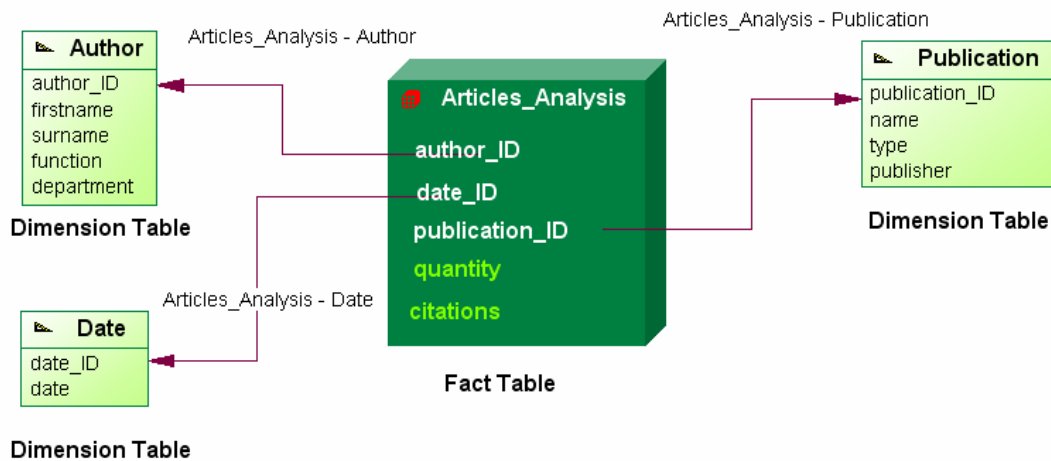


Fig. 3. Star schema for the multidimensional database.

Source: made by the author.

The *snow flake* schema (Fig. 4) that contains three more additional tables (*Function*, *Publication_Type* and *Department*) is obtained by normalizing the *Author* and *Publication* tables.

⁴ Airinei, D., *Depozite de date*, Polirom Publishing House, Iași, 2003.

⁵ Filip, F. Gh., *Decizie asistată de calculator: decizii, decidenți – metode de bază și instrumente informatice asociate*, Technical Publishing House, Bucharest, 2005.

⁶ Inmon, W. H., *Building the Data Warehouse, 4th Edition*, Wiley Publishing, 2005.

⁷ Kimball, R., Ross, M., *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition)*, Wiley, 2002.

⁸ Liu, L., Özsu, M. T., *Encyclopedia of Database Systems*, Springer, 2009.

⁹ Lungu, I., Bâra, A., *Sisteme informatice executive*, ASE Publishing House, Bucharest, 2007.

¹⁰ Vaisman, A., Zimányi, E., *Data Warehouse Systems Design and Implementation, Data-Centric Systems and Applications*, Springer, 2014.

¹¹ Nguyen, T. M., *Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development: Innovative Methods and Applications*, Information Science Reference, 2011.

¹² Pătrașcu, A., Tănăsescu A., "A Document Warehouse Architecture Model", *The Tenth International Conference On Informatics In Economy, Education, Research & Business Technology*, Bucharest, May 5-6, 2011.

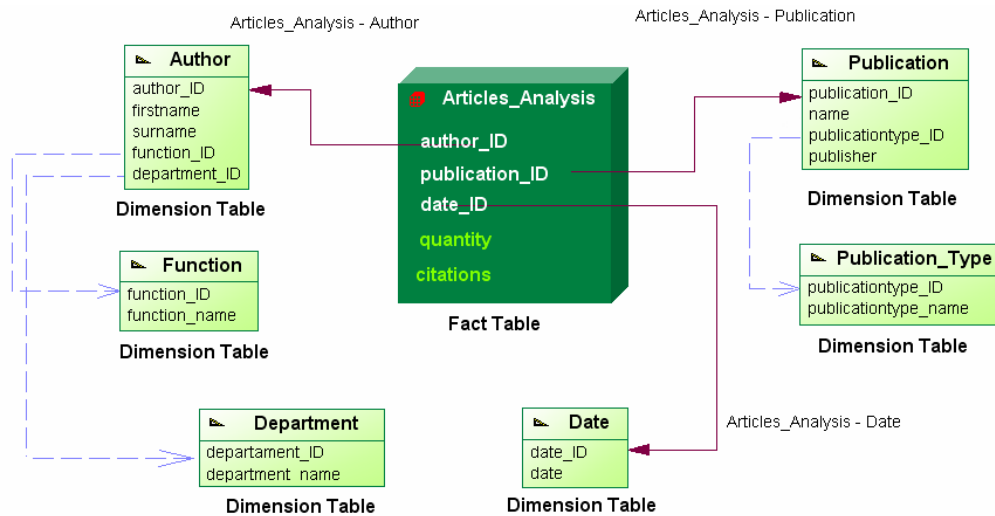


Fig. 4. Snowflake schema for the multidimensional database.

Source: made by the author.

Comparative Analysis between Multidimensional Databases and Relational Databases

Analysing the specialty literature^{13,14,15,16,17,18,19}, it cannot be stated that a certain database type is much better than the other or that the relational databases (RDB) will be entirely eliminated from the market. Both database types present advantages and disadvantages and it is important to identify the situations where a certain database type is adequate to be used.

The main differences between those two database types are emphasized in table 1.

As it can be seen in table 1, the MDB store both detailed data and aggregated data, in a multidimensional array format, being used, especially, for data analysis applications. Unlike them, in the RDB the data are stored in tables, the RDB representing the key element in the traditional transaction processing systems.

In the early 90s²⁰, there was a data query language only for the RDB, this language is the Structured Query Language (SQL), an international standardized language. In 1997 Microsoft has elaborated the Multidimensional Expressions language (MDX) as a part of the OLE DB for Olap specification. Since 2001 XML for Analysis (XMLA) has become the standard for accessing Olap data, thus, standardizing the MDX language too.

¹³ Dodds, D., Hasan, H., Gould, E., Relational versus multidimensional databases as a foundation for online analytical processing, *IRIS 22: Proceedings of the 22nd Information Systems Research Seminar in Scandinavia*, 1999, pp. 281–288.

¹⁴ Inmon, W. H., *op. cit.*

¹⁵ Kimball, R., Ross, M., *op. cit.*

¹⁶ Raden, N., Data, Data Everywhere, *Information Week*, October 30, 1995.

¹⁷ Rainardi, V., *op. cit.*

¹⁸ Vaisman, A., Zimányi, E., *op. cit.*

¹⁹ Thomsen, E., *Solutions: Building Multidimensional Information Systems*, Wiley, 2002.

²⁰ Raden, N., *op. cit.*

Table 1. MDB versus RDB

Characteristic	MDB	RDB
Data Organization	multidimensional array, cell, dimensions, fact table	table
Data Level	detailed data and aggregated data	detailed data
Typical Operation	analysis	update
Query Language	MDX	SQL
Disk Space Use	low	large
Performance	good	weak
Flexibility	low	high
Data Processing Time	high	low
Scalability	low	high
Design Scope	limited	very large
Data Access	support direct access to data	support indirect access to data

Source: made by the author.

The MDB present *advantages* such as:

- use a *less disk space* because they are compressed and do not use indexing in order to search data;
- have *better performances* in servicing the user needs because the aggregated data are precalculated and the storage format reduces the input-output operations number;
- ensure a higher performance because they are modelled taking into account only the requirements of a single users group.
- support a direct, efficient and fast access to data.

However, the MDB have *disadvantages* too:

- *the data processing time is high* as the aggregated data need to be recalculated when the operational data sources are updated;
- *the scalability decreases* as the number of dimensions increases or when the database is very large;
- *the MDB flexibility is low*;
- *the design scope is limited*, thus, as the number of users groups whose requirements must be taken into account increases, the model risks to fail.

The Multidimensional Database Design

The multidimensional database that can be used to *manage information regarding the articles publishing activity of the academic staff from a faculty* was built using *Palo for Excel* tool, created by the Jedox AG company^{21,22}. The product premium versions were named Jedox at the end of the 2011, but, currently, both open-source versions and premium versions are called Jedox.

²¹ Jedox AG, *First Steps with Palo for Excel 3.2, Palo plan analyse report*, 2011.

²² <http://www.jedox.com/en/> [accessed on 10th of June 2015].

Palo for Excel is a MOLAP database, composed of the Palo Olap server and an AddIn for Excel or OpenOffice Calc component.

Palo is a *cell-related, in memory, multidimensional and hierarchical* database²³.

I have used the *Modeller* tool (Fig. 5) in order to create the data cube, the cube dimensions, the dimensions' base and consolidated elements.



Fig. 5. Olap data cube design with Modeller tool in Palo for Excel.

Source: made by the author.

Thus, the *Statistics* data cube 8 dimensions are displayed in Figure 5:

- *Author, Function, Department, Publication, Publication_Type* that correspond to the five homonyms dimension tables from the snow flake schema;
- *Year* and *Month* that detail the *Date* dimension table;
- *Measure* that contains the two measure type attributes (*Quanty* and *Citations*) from the fact table of the snow flake schema.

The base elements and the consolidated elements of the *Publication_Type* dimension are shown in Figure 6.

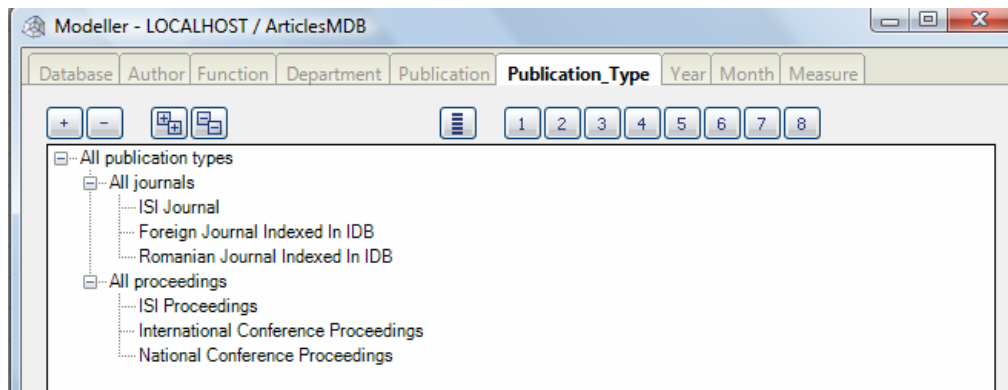


Fig. 6. Adding elements to dimensions.

Source: made by the author.

²³ Jedox, A.G., *op. cit.*

The data from the Olap data cube that will be used to manage the information regarding the articles publishing activity can be manually added or can be imported from an external data source (flat text file, ODBC query, other data cube), using the *Data Import Wizard* tool.

The Olap data cube described in this paper was populated by importing the data from the *EvidentaArticole* relational database, with the SQL query presented in Figure 7.

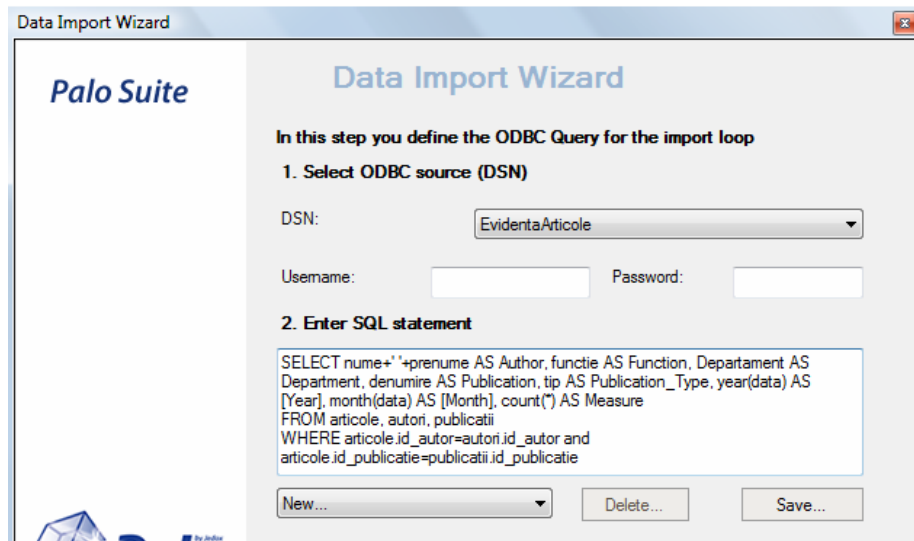


Fig. 7. Importing data in Palo from a relational database.

Source: made by the author.

Complex multidimensional analyses on the data from the data cube can be realized based on the seven dimensions that contain only descriptive elements.

A multidimensional analysis example of the data from the multidimensional database is presented in Figure 8 where the articles publishing activity of a professor is analyzed on each year and on each publication type.

	ISI Journal	Foreign Journal Indexed in	Romanian Journal Indexed in IDB	All journals	All journals	ISI Proceedings	International Conference Proceedings	National Conference Proceeding	All proceedings	All publication types
2003	0	0	0	0	0	0	1	0	1	1
2004	0	0	0	0	0	0	2	0	2	2
2005	0	0	0	0	0	0	1	0	1	1
2006	0	0	1	1	1	0	1	0	1	2
2007	0	0	0	0	0	1	1	0	2	2
2008	0	0	3	3	3	0	0	0	0	3
2009	0	0	0	0	0	3	1	0	4	4
2010	2	0	0	2	2	2	0	0	2	4
2011	0	0	0	0	0	0	1	0	1	1
2012	0	0	0	0	0	0	0	0	0	0
2013	0	0	1	1	1	0	0	0	0	1
2014	0	0	1	1	1	0	0	0	0	1
2015	0	0	0	0	0	0	0	0	0	0
All years	2	0	6	8	8	6	8	0	14	22

Fig. 8. Retrieving data from the multidimensional database.

Source: made by the author.

Conclusions

The MDB described in this paper can be easily updated and adapted in order to be used by the research activity responsible persons from any faculty, regardless of profile. It presents the following advantages:

- provides a high performance;
- supports complex analyses;
- uses a less disk space;
- supports direct access to data;
- is adequate to a limited design scope (department, faculty).

References

1. Airinei, D., *Depozite de date*, Polirom Publishing House, Iași, 2003.
2. Bucur, C., Opinion Mining Platform for Intelligence in Business, *Economic Insights – Trends and Challenges*, Vol. III (LXVI), No. 3, 2014, pp. 99 – 108.
3. Cambria, E. et al, New Avenues in Opinion Mining and Sentiment Analysis, *IEEE Intelligent Systems*, Vol. 28, No. 2, pp. 15-21, March-April 2013.
4. Dodds, D., Hasan, H., Gould, E., Relational versus multidimensional databases as a foundation for online analytical processing, *IRIS 22: Proceedings of the 22nd Information Systems Research Seminar in Scandinavia*, 1999, pp. 281–288.
5. Filip, F. Gh., *Decizie asistată de calculator: decizii, decidenți – metode de bază și instrumente informatice asociate*, Tehnical Publishing House, Bucharest, 2005.
6. Inmon, W. H., *Building the Data Warehouse, 4th Edition*, Wiley Publishing, 2005.
7. Jedox, A.G., *First Steps with Palo for Excell 3.2, Palo plan analyse report*, 2011.
8. Kimball, R., Ross, M., *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition)*, Wiley, 2002.
9. Liu, L., Özsu, M. T., *Encyclopedia of Database Systems*, Springer, 2009.
10. Lungu, I., Bâra, A., *Sisteme informatice executive*, ASE Publishing House, Bucharest, 2007.
11. Nguyen, T. M., *Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development: Innovative Methods and Applications*, Information Science Reference, 2011.
12. Pătrașcu, A., Tănăsescu A., “A Document Warehouse Architecture Model”, *The Tenth International Conference On Informatics In Economy, Education, Research & Business Technology*, Bucharest, May 5-6, 2011.
13. Raden, N., Data, Data Everywhere, *Information Week*, October 30, 1995.
14. Rainardi, V., *Building a Data Warehouse With Examples in SQL Server*, Apress, 2007.
15. Thomsen, E., *OLAP Solutions: Building Multidimensional Information Systems*, Wiley, 2002.
16. Vaisman, A., Zimányi, E., *Data Warehouse Systems Design and Implementation, Data-Centric Systems and Applications*, Springer, 2014.
17. <http://www.jedox.com/en/> [accessed on 10th of June 2015].