Research on Key Technology of Distributed Storage in Cloud Computing Environment

Lin Zhu[1], Li Zhuang[2], Caihong Li[3]

[1] Department of Computer Engineering, Southeast University Chengxian College, Nanjing 210088, China

[2] School of electronics and Computer Engineering, Southeast University Chengxian College, Nanjing 210088, China

[3] Southeast University Chengxian College, Nanjing 210088, China

Keywords: cloud computing environment; distributed storage; key technology

Abstract: With the rapid development of network technology, cloud computing plays an important role in information technology. It is a new type of technology that improves efficiency and security in data processing. Distributed storage technology improves the efficiency of data storage and helps to improve the cloud computing. This paper analyzes the background of distributed storage technology under the cloud computing, discusses the composition of distributed storage technology, and introduces the practical application of distributed storage technology in cloud computing environment from different perspectives.

Nowadays, the Internet has developed rapidly and plays a very important role in people's daily life, learning and work. With the development of network technology, the society is progressing, and the speed of information transmission is increasing. Cloud computing technology has emerged with the advent of a large amount of information, which has been widely applied in the information society. In different industries and fields, a large number of information should be calculated and stored. Therefore, in the link of cloud computing, the data center is the key. Now, based on the background of cloud computing, technicians should improve data analysis and control the unlimited data dynamically. Distributed storage technology is widely applied in people's life, reducing energy consumption and cost [1].

1. The generation of distributed storage technology

With the popularization of network technology, it is spread all over the people's work and life, and people have entered the era of Internet and large data. Information level is improving, massive information is surrounded by people every day, but the storage of information is restricted. Because storage medium has some limitations, it can not meet the needs of the masses. In view of this limitation, cloud computing has entered into people's life. Under the background of cloud computing, a large number of information resources can be reintegrated to achieve personalized mass demand.

Cloud computing is generated on the basis of the development of the Internet. It is a common way of computing, expanding the way of resource sharing, and designing a computable computer model. The service provider is fully intelligent in the application, and does not need a lot of intervention, so the integration of all data can be perfected. In the links of accessing and storing network resources, a lot of information is integrated on the basis of network nodes, forming the way of network database, so that it can satisfy different users' needs. In the construction of infrastructure, all kinds of software and storage services are needed. The form of cloud computing is very large, and it has formed a large scale, but the cost of using it is very low. In the new computing model, cloud computing is fully applied to information resources with the help of data from different fields. The common service of hardware and software to cloud computing promotes the development of resources and makes the resources fully utilized and shared. In cloud computing, data center plays a very important role. Hardware is an important part. It has perfect storage capacity and reduces the

cost of data center.

2. Key technology of distributed storage in cloud computing environment

In the high-speed development of Internet technology, the emergence of server equipment has a certain impact on traditional service mode, and the overall benefit of network technology is improved. The server based structure connects the servers through the network card through the connection of the network lines, so that they are in a whole network system, so as to realize the interconnection between servers. In this mode, using scientific data processing mode, we can find corresponding storage information, transmit packets timely, and achieve data sharing in a timely manner. This kind of structure and other structure comparison, the center server structure is relatively simple, without the aid of data transfer data in high speed forwarding equipment can make the whole structure and the circuit connection mode is more direct and simple, effective to prevent the problem of single tree structure. In the operation of the server, the transmission of mass information can be realized through the transmission of data information at the grass-roots level. But the server center structure will also have some problems, if there is compensation in the case of network operation, the load on the server running will be very large, cause the system equipment is not stable operation, also causes the operating costs increase, resulting in equipment overall effect can not fully play. Therefore, in the central structure of servers, the CAM CUBE structure is generally adopted, which realizes three dimensional nodes, and each node will connect to the server to form a three-dimensional network structure [2].

In the central structure, the switch is usually the central structure, and it is a kind of traditional type of distributed storage. With the continuous improvement of information construction, this kind of structure also presents a kind of limitations in the application, and can not achieve the diversification of the structure function. This kind of structure mainly uses switch as the main base point, and effectively connects the corresponding servers, but the server can only pray to store the effect of network information. This type of structure is centered on the switch and forms a tree structure, which is composed of the polymerization layer, the core layer and so on. The structure is not complex, and the way of connection is very simple. But nowadays, under the background of informatization, there are many data, huge data system, limited storage of information in tree structure, unable to effectively store massive information, and unable to satisfy all kinds of users' needs.

Through the analysis of some aspects, the server as the center, the switch as the central structure, it has some limitations. The combination of the two can effectively make up for their shortcomings, and its scalability is effectively played, and routers and network structure are unified. In the network structure, servers and switches realize effective combination, make the function more diverse, better play the router's function, and the network structure is more flexible. In the process of cost construction, the cost of the operation of the equipment is effectively reduced. In extensibility research, this kind of hybrid structure has a certain relationship with the number of server network cards, and its space is limited. In specific aspects of the application, this kind of structure shows a mixed function, based on the high-speed transmission, information transmission in accordance with the order, in order to improve the efficiency and quality of data, make the transmission process more simple, to combine the superiority of various equipment reform.

3. Key technology of distributed storage in cloud computing environment

In the traditional method of data storage, it is necessary to use efficient servers and specialized storage devices to improve the fault-tolerant effect of the technology. But in the specific technical analysis, the operation cost of the equipment will be very high, and the accuracy of the information data storage can not be improved effectively. In the link of cloud computing, it can effectively avoid the problem of data storage failure. In the related literature analysis, the node failure occurs during the operation of the system equipment. In the process of data collection, every 6 hours, there will be a node failure, which leads to the failure of data collection. The accuracy of data acquisition is not

high, resulting in economic loss. Under this problem, we should optimize the distributed storage technology, improve the efficiency of the operation, improve the fault tolerance of the system, and reduce the production of economic loss.

In the traditional distributed storage structure, small scale data can only be processed, and a large amount of cost will be generated. In the background of information technology, the links between countries are very close, the development of computer network is very rapid, and the scale of data is constantly improving. At this time, the information enterprises should face more serious problems, and more economic profits should be obtained by reducing energy consumption and cost [3].

4. Key technology application of distributed storage in cloud computing environment

Under the background of cloud computing, the key technologies of distributed storage will bring different problems in the application process, such as the cost can not be effectively controlled, the scalability is limited, and the fault tolerance rate is low. But with the rapid development of information technology, these problems can be effectively solved in the later stage, which can effectively improve the efficiency of data analysis and promote the development of information field.

Fault-tolerant technology is an important product of the development of the times, and it is also one of the key technologies on the basis of the development of cloud computing. Fault-tolerant technology can improve the reliability of the system and equipment to a certain extent, so that the mass of information can be excavated. Fault-tolerant technology is an important way to improve system stability in the operation of computer system. Even if the error is activated, the corresponding system can provide users with high-quality information service plan. In the related fields and industries, data fault-tolerant technology makes the application of computer technology more flexible. In the information data of honor, we should take practical measures to optimize the storage of resources in combination with specific circumstances. The characteristics and characteristics of fault-tolerant technology are analyzed, which can be divided into two categories, one is erasure code fault tolerance, and the other is the replication fault-tolerant technology.

This kind of technology mainly realizes data transmission by means of channels, and makes data clearer and more data blocks lost in time by encoding, and reduces the corresponding storage space by way of distributed storage. In the process of coding operation, different information can be integrated into it, and the new coded data object will be generated. The amount of storage is reduced by copying information. In the specific application, the cost of erasure code fault tolerance is relatively high. If the information is distorted, it takes a lot of money to repair it. In this case, in the link of erasure code fault tolerance, specific analysis and operation should be carried out to simplify the operation process [4].

In the link of replication fault-tolerant technology, we should create different modules for one data, and distribute them reasonably in different storage nodes. If there is a distortion of a data object, the user can get the data information through the same module. The application of replication fault-tolerant technology is very simple, and its reading efficiency is very high. It is very popular in practical applications. But the replication fault-tolerant technology has some limitations, and each data will form a number of data modules, so it will take up a lot of space.

In the related investigation, the space consumption of the storage system accounts for about 45% of the total consumption of the storage system during the normal operation. In the links of energy saving technology research, energy consumption has aroused people's attention. The way of dynamic control can effectively control the cost of equipment system, and play a certain ecological and environmental benefits. In the background of the use of cloud computing, the size of the underlying data stored in the distributed storage is very large, and the cost of distributed storage will be improved. The cost includes not only the purchase cost of the hardware, but also the IT equipment and power consumption, and the refrigeration equipment is also the key cost. In order to reduce energy consumption, it is necessary to choose the nearest place from the power station, but the greenhouse effect has a negative impact on people's life. Therefore, whether it is the cloud computing supplier to reduce the cost or to pursue higher profits, reduce energy consumption is the

key problem that should be studied in the field of cloud computing.

In the cloud computing environment, distributed storage applications are generally in the large-scale data center, in order to prevent a large consumption of the data center, the most direct way is to reduce the energy consumption of a single computer node, but in reducing energy consumption in the link, will result in poor performance. In order to better improve performance, we need to analyze the relationship between performance and energy consumption. We need to improve the proportion model and the two segment model based on the establishment of a single computer energy consumption model.

The proportional model is a simple and direct model, which mainly analyzes the energy consumption of the central processor, disk and switch. When the equipment is not in the state of operation, the dynamic frequency and voltage adjustment will be adjusted by the proportional model, and energy consumption will be saved. The dynamic load changes will be established to reduce the energy consumption of the central processor. In the two stage model, the energy consumption of the computer is mainly produced in two parts, one is fixed energy consumption, and the other is variable energy consumption. Fixed energy consumption is mainly divided into energy consumption produced by fans, mechanical drives and diodes and other equipment. Variable energy consumption is mainly divided into central processing and operation, and energy consumption is also increasing with the increase of disk speed. The two section model also believes that the load can be generated in the idle state of the computer, and this kind of load can not be ignored. The two segment model can make a more accurate analysis of the computer's specific energy consumption, so that some unnecessary energy consumption is removed[5].

First, it adopts the technology of software energy saving. In the specific practice, the application of software energy saving technology can ensure the overall operation performance of the device, reduce the energy consumption of data storage, and make it more scientific in node management. Under the background of cloud computing, the relevant personnel can analyze the distribution of storage nodes more comprehensively and objectively, improve the optimization design, and effectively reduce energy consumption and achieve effective utilization of resources through automatic analysis.

Secondly, the hardware energy saving technology is adopted. This technology should start from the point of view of hardware and improve the technology of data center and computer as a whole. In the specific application, data center technology, high-performance equipment can gradually replace the low performance equipment, so that the efficiency of the industry has been steadily improved. In the overall operation of the computer, the high energy consumption structure in the cloud computing background will replace the low energy structure.

5. Conclusion

There are opportunities for the application of data distribution storage technology in cloud computing environment, and there are also some challenges. The storage application should be further optimized and the data and central network structure should be perfected. Different types of applications will form different features. In the process of data access, the optimization of the routers should be improved. In the design, the effective organization and management of the data is formed to improve the throughput of the system. In the actual system deployment, we should mine the system data, optimize the underlying data, improve the data center network topology, and enhance the effect of access [6].

References

[1] Wang Y J, Sun W D, Zhou S, et al. Key Technologies of Distributed Storage for Cloud Computing [J]. Journal of Software, 2012, 23(4):962-986.

[2] Jiang G S, He X L. Research on Key Technology of Distributed Data Storage in Cloud Computing Environment[C]// International Conference on Advanced Design and Manufacturing

Engineering. 2017.

[3] Cheng Z. Research of Distributed Data Optimization Storage Model in the Cloud Computing Environment[C]// International Conference on Smart City and Systems Engineering. IEEE Computer Society, 2017:193-198.

[4] Shan S, Wu D. Key Technologies Research of Distributed Storage Based on Cloud Computing[C]// International Conference on Information Sciences, Machinery, Materials and Energy. 2015.

[5] Pan Q. Analysis of distribution storage key technology based on Cloud Computing [J]. Electronic Test, 2014.

[6] Luo Z, Xu G. Research of Provenance Storage in Cloud Computing Environment[C]// IEEE, International Conference on Trust, Security and Privacy in Computing and Communications. IEEE, 2015:660-664.