

# OLAP SYSTEMS AS THE MODERN DATA PREPARATION TOOLS FOR OUTDOOR ADVERTISING DATA MINING

Oleksandr Shelest<sup>1</sup>, Bella Holub<sup>2</sup>

**Abstract.** Today, most organizations use databases and at worst text documents and spreadsheet files as sources for data analysis, which prevent correct and error-free analysis. At best, the data can be constantly adjusted due to ambiguities and inaccuracies. The subject of the study is the intellectual analysis of outdoor advertising data. The methodology of successful data analysis is the correct storage of data, which is the basis for clear data analysis. Modern computer systems and computer networks allow the accumulation of large arrays of data to solve problems of processing and analyzing. Unfortunately, the machine form of data presentation itself contains the information that a person needs in a hidden form, and you need to use special methods of data analysis to obtain it. In order to get what you want, you need to create not just a database, but a data warehouse with a special storage structure. Thus, the data warehouse allows you to collect data from various sources, databases, table files and other things, store them throughout history and, unlike conventional databases, allows you to create systems for fast and accurate data analysis. Data warehouse is the basis for building decision support systems. Operational data is checked, cleared and aggregated before entering the data warehouse. Such integrated data is much easier to analyze. Different sources of operational data may contain data describing the same subject area from different points of view (for example, from the point of view of accounting, inventory control, planning department, etc.). A decision made on the basis of only one point of view can be ineffective or even erroneous. The goal is to use a data warehouse to integrate information that reflects different perspectives on the same subject area. Focus on the object, which will also allow the data warehouse to store only the data you need to analyze it. It will also significantly increase the speed of data access both due to the possible redundancy of the stored information and due to the exclusion of modification operations. Conclusion: the decision support system will ensure reliable storage of large amounts of data. Tasks will also be assigned to prevent unauthorized access, data backup, archiving, etc.

**Key words:** database, data warehouse, OLAP, data mining, decision support system.

**JEL Classification:** C80, Y40, M37

## Nomenclature

DB	Database
ODS	Operational data store
DW	Data warehouse
OLAP	Online analytical processing
OLTP	Online transaction processing
DSS	Decision support system
ROLAP	Relational OLAP

## 1. Introduction

The practice of using OLTP systems has shown the inefficiency of their use for full analysis of information. Such systems quite successfully

solve the problem of collecting, storing and retrieving information, but they do not meet the requirements necessary for modern DSS (Barsehyan, Kupryyanov, et al., 2009). Therefore, to enable the analysis of accumulated data, it is necessary for organizations to create data warehouses, which are integrated data collections accumulated from different systems of operative access to data.

In the process of data accumulation, the need to use data mining methods as an effective aid increases; this allows the researcher to gain additional knowledge on the subject area, in which

*Corresponding author:*

<sup>1</sup> National University of Life and Environmental Sciences of Ukraine, Ukraine.

E-mail: [Oleksandr.v.shelest@gmail.com](mailto:Oleksandr.v.shelest@gmail.com)

ORCID: <https://orcid.org/0000-0001-7240-4993>

<sup>2</sup> National University of Life and Environmental Sciences of Ukraine, Ukraine.

E-mail: [bella.golub55@gmail.com](mailto:bella.golub55@gmail.com)

ORCID: <https://orcid.org/0000-0002-1256-6138>

he or she works, and must make informed decisions (Serheiev-Horchynskyi, Ishchenko, 2018).

Most organizations take the wrong approach to analyzing data, resulting in errors and slower execution. The data is constantly being revised due to inaccuracies.

Analyst is a specialist who analyzes data, conducts research and generalizations in a particular field of activity. For the analysis to be successful, the data must be properly processed and structured.

Data analysis is a part of an interactive automated system designed to help and support various human activities in making decisions about solving structured or unstructured problems.

## 2. Formulation of the problem

The first step is to determine what data can be attributed to the detailed data. These are data that are transferred directly from ODS. They correspond to elementary events and are fixed by OLTP systems (for example, sales, prices, etc.).

The second step is to divide all the data into measurements and facts. Measurements are data sets needed to describe events (for example, cities, goods, people, etc.). Facts are data that reflect the nature of the event (for example, the number of goods sold, the efficiency of placement, etc.). Actual data can be presented as numerical or categorical values.

The third step for convenient operation you need to get information about the DW data. Such information is called metadata. According to the concept of J. Zahman, metadata should answer the following questions: what, who, where, how, when and why.

The fourth step is to organize the data transfer process, which includes the stages of extraction, conversion and download – ETL process.

The fifth step is to transform the data; at this stage, the following procedures are performed: data aggregation, translation of values, creating fields and data cleaning.

- Data aggregation will provide a smaller number of short records that will be transferred to DW.
- Translation of values replaces encoded data with clearer descriptions.
- Creating fields allows you to create not just ordinary fields, but also special fields that will exclude additional calculation operations.
- Data cleaning is aimed at detecting and removing errors and inconsistencies in the data in order to improve their quality.

## 3. Literature review

What is OLTP? Transactional system is a real-time transaction processing, known as OLTP, which supports transaction-oriented programs in a three-tier architecture. OLTP administers the organization's daily transactions.

Any data warehouse is an OLAP system. The use of OLAP is as follows, for example: a company can compare its mobile phone sales in June with sales in July, and then compare these results with another location that can be stored in a special database.

An example of an OLTP system is an online store. Let us say that several people want to buy the same product at the same time, although the number of products is limited.

The key difference is that OLAP is a category of software that analyzes data stored in a database, while OLTP supports transaction-oriented programs in a three-tier architecture (Figure 1).

OLAP creates a single platform for all types of business analysis needs, which includes planning, budgeting, forecasting, and analysis, while OLTP is useful for managing an organization's day-to-day transactions.

OLAP is characterized by a large amount of data, while OLTP is characterized by a large number of short online transactions.

The data warehouse in OLAP is created uniquely so that it can integrate different data sources to build a consolidated database, while OLTP uses traditional databases.

Advantages of using OLAP services:

- OLAP creates a single platform for all types of business intelligence needs, which includes planning, budgeting, forecasting and analysis.
- The main advantage of OLAP is the consistency of information and calculations.
- Simply build security restrictions on users and objects to enforce rules and protect sensitive data.
- The use of OLAP system gives the organization the ability to predict and analyze various situations related to current activities and prospects. These systems can be considered as an addition to enterprise level automation systems. All the advantages of OLAP systems directly depend on the accuracy, reliability and amount of source data.
- Based on detailed data, aggregated data can be obtained. Aggregation occurs by summing numerical actual data for certain dimensions.

Most DSS users do not work with detailed, but with aggregated data. DW architecture should provide a fast and convenient way to obtain

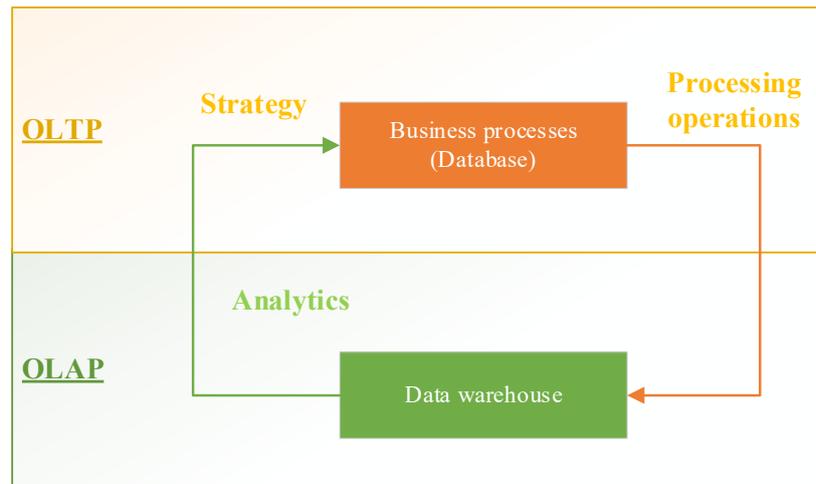


Figure 1. OLAP and OLTP

information of interest to the user. For this purpose, it is necessary to store a part of the aggregated data in DW, instead of to calculate them at performance of analytical inquiries. Obviously, this leads to redundancy of information and an increase in the size of DW. Therefore, when designing such systems, it is important to achieve the optimal ratio between the calculated and stored aggregate data. Data that is rarely accessed by users can be calculated during the execution of analytical queries. Data that is needed more often should be stored in DW.

#### 4. Research

After researching organizations engaged in outdoor advertising in Ukraine, their decision support system, if it can be called that, was analyzed (Figure 2). Also, the disadvantages of this system are presented in the publication (Shelest, Holub, 2019).

The most important criterion for the use of advertising is the number of its consumers. Outdoor advertising is one of the most effective advertising media. Where to place an ad, what size it should be, its design – all these are parameters that affect the effectiveness of its use. There is currently no full-scale system for monitoring and analyzing the process of using outdoor advertising.

As shown in Figure 2, there are two operational data stores that generate certain files for analytics on the server. Each client has local software installed, which periodically downloads files from the server and performs certain manipulations with them to obtain reports and perform data analysis.

Such a system is difficult to call a decision support system. At each client, the data for the analysis

contain locally, they are not protected in any way. If you continue to use this structure, it may lead to the loss of correct data.

In order to ensure security, clarity and accuracy in the data and analysis, the structure of the data warehouse is presented (Figure 3). Data come from ODS (operational data store). DW has a denormalized data structure for quick access to data. In the course of work the structure of DW can change.

Separate fact tables are created for possible combinations of generalization levels of different dimensions. This allows for better performance, despite the complexity of the database structure, and the large number of fact tables.

The relational system – ROLAP (Figure 4) is perfectly suitable for this OLAP system.

ROLAP servers use relational databases. According to E. Codd, “relational databases have been, are and will be the most suitable technology for data storage. There is no need in new database technology, but rather in analysis tools that complement the functions of existing databases and are flexible enough to anticipate and automate different types of intelligent databases analysis inherent in OLAP” (Barsehyan, Kupryyanov, et al., 2009).

In case of change of structure of measurements, the ROLAP system with dynamic representation of dimension is the optimum decision as such modifications will not lead to physical reorganization of a DB.

Relational databases provide a much higher level of data protection and good opportunities for delimitation of access rights.

ETL processes are presented for data transfer from different databases to DW (Figure 5).

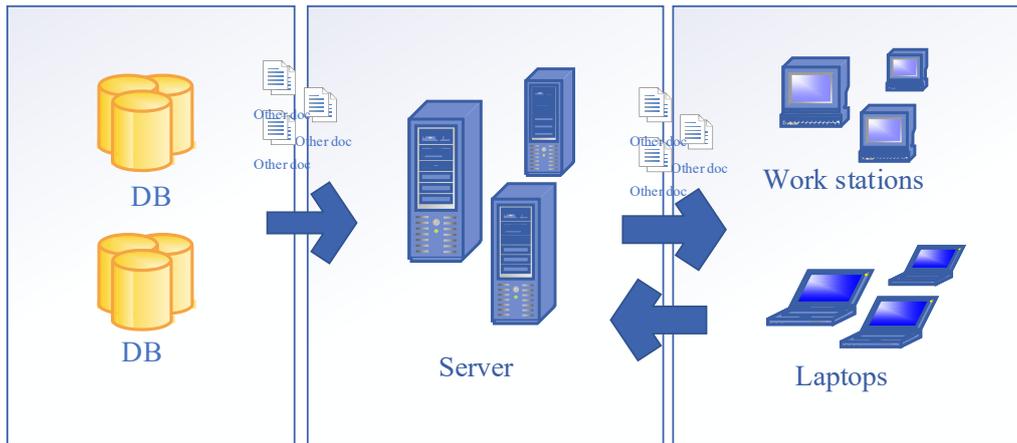


Figure 2. System of analysis of outdoor advertising

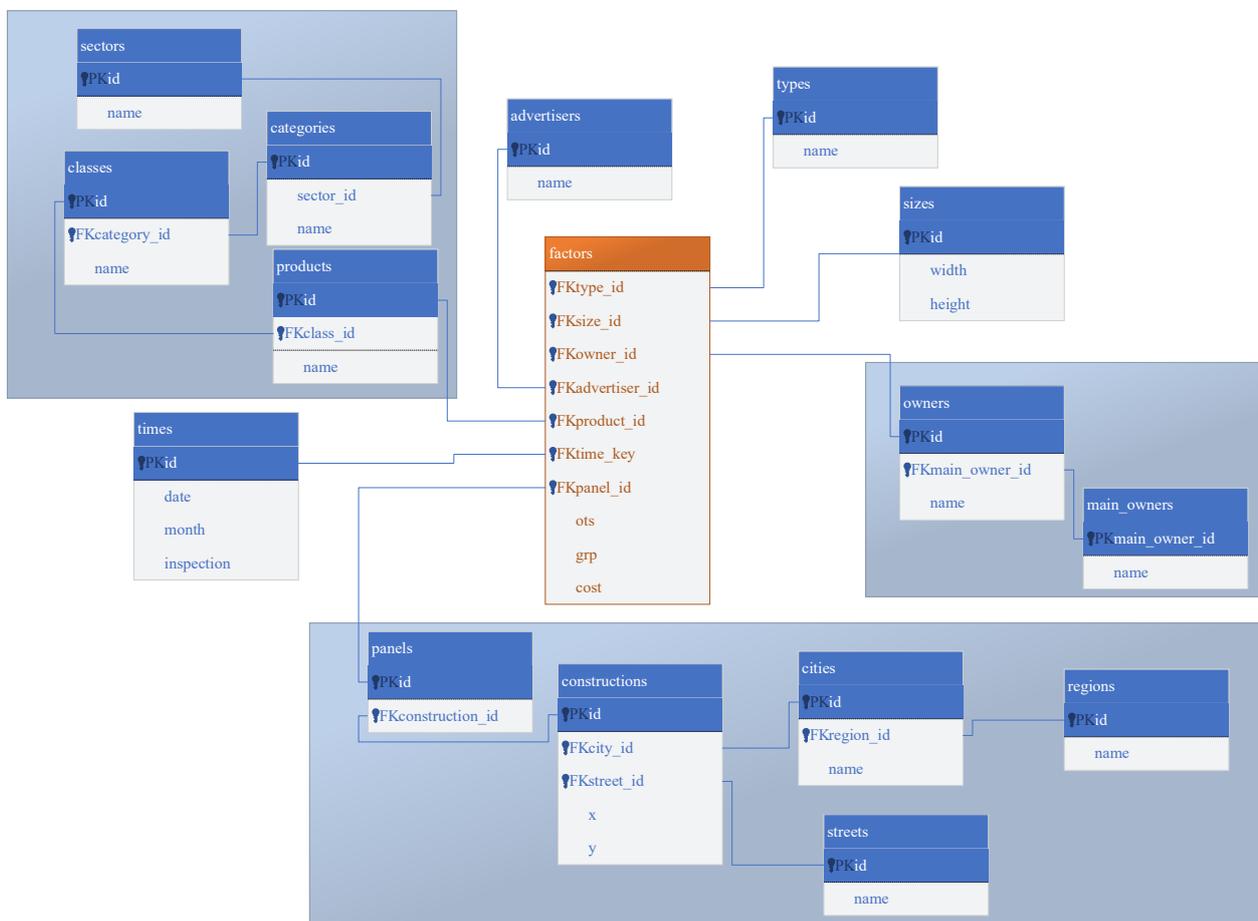


Figure 3. Data warehouse structure

Traditionally, ETL systems have been used to transfer information from outdated versions of information systems to new ones. Currently, ETL process is increasingly used to transfer data from ODS to DW.

To start ETL process, you need to extract data from our databases and prepare them for the conversion

phase. The method of obtaining information by auxiliary software directly from the database is used.

There are the following advantages of this method of obtaining information:

- no need to expand the OLTP system;
- data can be deleted based on the needs of the transfer process.

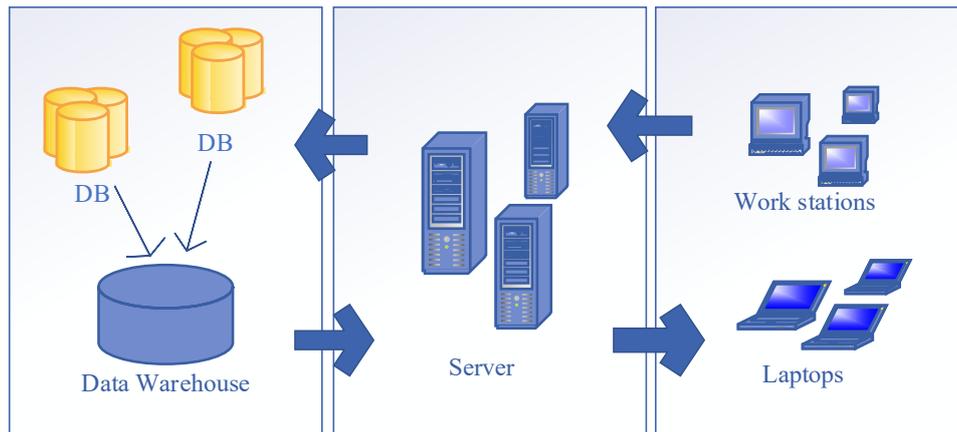


Figure 4. ROLAP DSS outdoor advertising system

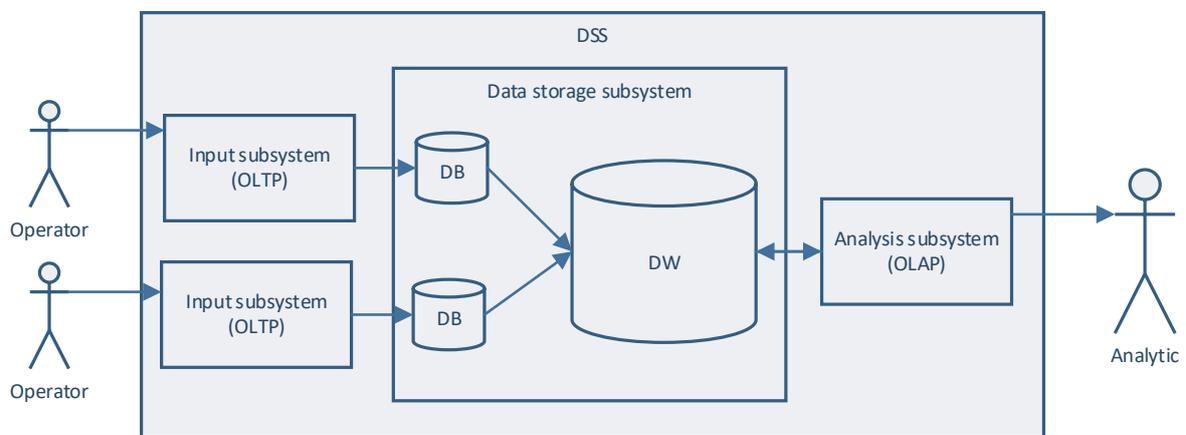


Figure 5. ETL processes

In the publication (Shelest, 2020), the intellectual analysis of the data of outdoor advertising by means of Power BI is presented.

With the help of Excel, it is possible to analyze the data (Figure 6).

Based on the data extracted from DW for certain filters, we can observe the rate of OTS panels of outdoor advertising on the graph (Figure 7). We see that in May 2020, the OTS rate in Kyiv is much higher than in other cities.

Each type of panel has its own mathematical model for calculating OTS.

OTS – Opportunity to See.

### 5. Conclusion

DSS solves three main tasks: collection, storage and analysis of stored information. At the same time there are: information retrieval analysis, operational analytical analysis and intellectual analysis.

The concept of data warehouses is used to unite OLTP subsystems and analysis subsystems

within one system. There is a database for OLTP subsystems and a database for analysis.

DW allows you to integrate information that reflects different points of view on one subject area.

Data in DW after download is read-only. This allows you to significantly increase the speed of access to data, both due to the possible redundancy of stored information, and by eliminating modification operations.

When loading information from ODS to DW, the data is filtered. Many of them do not fall into DW because they are meaningless in terms of use in the analysis procedures.

The concept of DW provides for the separation of data storage structures for operational processing and execution of analytical queries. This allows one DSS to combine two subsystems that meet conflicting requirements.

The advantages of ROLAP are the ability to work with existing relational databases, more economical use of resources and greater flexibility in adding new dimensions.

Year	(All)						
Sum OTS	Month						
City	Apr	Feb	Jan	Jul	Jun	Mar	May
Borispol	84300	475513	489676	73594	80760	504153	52204
Vinnytsia	37260	380393	403744	96534	23190	404860	15748
Dnieper	546780	1496487	1642907	156829	568920	1645635	615629
Zhitomir	114780	444686	472843	91605	58470	473091	42222
Zaporizhzhia	53280	1076683	1148333	92225	38070	1155866	15686
Ivano-Frankivsk	37740	539545	577065	54250	25440	259656	24459
Kamenskoe		144652	154752	372		155434	
Kiev	2192495,15	8020508,95	8518982,32	3324735,9	2492190,17	8010252,21	1582633,47
Kiev oblast	22740	24882	26598	496	2610	23498	2108
Krivoy Rog	14400	624341	669011	22909	14760	655185	8339
Kropyvnytsky	50940	386657	413323	34937	23190	406162	17329
Lutsk	30180	235857	252123	38750	26550	252123	18197
Lviv	536070	2325800	2480248	758942	532080	2483782	305071
Nikolaev	137490	1112324	1198460	152458	70650	1171862	46035
Odessa	343380	3103522	3306553	737056	590070	3294959	193967
Poltava	23490	608275	650349	50716	10050	648923	11036
Rovno	36750	425169	456382	58435	26370	451050	19778
Summy	85410	377870	404488	101370	68040	404302	34968
Ternopil	25020	421051	449562	59303	25170	450058	12276
Uzhgorod	55860	342113	367939	30845	19290	368528	9517
Kharkov	554340	1941115	2044915	814897	519630	2055083	458645
Kherson	47670	899783	965402	100285	33120	979197	22506
Khmelnytsky	64830	499003	531836	67766	50250	526535	51336
Cherkasy	27420	339271	363134	66371	21660	364498	12462
Chernihiv	19620	271469	288672	91295	41220	287711	16988
Chernivtsi	40380	500395	544143	65379	30540	540547	38006

Figure 6. Data obtained from DW in Excel

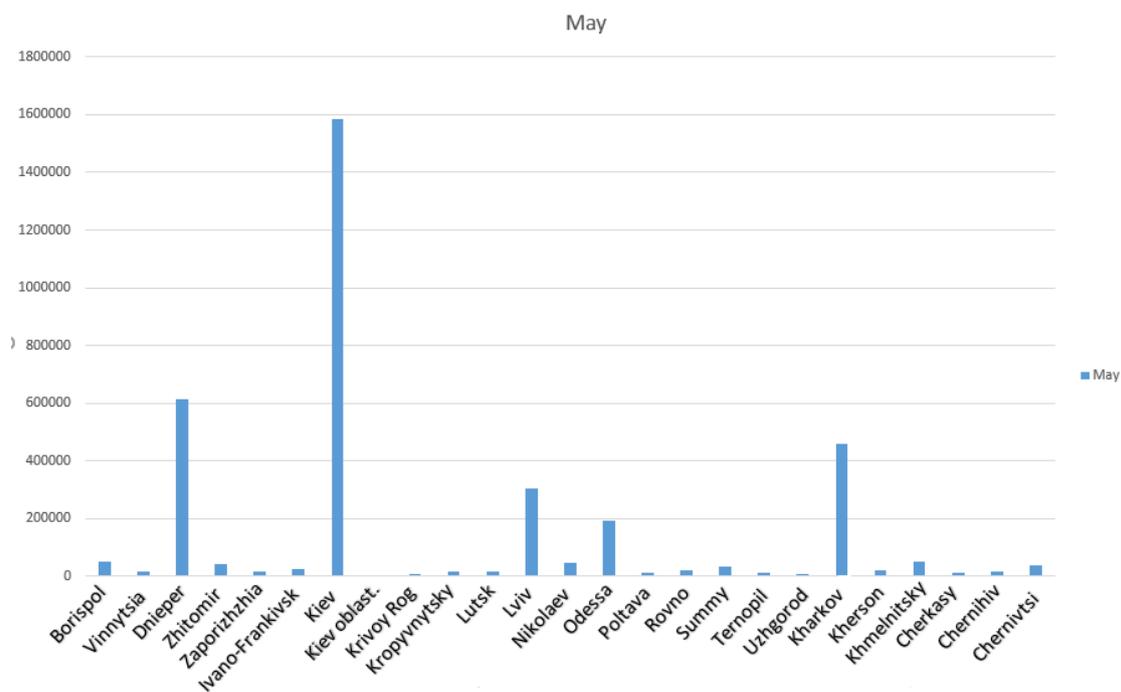


Figure 7. Chart of indicators of OTS panels for May 2020

## References:

- Barsehyan, A. A., Kupryyanov, M. S., Kholod, Y. Y., Tess, M. D., & Elyzarov, S. Y. (2009). *Analiz danykh protsessov: ucheb. posobie* [Analysis of data and processes: Textbook]. St. Petersburg: BKHV-Peterburh. (in Russian)
- Shelest, O. V., & Holub, B. L. (2019). *Rozrobka informatsiynoi tekhnolohii intelektualnoho analizu danykh zovnishnoi reklamy* [Development of information technology for data mining of outdoor advertising]. (PhD Thesis). Kyiv: National University of Life and Environmental Sciences of Ukraine. (in Ukrainian)
- Shelest, O. V. (2019). *Dodatkovy moduli u systemi analizu danykh zovnishnoi reklamy "Analitik" ta "Infopanel" dlia analizu tsyfrovoykh panelei* [Additional modules in the system of analysis of outdoor advertising data "Analitik" and "Infopanel" for analysis of digital panels]. (PhD Thesis). Kyiv: National University of Life and Environmental Sciences of Ukraine. (in Ukrainian)
- Serheiev-Horchynskiy, O. O., & Ishchenko, H. V. (2018). *Intelektualnyi analiz danykh, kompiuternyi praktykum* [Intellectual data analysis computer practicum]. Kyiv: National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute". (in Ukrainian)
- Shelest, O. V. (2020). *Intelektualnyi analiz danykh zovnishnoi reklamy zasobamy Power BI* [Intellectual analysis of outdoor advertising data by Power BI]. (PhD Thesis). Kyiv: National University of Life and Environmental Sciences of Ukraine. (in Ukrainian)
- Overview of Online Analytical Processing (OLAP). Available at: <https://support.office.com/en-us/article/Overview-of-Online-Analytical-Processing-OLAP-15d2cdde-f70b-4277-b009-ed732b75fdd6?ui=en-US&rs=en-US&ad=US>
- Upravlinskyi analiz: website. Available at: [https://studme.com.ua/1228112810027/ekonomika/metody\\_intellektualnogo\\_analiza\\_danyh.htm](https://studme.com.ua/1228112810027/ekonomika/metody_intellektualnogo_analiza_danyh.htm)
- Online analytical processing (OLAP). Available at: <https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/online-analytical-processing>
- DOORS Consulting: website. Available at: <http://www.doors-c.com.ua/>
- Power BI. Available at: <https://powerbi.microsoft.com/ru-ru/>