Investigation of methods location and organization distributed access to data cloud storage system of distance learning*

Irina Bolodurina, Denis Parfenov

Dept. of Applied Mathematics ,Orenburg State University, 13 Pobedy ave., Orenburg, Russia

prmat@mail.osu.ru, fdot_it@mail.osu.ru

Abstract. *Every year we are seeing an increase in the number of information systems and services designed to provide interactive services for Internet users. With this you can bind a significant increase in the flow of information to be processed and data transfer over the network. Currently, for these purposes, as a rule, used rented computing power, located in the data center (DC). In this case, the actual problem is the research and development of algorithms and methods of performance management and optimization of hardware and software resources provided for the operation of services. In the framework of the present study we have constructed a model of storing and organizing shared access to data using the cloud platform used to provide access to multimedia educational resources deployed in the distance learning system.*

Keywords

Distance learning, multimedia educational resources, cloud computing, efficient access, load distribution.

1 Introduction

Now due to the need to build a scalable solutions for high-loaded services using the cloud model resources received wide distribution. This approach is very effective for the organization indefinitely extensible web services, as well as services via scalable repository for access to large volumes of data. One aspect of these services is the creation of multimedia resources to broadcast the video in real time and on demand. The most widely used of these resources received during training using distance learning technologies. The main advantages of cloud computing are:

- unified access - users have access not only to the data, but also to the resource as a whole, which is very important in the construction of highly loaded applications;
- flexibility - the consumer of resources alone can determine and change the computing needs, such as server time, access speed and data processing, the required amount of data to be placed;
- pooling of resources - the ability to consolidate multiple services using the exact same hardware base, while managing the distributed computing power between consumers in terms of the dynamic of demand resources.

Moreover, one of the advantages of using cloud is a possibility to choose architecture (infrastructure as a service, platform as a service, program as service, etc.) for the organization of access from the outside to the shared resource pools. However, this, in turn, imposes some limitations on ability to control and use of shared computing power of the hardware platform. The most common use of cloud at the moment - it is a platform as a service. In addition to unlimited scalability, this architecture allows users to control the flow of requests, not only at the application level. Users can use network and hardware level too.

Despite the universal architecture of clouds, some services require a special approach. Thus, in the considered area (educational multimedia resources) there are some of factors that can affect the performance of the resource:

*Authors thank for support the Ministry of Education and Science of Russian Federation (projects 14.B37.21.1881 and 14.132.21.1801 by the federal target program "Scientific and scientific-pedagogical personnel of innovative Russia")

- The load on the servers is periodic. The access to multiple resources with different types of content occurs simultaneously. In most cases, existing equipment cannot serve all customers without the use of load balancing.

- Up to 90% of the load is predetermined, since access to the resources used by pre-registration (subscription to services). Using standard tools can not account for a predetermined load and allocate it to the limited resources.

In addition, it is worth noting that 80% of the resources in demand is only 20% of the time. Just as we have found that a single point of aggregation is the system storage (NAS), which provides the processing of requests from consumers for multimedia educational services. Consequently, the efficiency of multimedia service, as well as the quality of services provided depends on the performance of the data warehouse.

This study aims at improving the efficiency of storage as elements of cloud services through the development of algorithms for scheduling cells and migration management in distributed data nodes.

2 Theoretical Part

A key differentiator of multimedia data storage is heterogeneity placed information (text, audio, video or data), and as a consequence of different approaches to provide access to it. When using a cloud service sends the user to the application are placed in a global queue handled by the controller. The controller uses the queue inherent in it a set of rules of service requests and distributes the application on the compute nodes (figure 1).

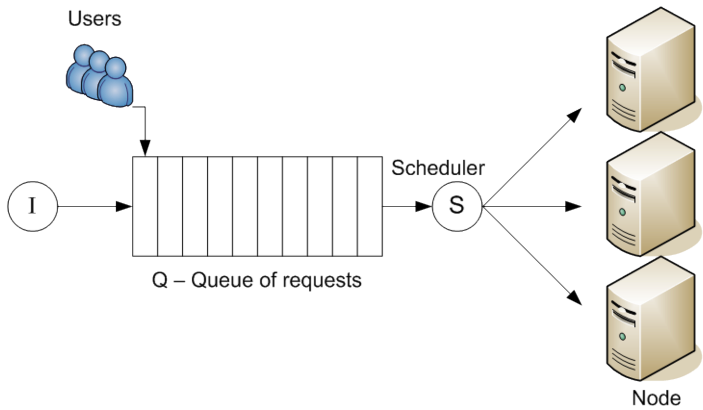

Fig. 1. Scheme query in the cloud

In addition to data access methods is essential to the intensity of treatment to the different elements, which can be obtained with the use of intra-user identification algorithms, which in turn allows us to estimate and forecast the demand load on the storage system. In this regard, an important aspect of management system, with a significant increase in downstream applications is competently arrange the placement and distribution of data elements for devices [2,3]. In this case, the parameters of assessment include: the waiting time in the queue and the execution time of the application.

A distinctive feature of cloud storage is the reconfiguration of the structure depending on the resources consumed. This in turn enables you to implement optimization algorithms in terms of placement of data within the disk space, and manage change in the number of the system's devices. The process of optimizing the placement should not lead to a decrease in the quality of customer service storage, this must be considered in the algorithms of network capacity and the maximum amount of data that can be transmitted at one time. [4] Additionally, you must take into account the current load on the devices themselves, as well as their location relative to each other and to clients that connect to them.

Given the specificity of the multimedia service, and using statistics of receipt of applications in the system based on the typical model of customer behavior (the user), we have developed an algorithm to optimize the data access mechanism. Depending on the nature of the query identifies the following conditions of the algorithm:

- access to dynamic (streaming) data (audio and video);

- access to static data (text, documents, archives).

To optimize data access mechanisms necessary to construct a general model of user requirements for the storage system. Suppose R=(U,M,Q), where $U = \{u_1,u_2,\ldots\}$ – set of users; $M = \{m_1,m_2,\ldots\}$ – set of unique data elements to be placed in storage devices. The minimum unit of data we assume m_i file with mandatory property h - size.

To ensure safe storage and load balancing between storage devices define the distribution function of data elements for this we introduce the set M_c={ $m_1^{j_1}, m_1^{j_2}, m_1^{j_3}, \ldots, m_2^{j_1}, m_2^{j_2}, m_2^{j_3}, \ldots$}, where $m_i^{j_k}$ - k-th placed a copy of the data element (m_i) on j_k storage device, provided $k \geq 3$.

T Then the distribution of data items for storage devices becomes $P:M_c \rightarrow D$.

Based on the above, we write the user requirements for data items.$Q:U \rightarrow X \subseteq M_c$, where X – set of data requested by set of users U. Then, data storage can be written as cortege $S=(M_c,D,P,L,C,R,G)$, where

$D=\{d_1,d_2,\ldots\}$ – set of device storage;

$L=\{l_1,l_2,\ldots\}$ – set of values which characterizes load of each storage device (number of simultaneous user requests to a specific device);

$C=\{c_1,c_2,\ldots\}$ – set of values that characterizes the amount of each of the devices in the store;

$G \in N$ – natural factor specifying the geographical (topological) priority use a storage.

Typically, for large cloud structures used consolidated storage, consisting of farms spanning multiple repositories in a single array. We represent it as $S_{farm}=\{S_1,S_2,\ldots\}$.

Since the performance of user requirements change over time, transform cortege requirements $R(t)=(U,M_c,Q(t))$. Then $Q(t):U \rightarrow X \subseteq M_c$ – user requirements for data elements that change over time. Because in addition to the activity of users to edit the repository, write cortege storage depending on the time $S(t)=(M_c(t),D(t),P(t),L(t),C,R(t),G)$, где

$D(t)=\{d_1,d_2,\ldots\}$ - set of storage devices, which change over time, such that $\forall t$, $D(t)>0$;

$P(t):M_c \rightarrow D$ – distribution function of the data items to a storage device, a time-varying.

At the same time to optimize the cost of the hardware resources and reduce device being used simultaneously introduce cortege relations $S_{cloud}(t)=\{S(t),D(t),D_{use}(t)\}$, где $\forall t$, $D_{use}(t) \subseteq D(t)$ Set of storage devices used in a scalable storage S at time t. In addition, scaling storage and data migration, the condition $\forall t$, i,j $i \neq j \Rightarrow D_i(t) \cap D_j(t)=0$, when migrating data store should not use the same device. This will ensure that as the speed of processing information, and ensure reasonable reconfiguration.

Thus, to minimize the number of simultaneously used storage devices in a single, scalable storage, and maximize the number of requests processed per unit of time users, we introduce an objective function of the form:

$$\sum_{i=1}^{N} P_i(t) \rightarrow \min;$$

$$\sum_{i=1}^{N} L_i P_i(t) R_i(t) \rightarrow \max.$$

As a rule, to spread the load and improve performance scalable repository, duplication and transfer data between devices in the same system used cache areas (arrays of devices that permit rapid processing of read / write), built with solid-state SSD drives or high volume memory [5]. However, the algorithms and methods of use of such resources are not adequate. Most often, the device cache of filled most popular data is not used in the model user behavior. Typically, when accessing multimedia service client sends multiple requests sequentially to obtain data. As part of an educational multimedia service can predict a set of requested data and the order they are received, allowing you to build the forecast and make a reservation of computing resources to solve the problem.

Using the above, an algorithm to optimize the user's access to multimeydiynym data.

Step 1. Receiving input parameters.

When registering a new request, allocated nodes (storage, D), containing the necessary data and analyze their load (L), and the geographical priority over client (G). Determine the type (static, dynamic) and the rating of demand requested data, based on statistics of requests.

Step 2. Request processing.

For static data using the received data in step 1 (G, L), determines the optimal node.

For dynamic data is to find the necessary data item in the cache area. If it is not found, the procedure is performed with optimal data caching node, obtained with the use of indicators (G, L).

Furthermore, using the search algorithm relations, taking into account the demand for ratings of resources formed a list of items that can be requested by the client in the near future. In addition to an efficient algorithm, also looks the least loaded nodes in the system containing the necessary data, which in turn allows you to partially insulate the process from the main cache of operations performed by the system. Using this data, performed the procedure cache. The number of elements depends on the relevance of the initial data and the total load on the system.

Step 3. Data transfer. Requested at the current time data is sent to the user from the selected source.

Step 4. Post-processing of results. At the end of the algorithm stored in the database is updated resource demand ratings used in the processing elements.

The proposed algorithm and the requirements model as the basis of the developed load-balancing mechanism and element distribution data storage devices. Simulation model developed by the storage system allows the study to establish the dependence of the performance of the distribution of data in the storage. To evaluate the effectiveness of the algorithm we used to simulate the system with various options. One of the commonly used models is the initial placement of data in a single copy on a minimum number of devices. At the same for all devices in the research process, a limitation on the maximum capacity L. The effectiveness of the algorithm is evaluated by comparing the length of the queue of the total number of applications simultaneously in the system, and the number of dropped requests. The modeling revealed significant congestion nodes, suggesting an inefficient distribution of load (**Fig. 2**).

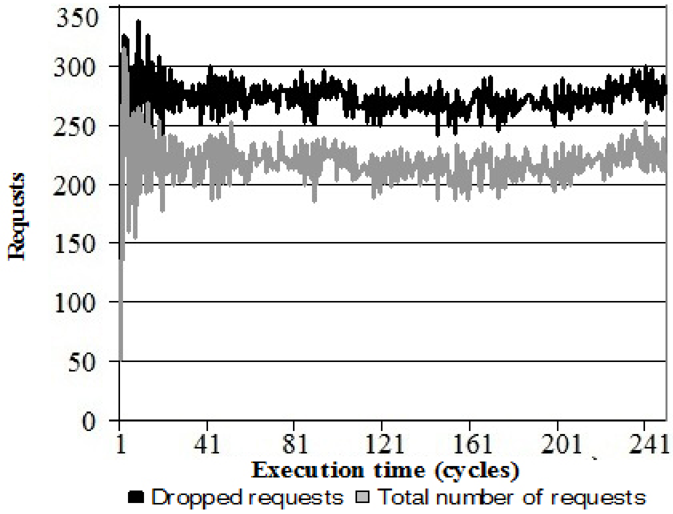

Fig. 2. Queuing diagram for a single copy of data

When operating systems are often highly loaded used method of redundancy (backup) data to ensure the integrity and organization of distributed repositories. The number of copies of each of the elements to be placed may vary, which can cause an imbalance of resource consumption. However this study using simulated storage showed that the increase in the number of copies of items placed data provides performance gains (number of applications at the same time served) on average 1.5 times (см. **Fig. 3**).

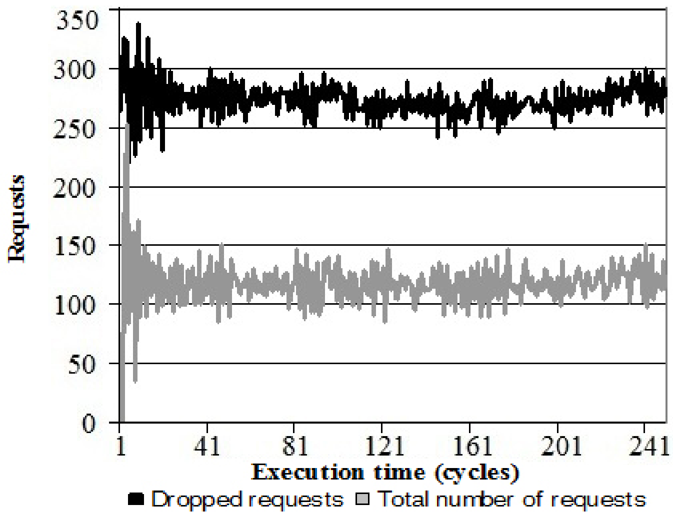

Fig. 3. Diagram applications service with the minimum number of devices with multiple instances of data

However, using this approach to the placement of data the following drawbacks:

- continued scaling of data elements and the positions they physically devices is not always possible;
- some portion of the devices is overloaded, the other is unloaded, which means inefficient use of storage resources.

Thus, this approach contradicts one of the goals we have solved the problem - reducing the use of resources. During the simulation the load on the storage we obtained pattern affecting system performance. Devices that store large volume (h) data remain under load longer than devices that contain data of small size. Most industrial storage analyze only the total volume of space used on the device when choosing IPO data element, which later caused problems during migration and back up items to other devices. To optimize the placement of elements in the data storage system simulator we added balancer, which produces a qualitative assessment of the location of data on the device by the amount of disk space occupied by the data elements as the number of large and small files. The relative evaluation of large and small file size is analyzed both in terms of devices and across storage systems in general. The simulation using the figures found on a minimum number of devices, to obtain additional gains of up to 5-7% (depending on the input data stream, and the total volume of data to be placed), subject to store multiple copies of data (**Fig. 4**).

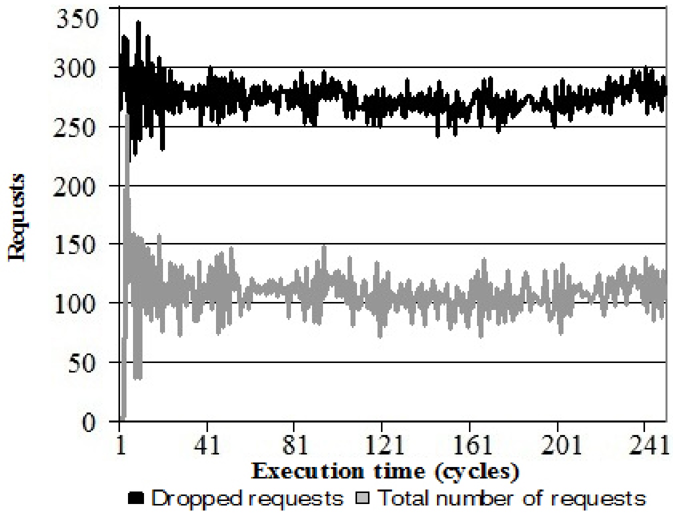

Fig. 4. Diagram applications service with the minimum number of devices with intelligent data placement

It should be noted that the application of these techniques can be complemented proposed in this paper predictive caching algorithm. To assess the effectiveness of the results of his work as a load balancer, storage systems, we conducted a comprehensive simulation using all the methods described above. As a result, a further increase in performance by 4-9% (depending on the method of placing the initial data) (см. **Fig. 5**).

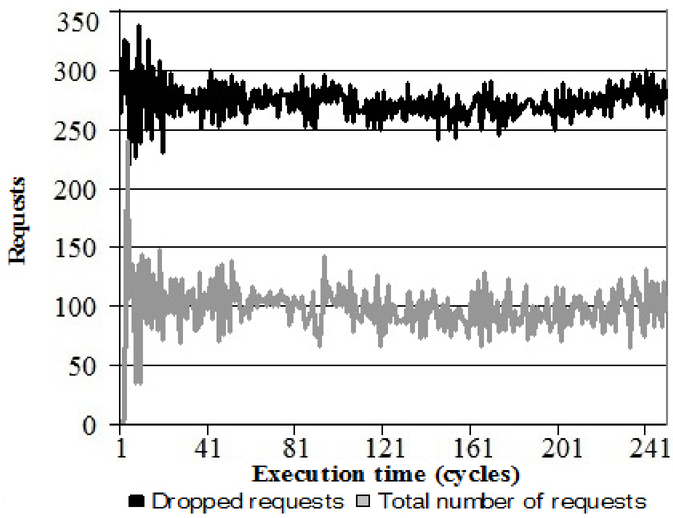

Fig. 5. Diagram applications service with the minimum number of devices with intelligent placement and data caching

3 Conclusion

Thus, assessing the overall performance of the load-balancing algorithm and data placement in the organization of access to educational multimedia resources, you can get the performance boost of 5 to 15% compared with the standard means that is highly effective at high intensity requests. In addition, the resulting objective function and built model can be used to further study the effectiveness of the use of hardware and software resources to improve the quality of service in distributed information systems of distance learning.

4 Acknowledgements

Authors thank for support the Ministry of Education and Science of Russian Federation (projects 14.B37.21.1881 and 14.132.21.1801 by the federal target program "Scientific and scientific-pedagogical personnel of innovative Russia").

References

[1] O.V.Gusev, A.V.Zhukov, V.V.Polyakov, S.V.Polyakov O. Gusev “The problem of an adequate assessment of the performance of Web servers in corporate networks in enterprises PPI”, Proceedings of the 6th Scientific Conference "New Information technology in PPI and energy"- Petrozavodsk, 2004., pp. 84-87.

[2] A.V. Zhukov “Some models of optimal control input of applications in the intranet systems”, Proceedings of the 6th Scientific Conference "New Information technology in PPI and energy"- Petrozavodsk, 2004., pp. 87-90.

[3] D.Parfenov “Video portal as an effective means of information exchange in educational process”, Proc. V International scientific-practical conference "University Infomedia XXI Century", Petrozavodsk, , 2011, pp. 141-144.

[4] I.Bolodurina, D. Parfenov, “Modeling the demands for resources in a distributed information system for remote support of the educational process” in “High technology, the economy, industry”, St. Petersburg, 2012, pp. 30-34.