

Research of the possibilities of application of the Data Warehouse in the construction area

*Alexandr Konikov**, *Ekaterina Kulikova* and *Olga Stifeeva*

Moscow State University of Civil Engineering, Yaroslavskoye shosse 26, Moscow, 129337, Russia

Abstract. Today, in information technologies, the direction associated with the use of Data Warehouse (DW) is evolving very dynamically. Using DW, it is possible to implement two types of data analysis: OLAP-analysis: a set of technologies for the rapid processing of data presented as a multidimensional cube; Data Mining is an intelligent, deep analysis of data to detect previously unknown, practically useful patterns (in our case, the construction area). It is noted, that of all the methods used in technology Data Mining, cluster analysis is especially useful for the construction area. At present, the role of DW has increased, significantly due to the fact, that many methods and approaches of Data Mining have formed the basis of a new, promising method of Big Data. We will specify that, that Data processing from the Data Warehouse with the help of technology Big Data, allows to deduce researches in a building area to the higher level. The purpose of this work is to research of the possibilities of application of the Data Warehouse in the construction area. The article suggests the new approach to data analysis in the construction area, based on the use of Big Data technology and elements of OLAP - analysis. In the section "Discussion" is considering the possibility of the new promising business in the construction field, based on the application of Data Warehouse and technology Big Data.

1 Introduction

Data Warehouse (DW) is widely used in data processing. Recently, due to the advent of intelligent technologies, particularly Big Data, their importance has increased significantly. Indeed, the Big Data technology has largely inherited the principles and methods of the earlier Date Mining intellectual technology, which in turn is based on the use of the Data Warehouse. It should be noted that of all the methods used in technology Data Mining and Big Data, cluster analysis is especially useful for the construction area. We can formulate the obvious conclusion: Big Data technology has very good prospects for the construction industry [1-4]. Accordingly, the role of the Data Warehouse in the field of construction has increased. It is possible to create the new forms of business, based on the use of technologies Data Warehouse and Big Data. The purpose of this work: research of the

* Corresponding author: a.konikov@gmail.com

possibilities of application of the Data Warehouse in the construction area and to explore the possibility of establishing new forms of business.

2 Materials and Methods

First of all, recall the basic principles of Data Warehouse [5]. Data is merged into categories and stored according to the areas they describe, not the applications they use. The data is combined so that it satisfies all the requirements of the enterprise as a whole, not the only function of the business. Data in the Data Warehouse is not created: that is, it comes from external sources, are not adjusted or deleted. The data in the vault is accurate and correct only when it is bound to a certain time.

For the construction industry, last principle mean that it is necessary to store data on the construction objects with time binding in DW. Important point: when you fill DW with data, the value of the total amount of information DW increases, respectively, you can get better results in the processing of information (first of all, is meant processing with Using the Big Data technology).

Let us first consider the "traditional" use of DW in Business Analytics, in the subject area under consideration-construction. It is assumed two types of analysis:

- OLAP Analysis: A set of technologies for the rapid processing of data presented as a multidimensional cube.
- Data mining is an intelligent, in-depth analysis of the data, for detection of previously unknown, practically useful regularities and knowledge necessary for decision making (in our case, for construction industry).

First, about OLAP - analysis. This term defines the category of applications and technologies that enable carry out the collection, storage, prompt processing, and analysis of multidimensional data.

The information is presented in the form of multidimensional cubes, where the measurements are the parameters of the object, and the cells contain the aggregated data [6]. As an example, figure 1 shows an example of a multidimensional cube in the construction industry: the X-axis is a type of building, the Y axis is the time interval Q1, Q2, Q3,Q4 (Q1 - January, February, Mart) , and the Z axis is the name of construction company. The cells contain specific indicators (for example, an integrated sum of investments in tens of millions of rubles).

For a multidimensional cube, for different axes are produce the slices - to bring data to tables, analyze them, and prepare reports based on that.

In Figure 1, as an example, the slice is produced on the z axis for Construction Company1. Once again, we note the main property of OLAP analysis - managers can prepare the necessary report in a relatively short time.

Unlike OLAP analysis, data mining technology requires much more time. This development technology allows to study the hidden depth patterns and on the basis of this, plan strategic approaches to the construction business.

Data mining is a deep analysis of data, for detection of previously unknown, practically useful regularities and knowledge necessary for making decisions in the construction area. Data Mining is based on the following methods: Associative rules, Decision trees, Classification algorithms, Artificial neural networks, Genetic algorithms, Memory-based Reasoning, MBR, Case-based Reasoning, CBR, Cluster analysis [7].

It should be noted that cluster analysis is especially useful for the construction area. In short, cluster analysis is a multidimensional statistical procedure that collects data that contains information about objects and then orders objects into relatively homogeneous groups. In the context of our consideration, this method allows the uniting of building objects into homogeneous groups and then purposefully exploring these groups.

Many of the Data Mining methods in particular cluster analysis has shifted to later technology – big data. In addition, a number of new approaches were used in big data, such as crowdsourcing [8], data fusion and integration [9] and others.

As a result, big data technology made it possible to quickly process structured and unstructured data of huge volumes and significant diversity. Often for short characteristic of technology Big Data use VVV, which mean: V-Volume: technology allows processing very large amounts of data; V-speed: high speed processing and obtaining results; V-Diversity: possibility of simultaneous processing of different types of data. Now on the subject "Big data" there is a large number of works in which this direction is investigated more in depth and in detail (for example, work [10]), therefore we will confine ourselves to the information set above.

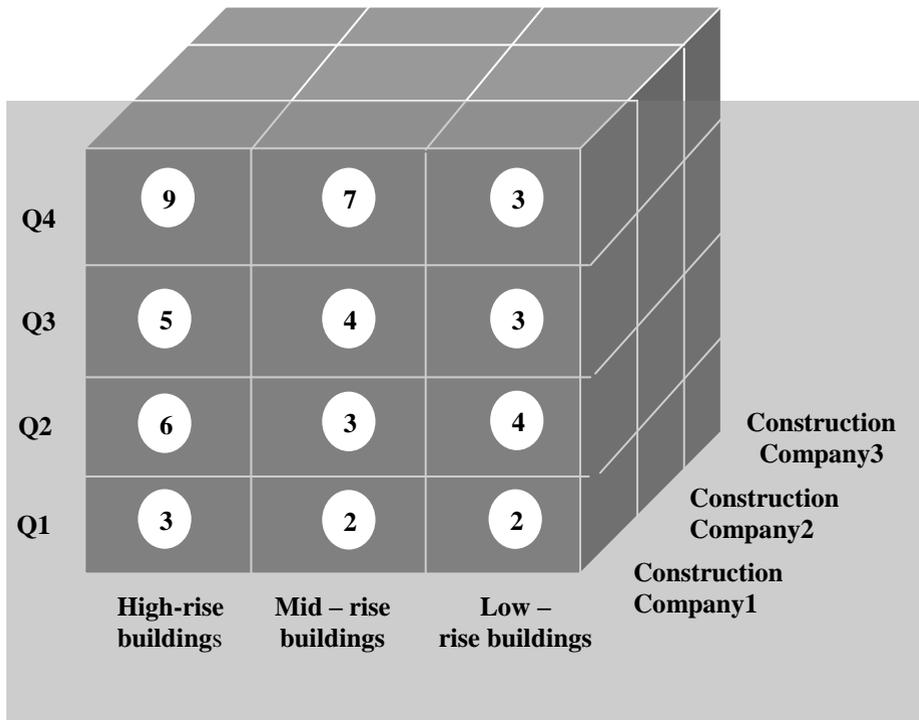


Fig. 1. The multidimensional cube.

As a result, big data technology made it possible to quickly process structured and unstructured data of huge volumes and significant diversity. Often for short characteristic of technology Big Data use VVV, which mean: V-Volume: technology allows processing very large amounts of data; V-speed: high speed processing and obtaining results; V-Diversity: possibility of simultaneous processing of different types of data. Now on the subject "Big data" there is a large number of works in which this direction is investigated more in depth and in detail (for example, work [10]), therefore we will confine ourselves to the information set above.

On further consideration, we will use elements of the theory of sets.

Note, that often in the construction area we have to deal with a very large amount of data. Let's consider, as an example, the problem of processing of the data received as a result of operation of multi-storey houses. Each of these objects is characterized by own dataset. They can be represented as:

$$Q_1 \{q_{11}, q_{12}, \dots, q_{1n}\}, Q_2 \{q_{21}, q_{22}, \dots, q_{2n}\}, \dots, Q_n \{q_{n1}, q_{n2}, \dots, q_{nn}\} \quad (1)$$

All data can be represented as sum of sets:

$$Q = Q_1 \cup Q_2 \cup \dots \cup Q_n \quad (2)$$

If you use for storage and processing data DW, each element from a set of Q_k should be represented as a set, whose elements bound to time values: t_1, t_2, \dots, t_p .

Thus, the data of a large-scale construction object (for example - a group of multi-storey houses) represent a huge amount of time-bound numerical data. This fact is one of the main reasons for the use of data warehouse in large-scale construction. The ability to store such large amounts of data was before the technology Big Data, but there were no technologies that would allow to process this data in a rather short period of time. Data processing from the Data Warehouse with the help of technology Big Data allows to deduce researches in a building area to the higher level. The following pattern is characteristic: as new data is added to DW, the quality of the analysis results makes better.

Consider other version of the application of DW and Big Data in the field of construction. Typically, a construction company enters into a contract with an IT company, that owns big Data technology, to get a number of results. For a construction company it would be more interesting to get some tool that allows analyze different variants by manipulating the data obtained from the Big Data analysis. For example, date can be represented as a multidimensional cube Data.

This solution is shown in Figure 2.

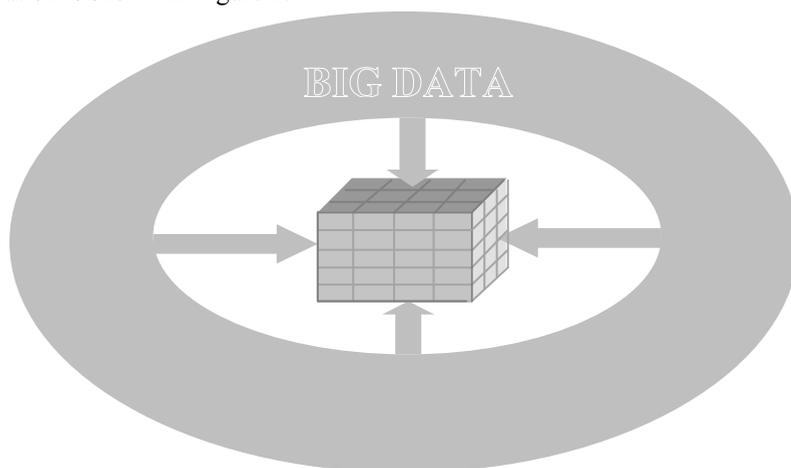


Fig. 2. The multidimensional cube obtained with the help of technology Big Data.

For clarity, we suggest that the X-axis is the district of the city, Y-year, Z -type of the building. The cells contain the cost of building (predicted with the help of Big Data technology). Then, for different axes may be produce the slices - to bring data to tables, analyze them, and get the results. For example, we can get a forecasted estimate of the value of a building for a certain year.

Despite the formal resemblance to OLAP-cube, the essence of analysis is fundamentally different: OLAP-technology deals with the operational data, and the proposed method is designed to work with data obtained as a result of processing large amounts of information from using Big Data technology.

3 Conclusions

In the paper the principles of Data Warehouse construction are briefly described: data is merged into categories and stored according to the areas they describe, not the applications they use. Data in the Data Warehouse is not created: that is, it comes from external sources, are not adjusted or deleted. The data in the vault is accurate and correct only when it is bound to a certain time. The corresponding practical interpretation is given: it is necessary to save in DW data on objects of construction, store time-bound data, to aggregate data on various objects [11].

The following pattern is noted: as DW is filled with data, the value of the total amount of information DW increases, respectively, you can get better results in the processing of information.

It is stated that it is possible to implement two types of data analysis using DW. Firstly, OLAP analysis: set of technologies for the rapid processing of data presented as a multidimensional cube. Secondly, Data mining is an intelligent, in-depth analysis of the data, for detection of previously unknown, practically useful regularities and knowledge necessary for decision making (in our case, for construction industry.) Methods included in the Data Mining technology are briefly discussed. It is noted that cluster analysis is especially important for the construction industry, as this method allows to combine construction objects into homogeneous groups and then purposefully investigate these groups [12].

It is noted that the role of DW has significantly increase due to the fact, that many methods and approaches of Data Mining have formed the basis of a new, perspective technology - Big Data. It is shown that this technology can be successfully used in the construction industry. The paper proposes a new approach to the analysis of data in the construction area, based on the use of BIG data technology and elements of OLAP-analysis.

4 Discussion

This section of the article discusses the following issue (see figure 3). There are a number of enterprises in the construction industry that have useful data (for example, companies that manage operation of the buildings). On Figure 3 these companies are marked with a digit I. The main operation tasks for these companies are the maintenance of the building in proper condition. Research and intellectual analysis, for most of these companies (especially small and medium) is not conducted, many data are eventually destroyed.

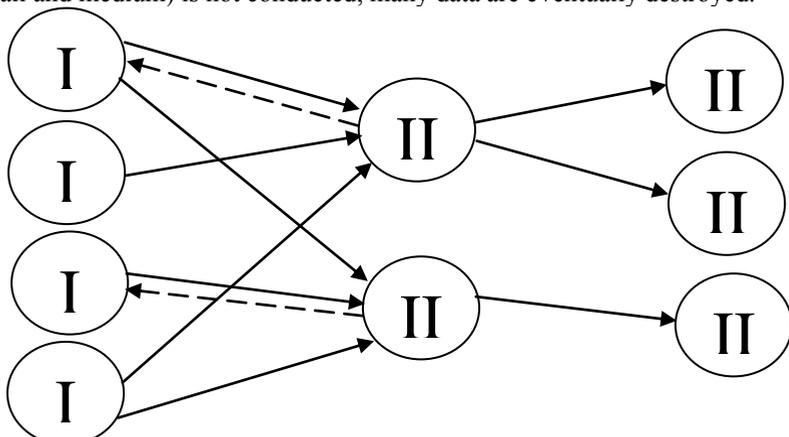


Fig. 3. Different types of companies working in the construction industry.

There are a number of enterprises (number II), which need the specified data in the processed form. Modern data processing involves the use of a Data Warehouse, an intelligent technology, such as Big Data. Companies that design new buildings can serve as type II enterprises. To improve the of results, the second type companies need data of the construction or operation of existing objects. But these companies do not have necessary amount of data for modern data processing facilities. (Note, that the use of DW, Big Data and other modern data processing facilities involves significant investments, for small and medium enterprises such a task is impossible). From the above, it is logical - need the companies of III type (figure 3) - are the companies owning, the above-mentioned technologies, in particular DW and Big Data. These companies acquire data from the campaigns I type, put them in the Data Warehouse. At the request of the type II companies, the data, stored in the DW, is processed, the results are transferred to the type II Companies.

Let's make two comments:

1) Now we consider only the statement of the problem, organizational, legal and other aspects of the task are not considered.

2) Companies of type I can make the request to companies of type III – to obtain results of data processing with the help of technology Big Data or other intellectual technology (on figure 3 the information is shown by a dashed line).

Thus, in the construction area, the following business is possible:

The company (in our terminology III), owning IT-technologies, invests the G1 funds to pay for the information services of companies I. At the requests of companies II, this company carries out the intellectual data processing, receiving the means G2. We see classic business rule – the difference between the invested and received means G2-G1 should recoup all expenses of the company III and give additional profit. In our opinion, this is quite a promising kind of business in the construction area.

References

1. Konikov A., Konikov G., 2017. Big Data is a powerful tool for environmental improvements in the construction business. IOP Conf. Series: Earth and Environmental Science 90 (2017), 012184.
2. Konikov A., 2018. A selective study of Information technologies to improve operations efficiency in construction. MATEC Web of Conferences, **170**, 01110.
3. Ivanov N., Gnevanov M., 2018. Big data: Perspectives of using in urban planning and management. MATEC Web of Conferences, **170**, 01107.
4. Valpeters M., Kireev I., Ivanov N., 2018. Application of machine learning methods in big data analytics at management of contracts in the construction industry. MATEC Web of Conferences, **170**, 01106
5. Gilemchanov, R., Bagautdinov, R., Kankhva, V. Autodesk revit and robot structural analysis in design of framed buildings. Advances in Intelligent Systems and Computing, **692** (2018)
6. Balalayeva E., Kukhar V., Tuzenko O., Glazko V., Prysiashnyi A., Artiukh V., Kankhva V. Researching of the stress-strain state of the open-type press frame using of elastic compensator of errors of “press-die” system. Advances in Intelligent Systems and Computing, **692** (2018)
7. Zuikov A., Orekhov G., Suehtina T., MATEC Web of Conferences **193**, 02024 (2018) doi.org/10.1051/mateconf/201819302024
8. Ahmad, Awais, Anand Paul, and M. Mazhar Rathore. "An efficient divide-and-conquer approach for big data analytics in machine-to-machine communication." Neurocomputing 174 (2016): 439-453.

9. M. Mazhar Rathore, Anand Paul, Awais Ahmad, Bo-Wei Chen, Bormin Huang, and Wen Ji. "Real-Time Big Data Analytical Architecture for Remote Sensing Application."
10. M. Mazhar Rathore , Anand Paul , Awais Ahmad , Suengmin Rho. Urban planning and building smart cities based on the internet of things using big data analytics. *Computer Networks* (2016), doi:10.1016/j.comnet.2015.12.023
11. Sarwat, M., Moraffah, R., Mokbel, M.F., Avery, J.L. Database system support for personalized recommendation applications. *Proceedings - International Conference on Data Engineering*, (2017) doi:10.1109/ICDE.2017.174
12. Zhang, W., Jiang, L. Algorithm Analysis for Big Data in Education Based on Depth Learning. *Wireless Personal Communications*. 102/ 4, (2018)