МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2012
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2012
Proceedings of the Forty First Spring Conference
of the Union of Bulgarian Mathematicians
Borovetz, April 9–12, 2012

WAREHOUSING AND OLAP ANALYSIS OF DATA ABOUT UNIQUE BULGARIAN BELLS*

Tihomir Trifonov, Tsvetanka Georgieva-Trifonova

In this paper, the system bgBell/OLAP for warehousing and online analytical processing data about unique Bulgarian bells is proposed. The implemented system provides a possibility for retrieving summarized reports and analyzing different characteristics of the bells with the purpose of extracting the previously unknown and potentially useful information.

1. Introduction. In recent years, the communication and information technologies have been introduced to all areas of the public life. The development of services, which give possibilities for maintenance and dissemination of information, obtained from the examination of a national cultural–historical inheritance of the separate peoples is very important and actual.

The aim of [8] is to propose a Web based approach to managing an audio and video archive for unique Bulgarian bells. The developed client/server system provides to the users a possibility for accessing information about different characteristics of the bells according to their specific interests. The data of the archive is accessible from [9]. The storage of the collected data about the bells in a database makes suitable conditions for their analyzing with the purpose of extracting the previously unknown and potentially useful information. This is the basic motivation for applying the data warehousing and OLAP (*online analytical processing*) technology on the data about the bells.

The main purpose of the system *bgBell/OLAP* is to provide a possibility for monitoring and comparing the characteristics of the bells.

More concretely, the implemented system can be utilized for the following:

- analyzing the data about the bells collected from the usage of the client/server system for managing an audio and video archive for unique Bulgarian bells;
- outputting the summarized reports about the sizes, the weights, the frequencies of the partials and the number of the bells by location, year of creation, state of the bells, type of the bells, material of the bells, creators of the bells, notes of the first partials of the bells;
- discovering the most frequent notes in the first five partials of the bells.

The basic features of the system *bgBell/OLAP* are divided by four groups:

- loading the data in the data warehouse periodically by a given schedule;
- calculating and maintaining the summarized data in the data cube;

*2000 Mathematics Subject Classification: 68U35, 68N99.
Key words: data warehousing, online analytical processing, bells.

- browsing the summarized data with the purpose of their analyzing by different dimensions in a tabular and a graphical view through Microsoft Excel application;
- exporting the summarized data in PDF, HTML, XML, others formats.

2. OLAP systems and bells.

2.1. OLAP systems. OLAP systems can be used for periodic reporting and data integrity checking. Analysts can interactively browse hierarchical and summarized data in order to extract new knowledge from the database. The traditional relational systems for database management that does not support OLAP, are appropriate for storing the data needed for daily activities and transactions processing. They are not suitable for performing complex queries that access large datasets and make multiple scans, joins and summaries, because they require more time for answer [1, 2]. Minimizing the response time of these queries proves crucial influence at designing OLAP applications.

2.2. The system bgBell/OLAP and other OLAP systems. Applying OLAP technology could solve important issues regarding databases storing information that is obtained from studding our cultural–historical inheritance. According to our research OLAP systems are not being implemented on the databases containing data about the bells. In this sense the proposed system *bgBell/OLAP* is unique. Its purpose is applying OLAP technology to exploring the data obtained from the client/server system for managing an audio and video archive for valuable Bulgarian bells. The developed system allows analyzing the sizes and the sounds of the bells by years, by locations, by creators, by types, etc.

3. Designing and implementing the system *bgBell/OLAP*. The development of the system *bgBell/OLAP* includes designing and implementing a data warehouse; a package for loading the data in the warehouse; a data cube; a client application for visualizing the results.

3.1. Architecture of the system bgBell/OLAP. The architecture of the system *bgBell/OLAP* is represented in figure 1.

The relational database BellDB is designed and created for the purposes of the client/server system for managing an audio and video archive for unique Bulgarian bells. The structure of this database is described in detail in [8].

The structure of the data warehouse BellDW and the implementation of the process of data extraction, transformation and loading (*ETL*) are represented in section 3.2.

On the basis of the dimension tables and the measures in the fact tables defined in the design of the data warehouse, the dimensions and the measures of the OLAP data cube BellCube are determined. The structure of the data cubes is described in section 3.3.1.

The usage of the OLAP data cube BellCube for analyzing the data about the bells is performed with an application that is implemented for the purpose with the means of Microsoft Excel. In section 3.3.2 an exemplary table and graphs, which visualize summarized data from the data cube, are represented.

3.2. Warehousing the data about the bells. The data warehouse serves for the data accumulation and organization with the aim of providing this data for analyzing. The purpose of the data warehouse determines the data model used for its designing.

3.2.1. Data modeling in data warehouse. The design of the data warehouses is based on the multidimensional model of the data [1, 2]. This model includes several numeric measures, which are liable for analysis. Each measure depends on a set of dimensions. The normalization of the data is used for designing databases in OLTP

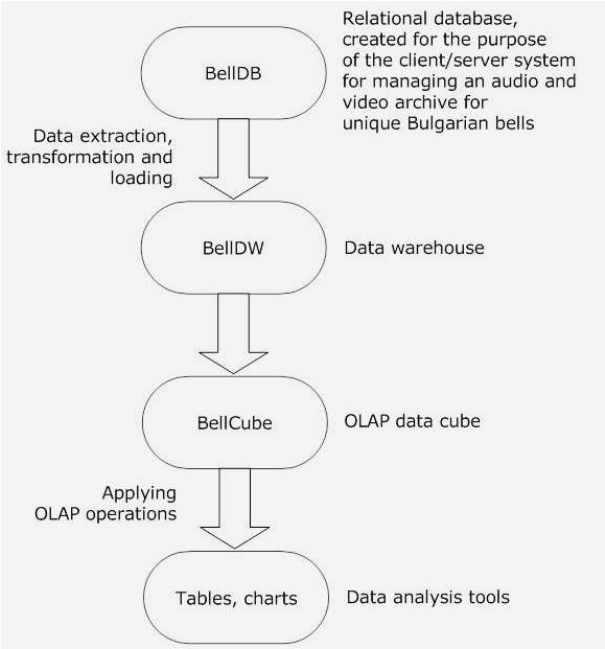

Fig. 1. Architecture of the system bgBell/OLAP

(*Online Transaction Processing*) environment, but it is unsuitable for designing data warehouses. The physical implementation of the multidimensional model requires two types of tables: dimension tables and fact tables.

The multidimensional model can be represented with a star schema, a snowflake schema or a galaxy schema. The model of the data warehouse BellDW is designed in conformity with the star schema (Fig. 2).

The dimension tables in the data warehouse BellDW store the data about the bell's location and its type; the year of the creation of the bells; the state of the bells; the type of the bells; the material of the bells; the creators of the bells; the notes of the first five partials of the bells. The fact table *Bell_facts* includes attributes which refer the dimension tables and the measure attributes: *Min_BellHeight*, *Avg_BellHeight*, *Max_BellHeight* – the minimal, average and maximal outer height of the bells; *Min_BellDiameter*, *Avg_BellDiameter*, *Max_BellDiameter* – the minimal, average and maximal bottom diameter of the bells; *Min_BellWeight*, *Avg_BellWeight*, *Max_BellWeight* – the minimal, average and maximal weight of the bells; *Min_FirstPartial*, *Avg_FirstPartial*, *Avg_First Partial*, *Min_SecondPartial*, *Avg_SecondPartial*, *Max_SecondPartial*, *Min_ThirdPartial*, *Avg_ThirdPartial*, *Max_ThirdPartial*, *Min_FourthPartial*, *Avg_FourthPartial*, *Max_Fourth Partial*, *Min_FifthPartial*, *Avg_FifthPartial*, *Max_FifthPartial* – the minimal, average and maximal frequency of the first five partials of the bells; *Count_Bells* – the number of the bells.

We have taken advantage of the database management system MS SQL Server [5, 6] to implement the data warehouse BellDW.

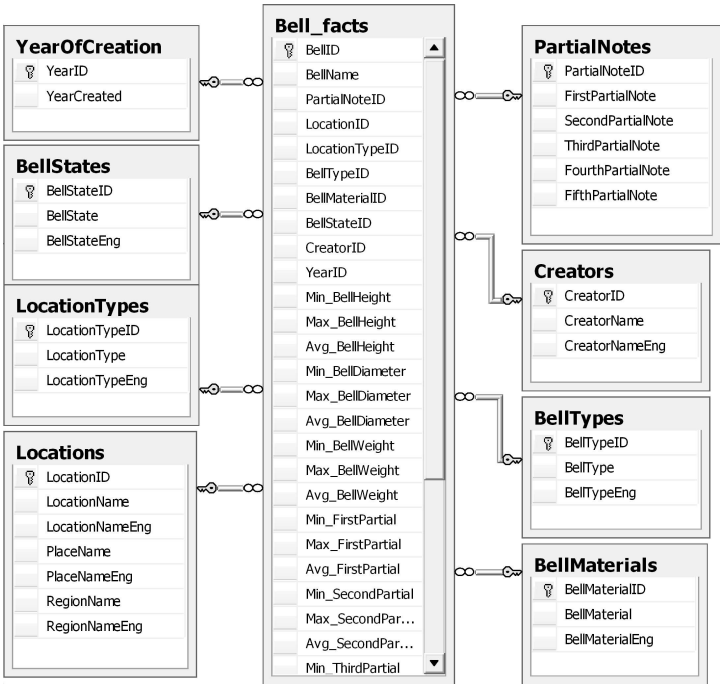

Fig. 2. Star schema of BellDW

3.2.2. Data extraction, transformation and loading in the data warehouse. The data loading into the data warehouse BellDW is performed with a package created by using *SQL Server Integration Services* [4].

The following tasks are included in the package:

- populating the dimension tables: *Locations*, *LocationTypes*, *BellTypes*, *BellMaterials*, *BellStates*, *Creators*, *YearOfCreation*, *PartialNotes*;
- populating the fact table *Bell_facts*.

The service SQL Server Agent provides a possibility for creating a package job, which includes performing the package for data extraction, transformation and loading on given schedule.

3.3. Online analytical processing the data about the bells. Computing and sorting the summarized data, which are stored separately from the data sources for online transaction processing, decreases the quantity of the data for processing, when it is necessary for the users to analyze large amount of information. The organization of the data in the data warehouse into the structures corresponding to the multidimensional model and their previously processing provides maximal performance for the queries, which summarize the data by different ways.

3.3.1. Designing and building the data cube. The data cube is a structure intended for providing fast access to the data in the data warehouse. It is a basic target for analytical processing the data. The data cube stores previously computed summaries of the data. The creation and the usage of the data cube eliminate the necessity from joining the

tables and re-computing the values returned from the most frequent executed queries.

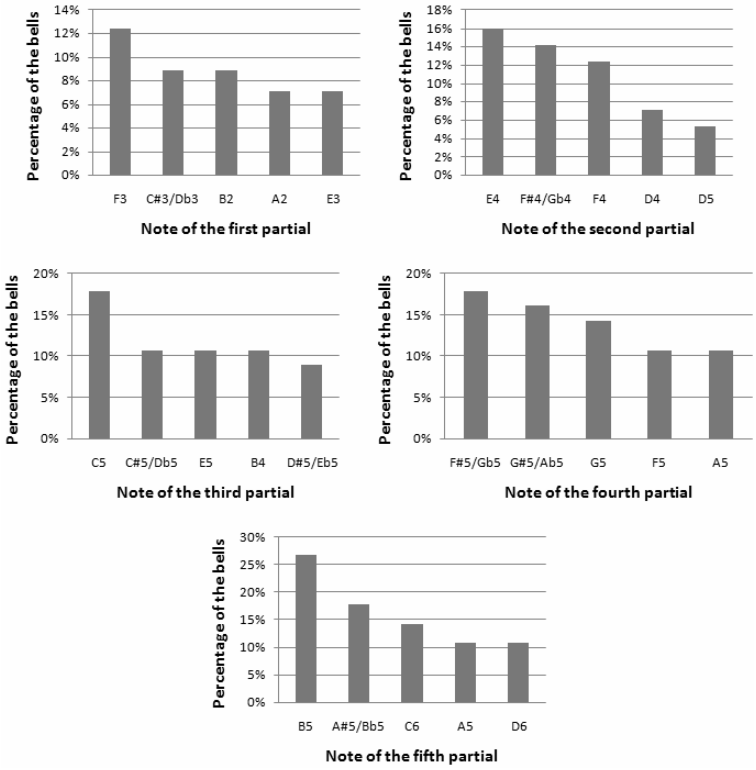

Fig. 3. The most frequent notes in the first five partials of the bells

The dimensions and the measures in the data cube are determined by the dimension tables and the measures in the fact table in the data warehouse. The values of the measures are obtained in correspondence with eight dimensions: bell's location – *Locations*; type of the location – *LocationTypes*; bell's type – *BellTypes*; bell's material – *BellMaterials*; bell's state – *BellStates*; bell's creator – *Creators*; bell's year of creation – *YearOfCreation*; notes of the first five partials of the bells – *PartialNotes*. A hierarchy is defined for the dimension of the location and the place, the region where the bells are situated. Therefore, the summarized data can be returned for chosen places and/or regions. The possible types of the locations are: churches, monasteries, museums, castles, etc.

The data cube BellCube is created with *SQL Server Analysis Services* [7].

3.3.2. OLAP analyzing the data about the bells. For the end user, an application is implemented with the means of Microsoft Excel [3, 7]. This application allows extraction of the summarized data from the data cube BellCube and their representation in a tabular view and a graphical view. The users can access multiple reports with the application and some of them are:

- the most frequent notes in the first five partials of the bells (Fig. 3);

- the frequencies of the first five partials and the number of the bells by the places (Fig. 4);
- the frequencies of the first five partials of the several bells; the frequencies of the first five partials of the three bells are shown in Fig. 5 – one of the bells (2_01_01) is situated in the church "St. Alexander Nevski", Sofia, and the others two (2_02_01 and 2_02_02) are from National historical museum, Sofia;
- the sizes and the weights of the bells by the notes of the first five partials of the bells;
- the sizes, the weights, the frequencies of the first five partials and the number of the bells by the type of the location (i.e. church, monastery, museum, etc.);
- the frequencies of the first five partials and the number of the bells by their year of creation.

	Values					
Row Labels	Max First Partial	Max Second Partial	Max Third Partial	Max Fourth Partial	Max Fifth Partial	Count Bells
Veliko Tarnovo	**450**	**750**	**920**	**1100**	**1330**	**19**
Kapinovo	450	600	920	1100	1240	6
Kilifarevo	330	750	810	1050	1330	2
Lyaskovets	190	380	650	870	1170	1
Veliko Tarnovo	330	600	750	910	1140	10

Fig. 4. The frequencies of the first five partials and the number of the bells by the places in the region of Veliko Tarnovo

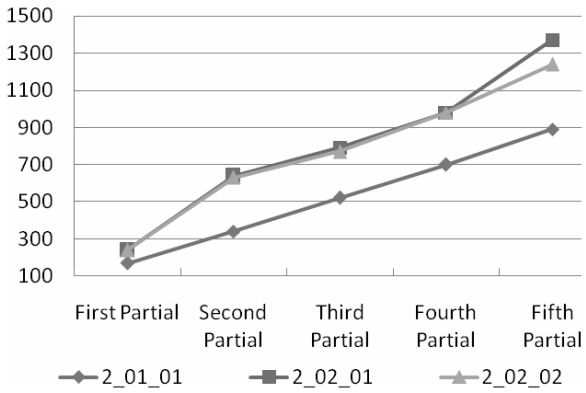

Fig. 5. The frequencies of the first five partials of the bells

4. Conclusion. In the present paper, an application of OLAP technology for analyzing the data about the unique Bulgarian bells is represented. The structure of the created data warehouse is described, as well as the implementation of the ETL process, the structure of the OLAP data cube, the features of the application designed for the end user.

Our future work includes applying the algorithms for data mining and development of an application providing possibilities for mining the data about the bells.

REFERENCES

[1] A. A. Barsegyan, M. S. Kupriyanov, V. V. Stepanenko, I. I. Holod. Technologies for Data Analysis: Data Mining, Visual Mining, Text Mining, OLAP. BHV-Peterburg, 2008.

[2] W. H. Inmon. Building the Data Warehouse. Wiley Publishing, Inc., 2005.

[3] J. Krishnaswamy. On Accessing Data From An OLAP Server Using MS Excel. http://www.aspfree.com/c/a/MS-SQL-Server/On-Accessing-Data-From-An-OLAP-Server-Using-MS-Excel, 2005.

[4] Microsoft Corporation, SQL Server Integration Services. http://msdn.microsoft.com/en-us/library/ms141026.aspx, 2008.

[5] J. Mundy, W. Thornthwaite, R. Kimball. The Microsoft Data Warehouse Toolkit: With SQL Server 2005 and the Microsoft Business Intelligence Toolset. John Wiley & Sons, 2006.

[6] V. Rainardi. Building a Data Warehouse: With Examples in SQL Server. Apress, 2007.

[7] J. Shumate. A Practical Guide to Microsoft OLAP Server. Addison-Wesley, 2000.

[8] T. Trifonov, T. Georgieva. Client/server system for managing an audio and video archive for unique Bulgarian bells. *WSEAS Transactions on Information Science and Applications*, **6** (2009) No 4, 660–669.

[9] The Bell Project "Research and Identification of Valuable Bells of the Historic and Culture Heritage of Bulgaria and Development of Audio and Video Archive with Advanced Technologies" Website, http://www.math.bas.bg/bells/belleng.html.

Tihomir Trifonov
University of Veliko Tarnovo
Faculty of Mathematics and Informatics
Department of Algebra and Geometry
e-mail: tihomirtrifonov@ieee.org

Tsvetanka Georgieva-Trifonova
University of Veliko Tarnovo
Faculty of Mathematics and Informatics
Department of Information Technologies
e-mail: cv.georgieva@uni-vt.bg

СКЛАДИРАНЕ И OLAP АНАЛИЗ НА ДАННИ ЗА УНИКАЛНИ БЪЛГАРСКИ КАМБАНИ

Тихомир Трифонов, Цветанка Георгиева-Трифонова

В настоящата статия е представена системата *bgBell/OLAP* за складиране и онлайн аналитична обработка на данни за уникални български камбани. Реализираната система предоставя възможност за извеждане на обобщени справки и анализиране на различни характеристики на камбаните, за да се извлече предварително неизвестна и потенциално полезна информация.