ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 1 , January 2021

Distributed Information Systems and Reliability Assurance

NishonboyevTuygun, Alimjon Dadamuxamedov

TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES NAMED AFTER MUHAMMAD AL-KHWARIZMI, Doctor of Technical Sciences, Professor Department of "Data Transmission Networks and Systems"Tashkent, Uzbekistan
INTERNATIONAL ISLAMIC ACADEMY OF UZBEKISTAN, Senior teacher of the " Department of Modern ICT",Tashkent, Uzbekistan

ABSTRACT: The article suggests the use of replication methods to increase the reliability and optimal presentation of data stored in distributed information systems. To do this, a modeling algorithm and software based on public service blinds were developed, and the effectiveness of the method was demonstrated through a computational experiment.

KEYWORDS: distributed information system, database, replication, query intensity, telecommunication network, datacenter,

I. INTRODUCTION

E.Tanenbaum in his book "Distributed Systems. The fundamental monograph "Ideas and Principles" [1] describes the distributed system as follows:"A distributed system is an interconnected set of independent computers using communication channels, which is viewed from the user's point of view as a single system based on special software."

The main purpose of a distributed system is to organize the resources of computers, software and applications in a single system. Resources include computing, memory, information, and network resources.

The user connects to the distributed system and provides information about the resources he needs. The required resource may consist of a set of resources located in several parts of the world. The system finds these resources, aggregates them, forms a "composite" resource, and presents it to the user as a single resource.Today, distributed systems are in the form of data processing centers based on Grid and Cloud Computing technologies, consisting of high-speed and large-capacity computer systems, as well as complex structured telecommunications networks and other modern hardware and software. [2,3].

At the heart of the idea of a distributed system based on grid technology is the creation of a new type of computer infrastructure. It addresses the global integration of distributed information and computing resources based on network technologies, as well as specialized software and standard services, and provides access to them to a wide range of users. Grid technologies are created on the top layer of the Internet. [2,3].

Cloud-based TTs are created using the capabilities of the idea of "cloud computing", in which computer resources are provided in the form of Internet services. [2].The user sets certain rules for the resource he needs, the cloud creates the optimal architecture of computer resources to meet the needs of the user and provides it to the user.

Cloud resources include the computing power of various types of computers, memory resources for storing large amounts of data, computing networks, databases, and software, all of which are converted into a single "link" in a heterogeneous or homogeneous environment using cloud computing technology. minutes.

Today, most cloud-based devices introduce virtualization technology, the essence of which is to enable multiple virtual computers to run on a single physical computer.In other words, virtualization allows computer resources to be distributed across multiple independent environments, resulting in the creation of several independent virtual machines within a single server computer.Using a virtual server allows you to place several operating systems and several applications on a single computer. The resources of cloud data processing and storage centers are interconnected on the basis of network technology, which today has a speed of 100 mbit / s and more. [4].

The problem of Distributed Cloud Data Center deployment of information resources in distributed (dislocated) memory repositories and their provision to a wide range of users is becoming increasingly important, as cloud storage infrastructure is currently at different levels for a large group of Internet users. is creating a new architecture that supports services.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 1 , January 2021

The information stored and planned to be stored on the memory disks of the dislocated autonomous data centers within the Distributed cloud data center differs significantly in terms of value and content. There is a greater demand for the same content and less for the same.These distributed databases result in higher loads in some autonomous data centers within the data center and in the same sections of the telecommunications network, and less in others.Therefore, it is important to ensure that the distributed databases are securely stored on dislocated memory disks within the data center and that fast and high-quality information is provided on request.

II. LITERATURE REVIEW

The strategy we used to create the search strings was as follows [2] [11]: • finding papers about distributed information system. • Listing keywords mentioned in primary studies, which we knew about. • Use synonyms word (usage) and sub subjects of network technology in data replications such as (distributed information system, database, replication, query intensity, telecommunication network, and datacenter). • Use the Boolean OR to incorporate alternative spellings and synonyms. • Use the Boolean AND to link the major terms from population, intervention, and outcome. The complete search string initially used for the searching of the literature was as follows: network technology AND data replications. It has been highlighted in [5] [11] that there are two main issues on conducting an SLR search which are the sensitivity and specificity of the search. In our preliminary search, when we used the complete search string defined above we retrieved a very high number of articles. For instance, Google scholar, Scopus, ProQuest education, IEEEXplore, Science Direct, Springer Link retrieved more than two hundred results. Therefore, we have deepened our search and used this search string: (Adoption OR Usage)AND (social network OR database) AND (query intensityOR telecommunication network). The revised search string has given us a reasonable number of studies and we finally selected relevant empirical studies

The main part. Reliability of information storage and functional stability of servers within distributed databases are one of the main factors determining the quality of modern distributed systems.A common way to protect yourself from data loss is to back up these databases. When data is lost from memory for a variety of reasons, the availability of a backup allows the database to be restored quickly.The backup process does not require additional hardware to be added to the system, as backups can be made on the server itself or on external storage media.

At the same time, the information stored in the distributed data center segments of the distributed databases must be delivered to the users within a short period of time. Distributed databases information stored in some segments of data centers may be located at a great distance from the user terminal, or the data transmission capacity of the network media (communication channels) between the segment where the information is stored and the central management system may be low. It takes a lot of time to provide the information required by the request.

The goal is to ensure that the distributed cloud data center reliably stores and delivers high-quality information of different content gathered at different "points" based on the use of replication methods.In distributed systems, the replication method allows you to copy information from one storage location to another, and then synchronize its contents [5,6]. The method can refer to multiple memory disks, rather than a single memory space, to process all queries. This allows one memory disk load to be distributed over several (Figure 1).

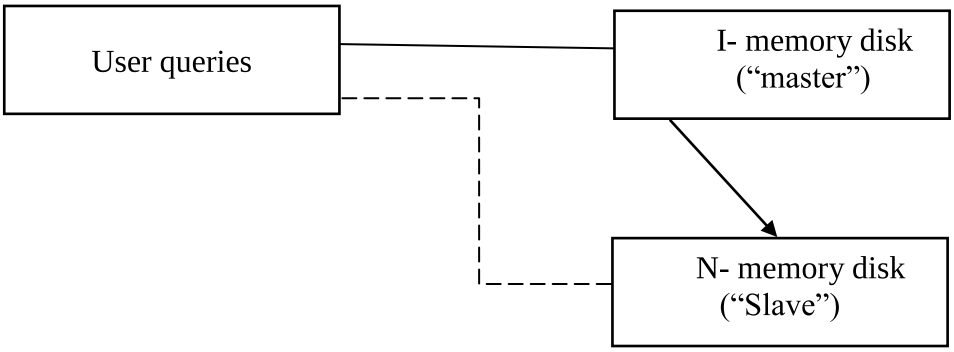

Figure 1. A query diagram for allocating queries to other memory disks. (July 2018 to June 2020)

There are basically two approaches to working based on the replication method:

- “Master-slave”
- “Master-master”.

In the first approach, a single primary database memory disk is allocated, which is called a master. All data changes (any MySQL INSERT / update / delete queries) are performed on it. In the master-slave method, the master always copies all changes.Select requests are sent from the application to the slave server. Thus, the Master server is responsible for the data changes, and the slave is responsible for the readings.

The advantage of this method of replication is that the user can use multiple "slave" memory disks.The method is often used to copy data. In this case, the Master Memory Disk Management System (or server computer) handles all system requests. The "slave" server operates in passive mode. However, when the Master fails, all operations are performed on the "Slave" server.

The ability to read and write data in the Master-Master replication method is available on both sides (Figure 2).

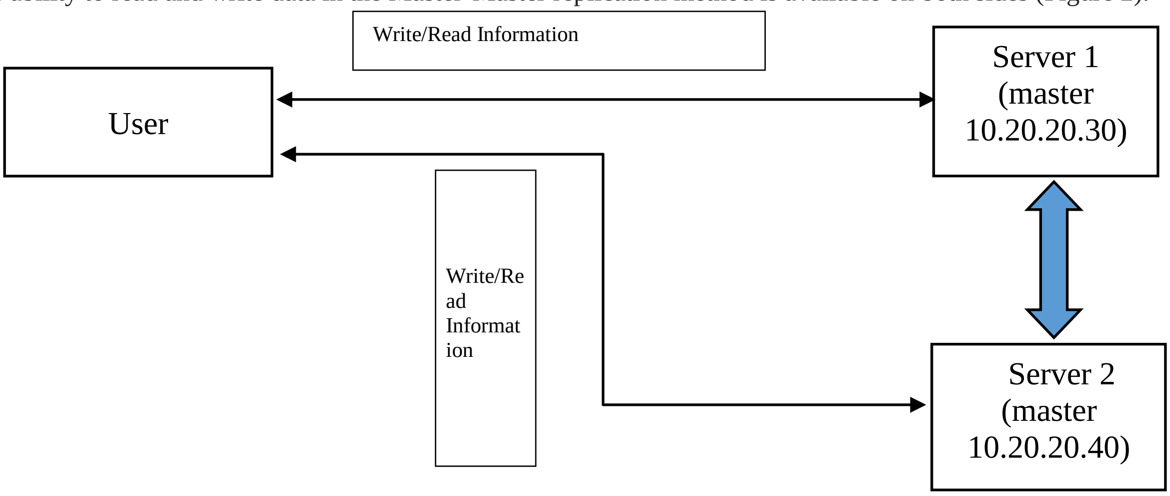

Figure 2. Master-Master replication scheme. (July 2018 to June 2020)

The following is a description of a model based on multi-channel public service theory to determine the effectiveness of the replication methoda[7].

According to him, Requests from n users to access the management system in a distributed management data center intensity$\lambda_i$the sum $\lambda = \sum_{j=1}^{n} \lambda_j$. is represented by the average values μof service intensity in the distributed management data center system.

The average values of the input currents and the service intensities in the system are expressed in units of time as follows

$$\lambda = \frac{1}{\overline{t_{ax}}},\ \mu = \frac{1}{\overline{t_{ovc}}},$$

The following conditions must be met in order for a public service system to operate stably, in a stationary state:

$$\rho = \frac{\lambda}{\mu} < 1,$$

The average number of orders in the system and the average time T in the system are calculated on the basis of the following formulas:

$$\overline{N} = \frac{p}{1-p}, T = \frac{1/\mu}{1-p},$$

The average time of ordering in the d system of a multi-channel system is calculated from the average time of waiting in the queue as well as from the average time of service delivery:

$$\overline{T_d} = \overline{s_d} + \overline{t_d}. \quad (1)$$

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 1 , January 2021

If there are r orders waiting to be serviced at the time the order arrives, then the service for newly received orders will not start before the previous r_d orders are serviced. When a new order arrives, the average waiting time in the queue is proportional to the average number of orders in the queue:

$$\overline{s_d} = \overline{r_d}\,\overline{t_d}, \quad (1)$$

$\overline{r_d}$ parameter value $\rho_d = \lambda_d/\mu_d$determined by the ratio. The flow rate for a service device is proportional to the number of service channels:

$\mu = v_d\, \mu_d,$

here v_d - number of service channels and $\mu_d$the service intensity of each channel.

Intensity of requests to a multi-channel (Public service system) with λ and its loading ratewill be equal to$\rho = \lambda/ v_d\, \mu_d$. According to this expression, the bandwidth of the system increases with the number of service channels.

Based on the above, the time of dispatching a request in a multi-channel Public service system can be expressed as follows

$$\overline{T_d} = \overline{s_d} + \overline{t_d} = \left(\overline{r_d} + 1\right)\overline{t_d} = \frac{\overline{r_d + 1}}{\mu_d} \quad (2)$$

Intensity of requests to each element of the system (flow rate) $\lambda_d = \lambda\, p_d$($p_d$ – the intensity of the current flowing into the system input λ a parameter that indicates the portion directed by the dispatcher to channel d will be $\sum p_d = 1$).

Dispatcher channel functions are limited to queuing requests to the appropriate service item.The time of the request in the multi-channel Public service system is determined by the sum of dispatching and service time:

$$T = T_d + T_0.$$

Order dispatch time is determined based on formula (2).After dispatching, the average order time in the system is calculated by the following formula:

$$\overline{T_0} = M[T_0] = \sum_{i=0}^{n} p_i \overline{T_i} = \sum_{i=0}^{n} p_i \frac{\overline{r_i}+1}{\mu_i}$$

where pi is the probability that the i-element of the public service system will be serviced after dispatching.Based on the above:

$$\bar{T} = \overline{T_d} + \overline{T_0} = \frac{r_d + 1}{\mu_d} + \sum_{i=0}^{N} p_i \frac{\overline{r_i} + 1}{\mu_i}$$

The last expression allows you to calculate the average time it takes to place orders in a multi-channel Public service system.The following is a description of the algorithm for placing and presenting information in segments of autonomous systems based on the method of replication in autonomous system memory warehouses in order to ensure the reliability of information storage processes in distributed information system segments (autonomous storage systems) and increase the efficiency of providing information to a wide range of users.

1. Initially, the autonomous centers of the Distributed cloud data centern contained information of different types and content stored in memory ombor {Vm}, the size of the "empty" fields in memory{ΔVm},the bandwidth of virtual channels between the central management server and the autonomous information centers is determined, (The distributed information system is based on the telecommunication network, which is based on the interconnection between the central control server and the autonomous memory areas, which are based on the serial connection of several switches with communication channels. This connection is made on the third network layer of the seven-level OSI model. This direct connection between the central management server and the autonomous storage space is called a virtual channel. The bandwidth of a virtual channel is equal to the bandwidth of a communication channel with the smallest bandwidth it contains.) its value μm (m = 1.2,…..M) is a parameter of the intensity of service of requests of the public service system[7].

2. The intensity of requests to the central management server of the information system, which is distributed to the information stored in the memory warehouses of autonomous information systems, is determined$\gamma_{1,m}$, based on their analysis, the value $\gamma_{2,m}$ of the intensity of requests on the access of virtual channels of the telecommunications network and storage disks of information storage and their loading rate$\rho_1 = \gamma_{1m}/\mu_{1m}$ $\rho_2 = \gamma_{2m}/\mu_{2m}$is calculated. In general, this process is done by calculating the bandwidth of the communication channels in the virtual channel, as well as the level of loading of memory disks with requests for other types of information.However, since the volume of information

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 1 , January 2021

requests at the input of virtual channel and memory disks is approximately the same and the processing speed of computer memory disks is several times higher than the virtual channel data transfer rate, it is also expected to perform the replication process based only on the value of ρ_1 the result will be achievable

3. A certain threshold value of the load level of the virtual channel is determined $\rho_{1,y}$, on the basis of which the segments of the distributed information system are divided into 2 sets: $\{M_1\}$- $\{\rho_1\}>\rho_ч$the number of segment sets whose condition is met $\{M_1 \subset M\}$, the information stored on their memory disks is replicated; $\{M_2\}$-$\{\rho_1\}<= \rho_y$the number of segment sets whose condition is met $\{M_2 \subset M\}$, the information stored on their memory disks is not copied,$M = \{M_1\}+\{M_2\}$.

4.Based on the above$\{M_1\}$and$\{M_2\}$ are arranged according to the value of the loads on the input of the autonomous information systems in the packages, that is $\rho_{1,\{1\}}>\rho_{1,\{2\}}>\ldots> \rho_{1,\{M1\}}and\rho_{1,\{1\}}>\rho_{1,\{2\}}>\ldots> \rho_{1,\{M2\}}and\{M_1\}$The replication process starts from the largest load segment in the package: $\{M_2\}$ the smallest loaded segment in the package to the memory disk $\{M_1\}$a copy of the information stored in the largest loaded segment of the set is written, and a certain part of the flow directed to it is directed to the smallest loaded segment of the set $\{M_2\}$, i.e.

$$\gamma_{\{1\}}^{\{M1\}} = \gamma_{\{1\}}^{\{M1\}} - \Delta_{\{1\}}^{\{M1\}}$$

$$\gamma_{\{1\}}^{\{M2\}} = \gamma_{\{2\}}^{\{M2\}} + \Delta_{\{1\}}^{\{M1\}}$$

in this $\Delta_{\{1\}}^{\{M1\}}$ - $\{M_1\}$of the stream directed to the segment in the set$\{M_2\}$the part of the set that is transferred to the segment; $\{1\}index$the information being copied is stored $\{M_1\}$the sequence number of the segment in the set; $\{2\}$index - being replicated $\{M_2\}$the sequence number of the segment in the set.

This process is performed for information stored in all segments of the package {M1}, depending on the situation, the information stored in one segment is written to several segments of memory in the package {M2} and vice versa to a single memory space of the package {M2} several types of information can be replicated.

5. - n + 1; the distributed information system simulates the process of entering the next request into the management system: it is serviced, queued, or the request is lost if there is no space in the buffer;

6. Determines the information resources specified in the current query in which segments of memory disks, the degree of loading of memory disks, the bandwidth of virtual channels of the telecommunications network connected to them;

7.An information resource with a minimum value of two indicators finds the stored segment and redirects the current request to it through a defined virtual channel and calculates the average value of time spent on the request in the dispatcher block;

8. Model the current query service process in the selected segment and calculate the average value of the service time and send control to 9 operators;

9. Calculates the total time spent on the next request and the cost of servicing the flour;

10. Checks that all requests are considered, if the condition is not met, passes control to 5 operators, otherwise calculates the final results, publishes and stops the algorithm.Based on the algorithm, software was created in the Paython programming environment and a computational experiment was performed based on it.It is assumed that the distributed system consists of 11 autonomous information centers, which are directly connected to the dispatch system using virtual channels of the telecommunications network.The systematization and service of queries are modeled according to the law of exponential distribution. The given parameters are the intensity of incoming request flows, the intensity of service of the request in the virtual channels and autonomous components of the distributed information system, the probability of distribution of the request depending on the type of information, etc.

Software-based query distribution processes were modeled without the use of a replication method and on the application of distributed information system components, the results of which are summarized in Figure 3.4.

Figure 3 shows the results obtained based on the distribution of requests for the five types of information stored in the 11 components of the Distributed Information System based on replication and traditional methods.When the system load level is relatively low ($\rho = 0,1 \div 0,4$), the total time T of the requests in the system is almost unchanged,however, starting from r = 0.5, it was observed that the total sum of the time T in the system of queries in the replication method decreased significantly compared to the traditional method (from 4.22 to 3.89).A similar situation was observed in Figure 4 at the value of the probability of loss of queries in the system: at $\rho = 0.5$ the value of P_y decreased from 0.1 to 0.07, and at $\rho = 0.8$ it decreased from 0.254 to 0.092.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 1 , January 2021

III.CONCLUSION

Thus, the use of replication techniques is effective in the optimal presentation of data stored at different "points" of distributed information systems.At the same time, replication increases the reliability of information - when a physical server or memory space fails, a copy of the data is stored in other memory areas of the Distributed Information System and is not lost, and when accessed, the data is provided from another memory area.

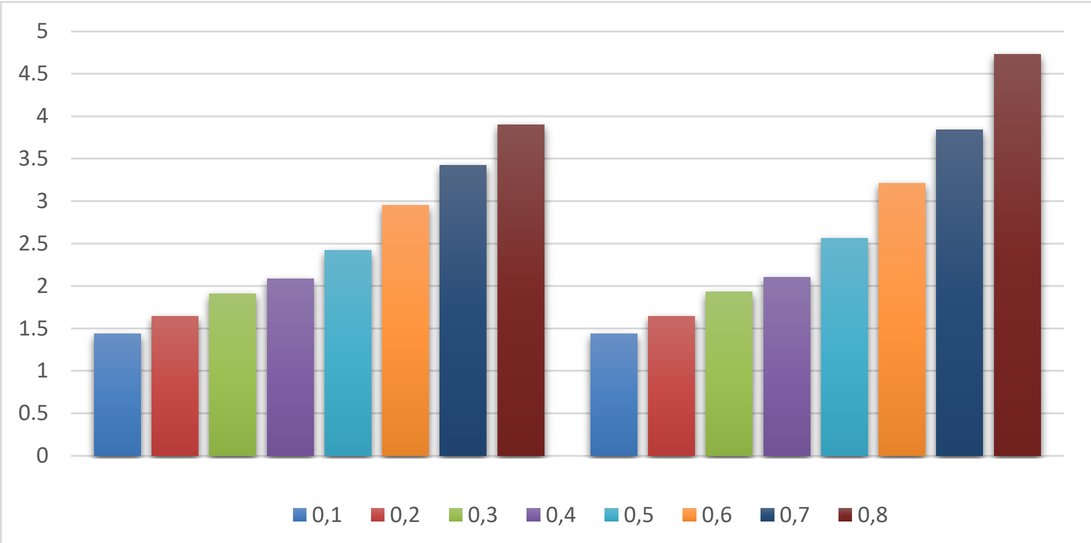

Figure 3. Diagram showing the effectiveness of query replication in the Distributed Information System (DIS) (on the left).

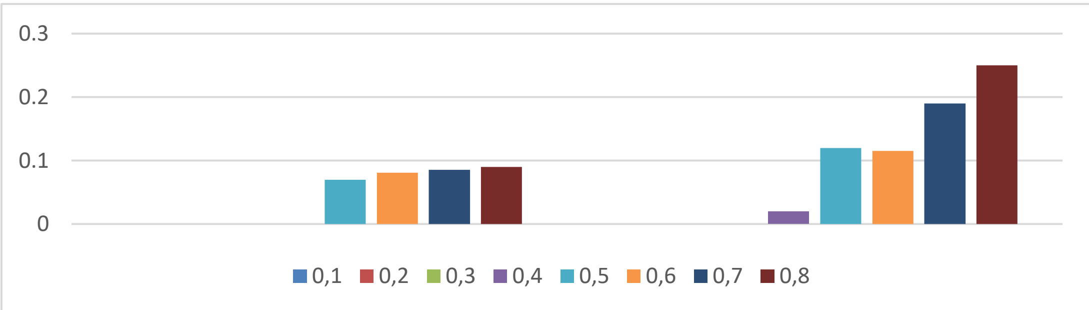

Figure 4. A diagram showing that the use of the replication method (left) in the Distributed Information System (DIS) reduces the probability of losing queries compared to the traditional method.

The application of this method in the creation of a single distributed system of religious information is of great scientific and practical importance.One of the most pressing issues today is the formation of a single centralized database of religious institutions operating in different regions of the Republic of Uzbekistan under the leadership of the Islamic Academy of Uzbekistan.Religious information with "deep" content, collected in religious institutions in the regions of the country, is created (and is being created) on the basis of various programming languages and management systems.Due to the lack of a single system, there are major problems in providing them to a wide range of users. Therefore, the goal is to create a distributed information system based on a single central management using replication methods.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 1 , January 2021

REFERENCES

1.Tanenbaum E, Van Steen M. Distributed systems. Principles and Paradigms) St. Petersburg: Peter, 2003, 877 p.

2. Nishonboev.T. Service-oriented distributed systems (monograph). Science and Technology Development Center Printing House 2017 (p. 300).

3. Nishonboev.T. "Software configured networks". Textbook. (Griffiths). TUIT Printing House named after Muhammad al-Khwarizmi. 2017 (186 p.)

4. Nishanbayev, T.N., Abdullayev, M.M., Maxmudov, S.O.The model of forming the structure of the 'cloud' data center. International Conference on Information Science and Communications Technologies: Applications, Trends and Opportunities, ICISCT 2019

5. Belousov V.E .. Data replication algorithms in distributed information processing systems.http://diss.rsl.ru/diss/05/0591/050591031pdf.

6.Belousov V, E. Features of building a queuing system within the middle level of a distributed information processing system "// System analysis, control and information processing: scientific and technical, collection of articles: 2005, 1, - Penza, PSU, 2005, -p. 23-31

7.Basharin G.P, et al. Analysis of queues in computing networks: theory and methods of calculation / G.P. Basharin, P.P. Bocharov, Ya.N. Kogan. - M .: Nauka, 1989.-336 p.

8. Irgashevich, D. A. (2019). Development of national network and corporate networks (in the case of Tas-IX network). International Journal of Human Computing Studies, 1(1), 1-5.

9. Irgashevich, D.A. (2020). Development of national network (tas-ix). ACADEMICIA: An International Multidisciplinary Research Journal, 10(5), 144-151.

10.Dadamuxamedov, A.I. (2017). Development of the national network and corporate network (on the example of the Tas-IX network). Actual scientific research in the modern world, (3-2), 133-137.

11. Dadamuxamedov, I. A. (2020). Cloud technologies in islamic education institutions. TheLightofIslam, 2 (23).