DOI: 10.37943/AITU.2022.53.85.008

A. Mukasheva
PhD, Associate Professor of the Department of Information Systems and Cybersecurity
a.mukasheva@aues.kz, orcid.org/0000-0001-9890-4910
Almaty University of Power Engineering and Telecommunications, Kazakhstan

D. Yedilkhan
PhD, Associate Professor of the Department of Computer Engineering
yedilkhan@gmail.com, orcid.org/0000-0002-6343-5277
Astana IT University, Kazakhstan

M. Aldiyar
Master of Technical Sciences, Leading specialist of the Software Development Department
aldiyarmyrzabek@gmail.com, orcid.org/0000-0002-2895-5197
Almaty University of Power Engineering and Telecommunications, Kazakhstan

MULTIDIMENSIONAL DATABASES IN INFORMATION SYSTEMS OF UNIVERSITIES

Abstract. The article is devoted to the description of the method of multidimensional database, which is an effective method of data storage, which allows analyzing data qualitatively, and most importantly in a short time. The article discusses the capabilities of multidimensional databases, in particular, multidimensional OLAP (On-Line Analytical Processing) cubes when analyzing large amounts of data. Provides an overview and features of a multidimensional database and discusses the steps you need to take with a multidimensional database to understand the structure and capabilities of an OLAP cube. To create a knowledge base, it describes the steps you can take to create and execute a multidimensional database that you can collect from various sources, save to a database, and then prepare a report using OLAP analysis. Various information system data processing technologies such as OLTP and OLAP were considered. The algorithm of the data storage process for analysis purposes was studied. A model of a multidimensional database in the form of a three-dimensional cube was presented. Examples of analysis and ways of obtaining information from the data cube were also given. The use of a multidimensional database in higher education institutions as a simple and effective method of data storage is considered. There are also illustrations of the structure of a higher educational institution to see the bulkiness of information, and what kind of information database operates in the educational institution.

Keywords: Information system, Database, OLAP, OLTP, three-dimensional, one-dimensional, data analysis.

Introduction

The development of information technologies is directly proportional to the current development of society. Humanity copes with the daily growth of information through information technologies, in particular, information systems that are able to store data, analyze and submit information when requested in a very short time, which simplifies work and saves

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 9, MARCH 2022

our precious time. Another factor is that improving data warehouses is one of the factors for effective management of an organization. For this reason, the information system (IS) is used in all areas of human activity [1]. And the field of education is no exception. Since higher education institutions have so many divisions (individual buildings, departments, research institutes, libraries, rectorates, administrative units, etc.), there is more than enough information to use information databases. This research will examine the structures, schemas, and purposes of multidimensional databases, as well as the use of these databases in higher education institutions.

If the basis for solving many problems is data processing, then information systems can facilitate this processing process. There is a data bank and a database [2]. All incoming data is stored in the IC cores, and this data can be reorganized into one or more databases. If a person opens a shelf of books, they can see the number and condition of these books. If this person is looking for something specific that has a connection to his or her subject area, then if available, he or she will write down or pull out of the shelf everything that he or she needs. In this case, the shelf is the memory of the computing system, not the database. A data bank is just a type of IS consisting of a database, a database management system, a data dictionary, an administrator, a computer system, and maintenance personnel.

Before dealing with a database, you need to create it, maintain it, and make it available to many users so that they can work together with the database. For this purpose, there is such a thing as a database management system (DBMS) [2], which is a set of language and software tools.

The IS can be used to process data or use it in more complex tasks, such as statistical processing of data accumulated over a certain amount of time, for decision-making purposes. In the first case, the data can be processed using the online transaction processing technology (OLTP technology) [2]. A transaction is a single operation associated with a database. A transaction is considered as a single whole action, i.e. it is either fully executed or not executed. The second system that deals with difficult tasks is called operational analytical processing (OLAP). The main advantage of the OLAP system is its speed and quality of analysis. To achieve such a high level, OLAP uses powerful multiprocessor techniques, sophisticated analysis methods, and specialized data warehouses. Figure 1 shows an illustration where you can compare the structures of OLTP and OLAP technologies.

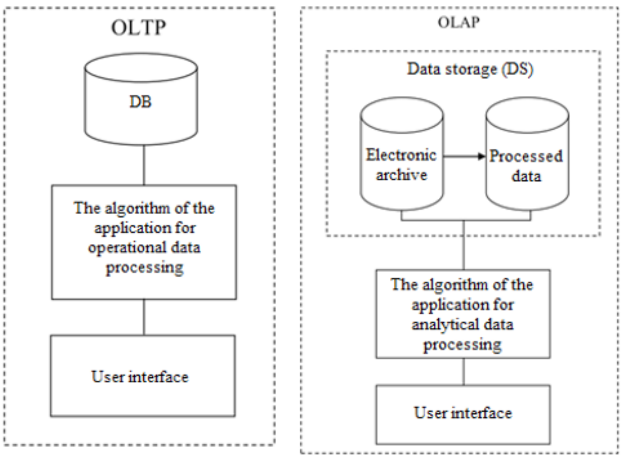

Figure 1. Structures of different information system technologies [1]

The OLAP structure can be represented as a cube [3], consisting of working data. The cube consists of many tables, which in turn contains key facts that are used to make queries. The data in these cubes is stored as aggregates, not as individual metrics. So for example, if you

DOI: 10.37943/AITU.2022.53.85.008
© A. Mukasheva, D. Yedilkhan, M. Aldiyar

imagine a cube of student grades for one semester, then one calculated grade result is stored in one cell, and not the number of all grades received in different cells. Let's take another example: if we imagine that a store's annual sales report is being prepared, then each cell will contain information for the month. Because the analysis will be difficult to count the results of daily sales. And that would take a lot longer. To simplify the task, the cube stores aggregates of results.

Materials and methods

Today, OLAP is the most common method of data analysis, which is based on two approaches [4]:

– Multidimensional OLAP;

– Relational OLAP.

The first approach is implemented on the basis of a multidimensional database, while the second approach is based on structured queries of a relational DBMS.

An OLAP cube can be considered as a logical model for representing multidimensional data characterized by indicators and measurements: G = <D, F> - hypercube (Figure 2).

F = $<f_1, f_2 \dots f_n>$ - hypercube indicators (measurements): each indicator has a set of values that quantify the analyzed process.

D = $<d_1, d_2, \dots d_m>$ - hypercube dimensions: each dimension is an ordered set of values of a certain type. Dimensions can be organized into an ordered hierarchical structure. Hypercube axes are created from a set of dimensions:

$$d_1 = d_1^1, d_1^2, d_1^{k_1}$$

$$d_2 = d_2^1, d_2^2, d_2^{k_2}$$

$$\dots\dots\dots\dots\dots\dots..$$

$$d_n = d_n^1, d_n^2, d_n^{k_n}$$

One of the main requirements of the technology OLAP is "transparency": a ready-made multidimensional cube should be presented in a user-friendly form, the cube controls should be intuitive, the names of the objects of analysis should correspond to terminology. The OLAP analysis process is characterized by the following set of actions with multidimensional data-consolidation, aggregation (grouping), cutting and rotation. The formation of a part of the cube consists in determining the value (values) of a certain dimension in which the size of the cube decreases. Part of the cube is the inner cube containing all the other dimensions. The rotation operation consists in changing the direction of the axes of the cube.

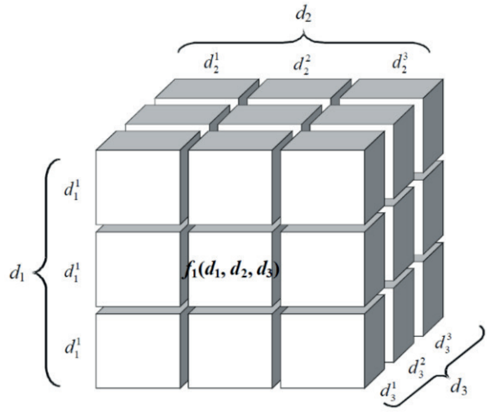

Figure 2. OLAP-cube

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 9, MARCH 2022

High flexibility in decision-making is provided for the end-user due to the possibility of changing the resulting appearance of the OLAP cube. The analyst is not given a strictly regulated report, but the opportunity to use a set of tools for creative research of the problem. Free data manipulation makes it easy to get the required data set. An OLAP cube can be considered as an abstract representation of a selected subset of a relational database.

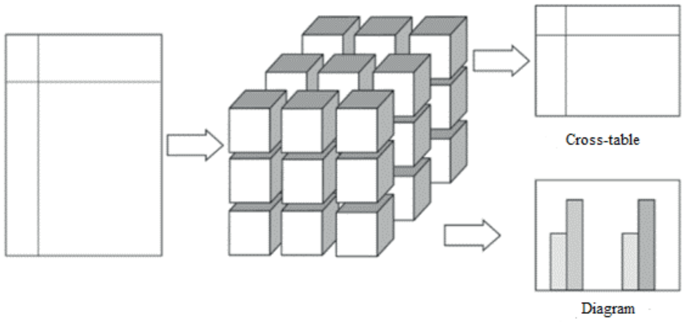

Figure 3. Operation diagram of a simple OLAP system [5]

You can build an OLAP algorithm based on Figure 3 above:

1. Getting data as a flat table or as the result of a structured query;
2. Caching data and converting it to a multidimensional cube;
3. Displaying the constructed cube using a cross-table or diagram [6].

Data warehouses are an integral part of any organization. And in the case of a university that has multiple buildings, it can store data centrally or in separate buildings. However, the data obtained from various structural elements of the institution are not unified, often contradictory, and the indicators used for analysis and management cannot be obtained directly from them. To avoid this, you need to use data warehouses as a specialized source for analytical processing of information. At the stage of data collection and integration, this allows you to combine data and unify it [7].

Although all the prepared information is stored in the data warehouse, processing methods depend on specific tasks. The multidimensional model is capable of fast data analysis, but it does not allow storing large amounts of information. In fact, the relational model has no restrictions on the volume of accumulated data, but it has a low speed of performing analytical queries [8].

The question arises whether these two approaches can be combined. First, there is rarely an operation that requires all the information stored in the repository for analysis. Usually, each analyst serves one of the business areas of the organization. The real volume of this data allows you to store it in multidimensional storage. The source for them should be the organization's central repository. Multidimensional storages play the role of small warehouses. A data kiosk is a specialized multidimensional storage facility that serves one of the company's business lines [9].

A multi-dimensional model is a cube whose main components are a dimension and a cell.

A dimension is a set of data of the same type. The most commonly used time dimensions are days, months, quarters, years, or geographical dimensions: cities, districts, regions, and countries. In the multidimensional model, dimensions play the role of indices [10].

A cell is a field whose value is uniquely determined by a fixed set of dimensions. The field type is usually numeric. In the figure below, you can clearly see the different number of students enrolled in different specialties during three consecutive academic years. Analyze and summarize for further action.

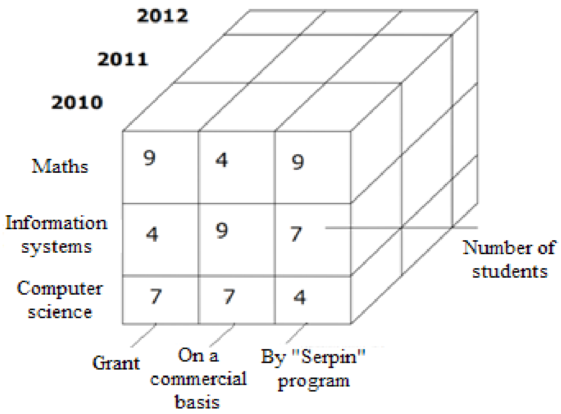

Figure 4. Three-dimensional data model

A multidimensional database is a database that has multiple dimensions. When all three intersect, we can find out the amount of a product in a particular store on a particular date. Multidimensional data structures such as k-d trees, quadrant trees, and range trees are comparison-based data structures. Weighted trees and multidimensional trees are related to each other [11, 12].

For analysis: first, we can get a single square from a cube. Here it is very important to understand that 7 is an aggregate that has already been calculated when entering it into cubes [4]. This unit consists of: Aimaganbetov, Temirbekova, Serikov, Smagulov and others. The method of obtaining a square from a cube is shown in Figure 5.

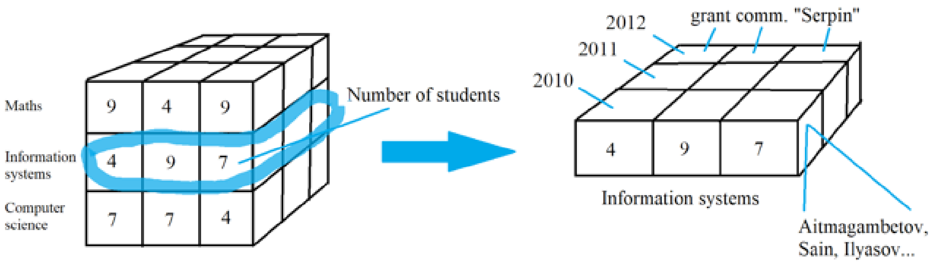

Figure 5. Visual representation of obtaining a two-dimensional measurement of data from a three-dimensional model

The second approach to get information from this cube is to look at only one characteristic, for example, the number of graduates in the field of information technology in 2011. At the intersection of the necessary data, we pull out a one-dimensional array from the cube. Figure 6 shows a method for obtaining a one-dimensional data array from a three-dimensional one.

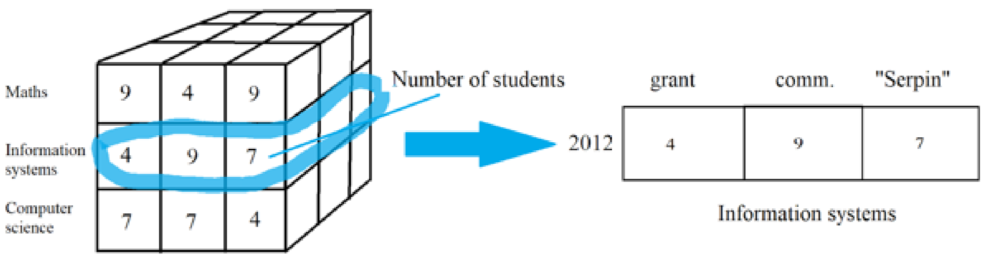

Figure 6. Visual representation of obtaining a one-dimensional measurement of data from a three-dimensional model

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 9, MARCH 2022

The numbers that we see as the number of students are aggregates. Aggregation occurs when information is entered in cubes. Here the question may arise: "why cube, why three-dimensional dimension?". You can create four-dimensional or more dimensions, but increasing the dimension will create a problem for you. To do this, I will give an example in a three-dimensional cube: if there was no enrollment of students in the specialty "Computer Science" on a grant basis in 2010-2012. These cube cells, which are a one-dimensional array, will be empty, but filled with zeros. As the dimension increases, the volume of the array will also increase. For example, it can become a square or cube of empty cells. For this reason, several cubes are used, which are filled with certain values, depending on which aggregates you may need. Data cleaning and validation are important steps in any data analysis, since the reliability of the conclusions drawn from the analysis depend on the quality of the input data. Data errors can occur for a variety of reasons, including erroneous encodings, faulty measuring equipment, and inconsistent data creation guidelines [13].

To use the multidimensional model in higher education institutions, you must first look at the structure of the university according to Figure 7.

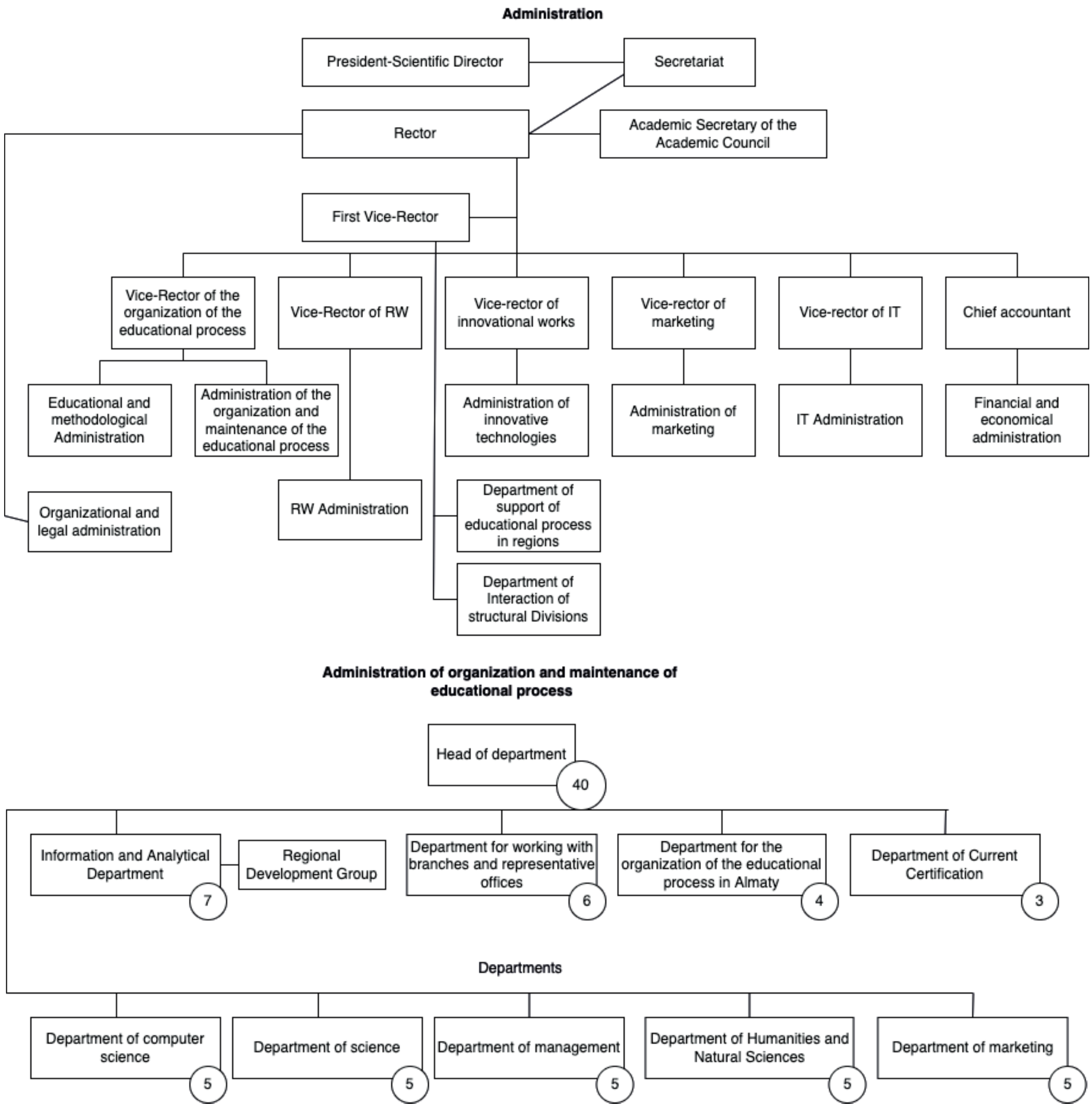

Figure 7. General structure of the educational institution and the educational process [15]

DOI: 10.37943/AITU.2022.53.85.008

If we talk about a holistic structure, then before the departments there are also faculties. For clarity, let's consider a multidimensional cube for the Faculty of Physics and Technology according to Figure 8. This cube stores the data of the teaching staff. We can say that the Faculty of Physics and Technology has three departments, which employ a certain number of professors, doctoral students, senior teachers, teachers, as well as the number of hours they lead, and, consequently, what salary they receive for the number of hours allocated to them.

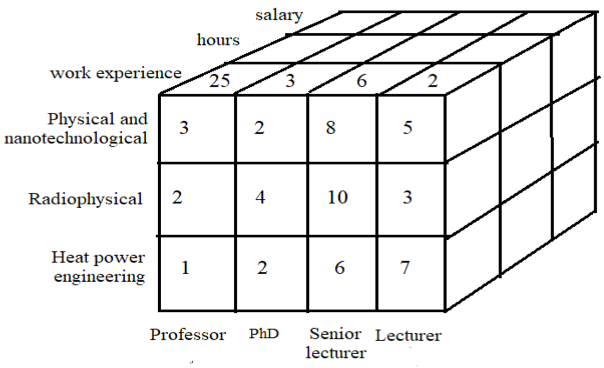

Figure 8. Multidimensional cube for the Faculty of Physics and Technology

If we want to do an analysis for all faculties, then the overall view will look like several cubes for individual faculties. Each branch and sub-division of this holistic structure has its own multidimensional cubes that allow you to analyze data efficiently and quickly. All this information is connected to the central data warehouse, as mentioned above, for the integrity and unification of all data [16]. The main reason for this is the fact that all information belongs to one educational institution.

A higher education institution is essentially a firm that generates material, physical, and intellectual resources and transforms them into the final product – knowledge and skills developed by university graduates. Digital transformation is the way forward for all structures. Technologies are developing at a rapid pace, and structures need to adapt to changes not only to take advantage of the huge opportunities they provide, but also in order to remain relevant in this world of instability, uncertainty, complexity and ambiguity [17]. The educational process is a complex process that consists of a large number of disparate subsystems and generally has no formal description. A significant recent technological development concerns the automation of work with knowledge and services as a result of advances in artificial intelligence and its subdomains. To describe this phenomenon, the authors used the term "intelligent automation". This development provides organizations with a new strategic opportunity to add business value. However, the academic research that examines these developments is scattered across a wide range of scientific disciplines, which leads to a lack of consensus on the main conclusions and implications. The authors conducted the first interdisciplinary literature review systematically characterizing the intellectual state and development of intelligent automation technologies in the field of knowledge and services.

The first is the conceptualization of intelligent automation and related technologies. Secondly, a business value-based model of intelligent automation for working with knowledge and services is proposed and twelve gaps in research are identified that prevent a full understanding of the process of implementing business value. Thirdly, a research program is proposed to address these gaps. The advent of using big data opens up a new wave of relationship management strategies in support of personalization and customization of

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 9, MARCH 2022

services. Big data requires new tools and methods for its collection, storage and analysis and is used to improve the decision-making process to improve management. The effectiveness of the functioning of the higher education system is determined by the efficiency of higher educational institutions. Using IS by universities leads to digital transformation. Concept of digital transformation considers as a process in which digital technologies create disruptions that cause strategic reactions from organizations that seek to change their ways of creating value, while simultaneously managing structural changes and organizational barriers that affect the positive and negative side results of this process [14].

Discussion

A systematic approach allows analyzing the development and functioning of the educational system, taking into account the requirements of the state, society and citizens, both in the short and long term. A systematic approach should be applied in every area and part of the educational process. Like the "blockchain" technique, where one circle with teeth drives the movement of another circle, and so on, the work of each department, department, administrative unit, and others as separate circles must resonate with each other in order to simply not stop. In this case, it is ifnormation systems that allow you to conduct university business by visiting each department of an educational institution online. As mentioned above, using centralized data storage, which means having access to all the components, you can analyze the work of each circle and compare them to output the final results.

In practice, when implementing a multidimensional model, the arguments in favor of its performance are confirmed by tests showing that multidimensional databases perform operations at least an order of magnitude better than relational databases for data queries. For example, in the case when it takes several minutes to process data in a relational database, a multidimensional database copes with the same operation in just a few seconds. However, at the moment at the moment, many enterprises still use analytical systems based on classical relational databases. The performance advantages present in the implementation of a database using a multidimensional model help facilitate the development of interactive decision support systems, when implemented in a relational environment, performance is often unprofitable. Despite the fact that most of the data manipulation actions that are possible it can be implemented using a multidimensional data model, it can also be implemented in a classical relational database, a multidimensional model can offer us some advantages. Among them: simplicity of data presentation and display – data is stored and displayed in an intuitive way to the user; great flexibility in building reports; and finally, last but not least – performance. The use of a multidimensional data model increases the speed of analyzing large amounts of information at times, in comparison with a relational database. Often, analysts face this problem when using a classical relational database for analysis. After all, in order to make a conclusion about a hypothesis, the analyst needs to get aggregated data, and if it takes a lot of time, the analyst's time is also wasted. Thus, the benefits provided by using a multidimensional data model, are critical in many cases. Therefore, enterprises for which analytics is important, i.e., those with large amounts of information on the basis of which the policy of the enterprise should be based, which currently use classical relational databases, should consider switching to the use of storage.

Conclusion

In conclusion, the authors note that a higher education institution is essentially a firm that generates material, physical, and intellectual resources and transforms them into the final product – knowledge and skills developed by university graduates. The educational process is a complex process that consists of a large number of disparate subsystems and generally has no formal description. The effectiveness of the functioning of the higher education system is determined by the efficiency of higher educational institutions. Using IS by universities leads to digital transformation. A systematic approach allows analyzing the development and functioning of the educational system, taking into account the requirements of the state, society and citizens, both in the short and long term. A systematic approach should be applied in every area and part of the educational process. Like the «blockchain» technique, where one circle with teeth drives the movement of another circle, and so on, the work of each department, department, administrative unit, and others as separate circles must resonate with each other in order to simply not stop. In this case, it is information systems that allow you to conduct university business by visiting each department of an educational institution online. As mentioned above, using centralized data storage, which means having access to all the components, you can analyze the work of each circle and compare them to output the final results. We have considered various data processing technologies of the information system, such as OLTP and OLAP. The algorithm of the data storage process for analysis purposes was studied. A model of a multidimensional database in the form of a three-dimensional cube was presented. Examples of analysis and ways of obtaining information from the data cube were also given. The use of a multidimensional database in higher education institutions as a simple and effective method of data storage is considered. There are also illustrations of the structure of a higher educational institution to see the bulkiness of information, and what a huge database operates in an educational institution. Summing up, it can be concluded that the use of a multidimensional database is an effective method of data storage, which allows analyzing data qualitatively, and most importantly in a short time.

References

1. Karpova I.P. (2009) *Osnovy bazy dannyh. Uchebnoe posobie.* Moskva: Moscow State Institute of Electronics and Mathematics (Technical University).
2. Palominos F. E., Córdova, F., Durán, C., & Nuñez, B. (2020). A simpler and semantic multidimensional database query language to facilitate access to information in decision-making. *International journal of computers communications and control. 15*(4). https://doi.org/10.15837/ijccc.2020.4.3900
3. Orlova M.A. (2021). Konceptual'noe proektirovanie mnogomernoj bazy dannyh importera produkcii. E-Scio, *11* (62), 99–105.
4. Blagodatskiy G.A. (2020). *Multidimensional database analysis.* Izhevsk, Russia: Izhevsk State Technical University named after M. T. Kalashnikov.
5. Cabibbo, L., & Torlone, R. (1998). *A logical approach to multidimensional databases.* Lecture Notes in Computer Science, 183–197. https://doi.org/10.1007/bfb0100985
6. Cabibbo, L., & Torlone, R. (1998). *Querying multidimensional databases.* Database Programming Languages, 319–335. https://doi.org/10.1007/3-540-64823-2_18
7. Shahi, C., & Sinha, M. (2020). Digital Transformation: Challenges faced by organizations and their potential solutions. *International Journal of Innovation Science*, *13*(1), 17–33. https://doi.org/10.1108/ijis-09-2020-0157
8. Iyengar, S. S., Rao, N. S. V., Kashyap, R. L., & Vaishnavi, V. K. (1988). Multidimensional Data Structures: Review and outlook. *Advances in Computers*, 69–119. https://doi.org/10.1016/s0065-2458(08)60257-0
9. Khatwani, G., & Kar, A. K. (2017). Improving the cosine consistency index for the analytic hierarchy process for solving multi-criteria decision making problems. *Applied Computing and Informatics*, *13*(2), 118–129. https://doi.org/10.1016/j.aci.2016.05.001
10. Petersen, A. H., & Ekstrøm, C. T. (2019). Your assistant for documenting supervised data quality screening in R. Journal of Statistical Software, *90*(6). https://doi.org/10.18637/jss.v090.i06

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 9, MARCH 2022

11. Liu Y.A., Stoller S.D. Knowledge of Uncertain Worlds: Programming with Logical Constraints.- Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11972, (pp.111-127).
12. Jennex, M. E. (2017). Big Data, the internet of things, and the revised knowledge pyramid. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 48*(4), 69–79. https://doi.org/10.1145/3158421.3158427
13. Zhang, Y., Lin, G., Gu, H., Zhuang, F., & Wei, G. (2020). Multi-copy dynamic cloud data auditing model based on IMB Tree. *Enterprise Information Systems, 15*(2), 248–269. https://doi.org/10.1080/17517575.2020.1812004
14. Arnas, D., & Rodríguez, M. (2020). Range searching in multidimensional databases using navigation metadata. *Applied Mathematics and Computation, 386*, 125510. https://doi.org/10.1016/j.amc.2020.125510
15. Anshari, M., Almunawar, M. N., Lim, S.A., & Al-Mudimigh, A. (2019). Customer relationship management and Big Data enabled: Personalization & Customization Of Services. *Applied Computing and Informatics, 15*(2), 94–101. https://doi.org/10.1016/j.aci.2018.05.004
16. Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems, 28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003
17. Roberts, N., Qahri-Saremi, H., & Vijayasarathy, L. R. (2021). Understanding IT value at the managerial level: Managerial ambidexterity, seizing opportunities, and the moderating role of information systems use. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 52*(3), 39–55. https://doi.org/10.1145/3481629.3481633