Research on Key Technologies of Distributed Storage in Cloud Computing Environment

Liu Tao[1, a]

[1]Chongqing College of Electronic Engineering, Chongqing, China,401331

[a]15823118520@139.com

Abstract. As a new commercial computing model which sharing resources, cloud computing is widely used in the field of data storage, it stores massive amounts of data in the data center and provides transparent, secure and reliable computing services and data storage for upper applications and services. It has been seen as the "revolutionary computing model." In this paper, we combined with the actual work and set improving the storage fault tolerance, scalability and reduce energy consumption under the cloud environment as a target, analyzed and elaborated the key technologies of distributed storage in cloud computing environment.

Introduction

Nowadays, cloud computing technology has become an inevitable trend of network technology development and it will also be an important basis for the development of next-generation computing model. At present, cloud computing technology in many areas have an extremely important role, such as computer science, business computing, has gradually become the focus of attention of all sectors of society. In the cloud computing environment, the relevant research data distribution and storage are concentrated in the data center and the data center is the responsibility of management and organization. In other words, distributed storage key technology is at the core of cloud computing technology. In general, the data is located in the storage center by the combination of many millions of nodes formed, with great storage specifications, PB-class and EB-class majority. However, distributed storage technology has always been the problem of data failure, which is the promotion and application of cloud computing has a very limited and hindered, making the cloud computing operating costs are relatively high. Therefore, the study of distributed storage key technology in cloud computing environment has become the top priority in the industry.

The Background of Distributed Storage Technology Based on Cloud Computing Environment

The 21st century is the era of network information and there is a wide range of applications of computer technology in various fields, the amount of data the size of the rapid growth. A large number of data is stored in the computer alone and it is not able to meet the development needs of modern society, cloud computing is also to adapt to the development of modern computer technology to solve a large number of data storage and communication technology and other social requirements.

Distributed storage technology in the cloud computing environment is the user through their own demand on the Internet computer storage system for analysis and then leases or purchases the Internet storage space to meet their own data storage requirements. How data is stored on the Internet is also a key issue. Based on the cloud computing environment, the different data center makes it have different nodes, a large number of data will be stored in the appropriate data storage node or data center, data center through the computer program to organize the data organization order Choreography, and then through a special computer port to convey to the user, the user then through a special computer port data will be stored in the data center. Cloud computing is a new computing model, based on Internet technology, its biggest feature is the ability to share the basic resources.

The Constitution of Distributed Storage Key Technologies in the Cloud Computing Environment

Data center is the fundamental guarantee of the cloud computing technology system safe and stable operation. Typically, the basic composition of cloud computing is divided into two parts, one of the hardware part of the computer equipment and related systems to provide the necessary support, the software part of the main operation for the normal operation of the data center. According to the data center data access path forwarding node function type cannot cloud computing environment network construction is divided into the following types, the following specific description:

In the cloud computing environment before the advent of distributed storage technology, switch equipment for a long time in the modern computer technology has been a very wide range of applications. In the cloud computing technology system, the switch can play the role of the data center, with the switch and you can connect the user and the entire network system to achieve data-free exchange. Under normal circumstances, the switch is connected with the computer server to play a role in order to smooth and orderly data packets in the Internet technology environment for forwarding and delivery. It should be noted that the computer server mentioned herein is essentially a storage device capable of data storage and reading operations. The switch is usually a tree structure, from the inside to the outside there are three layers, respectively, the polymerization layer, the edge layer and the core layer. Combined with the actual use of the switch, you can see that the switch in the use of the process is simple, easy to connect, expand the simple and significant features. However, the switch technology also has a fatal flaw, cannot meet the application of large amounts of data.

With the emergence of the Internet in the field of servers, network technology, the traditional service model was soon to be completely broken, then, the use of single computer equipment, you can achieve a smooth connection with the Internet. On the server, the main task is to the Internet environment in the data information object forwarding, the biggest technical advantage is that the internal structure is extremely simple, and the biggest drawback is that the data storage space is too small, which is limiting its wide range The biggest problem.

The Data Center Network Construction Technology

Data center is to ensure the normal operation of cloud computing core, in the logic application includes software and hardware two aspects. Software refers to the data center to provide services and the required installation procedures; hardware refers to the data center infrastructure, including computer equipment and support systems. Data center node structure and cloud computing data reliability and access efficiency are closely linked, through the data center network construction technology research, can effectively connect to the data center equipment and server nodes to organize and connect, And for the upper application and services provide a good interface. By classifying the node types in the data center, the data center network can be classified into the following three types.

Switch-Centric Architecture. Switch-centric network architecture is to connect the servers in the data center to each other using switches. Data packets are also forwarded through the switch. The server is only responsible for the storage and processing of data. The structure is generally connected into a three-tier tree structure, respectively, the edge layer, the polymerization layer and the core layer. Switch-centric tree structure is easy to operate, simple and intuitive, and easy to expand by using switch functions. However, there are also some problems such as low flexibility, low utilization of the lower server and serious waste of switch resources.

Server-Centric Architecture. The structure is installed in each service card, then use the network cable to connect the server structure. This structure does not use the switch as a data forwarding device, so the server not only need to deal with data storage, but also responsible for the packet forwarding function.

The server-centric architecture is simpler in terms of line connectivity and fabric composition because there is no switch equipment and no single point bottleneck in the tree structure, allowing

the server to interact with the underlying network and developing more efficient routing algorithms. The disadvantage of this structure is that it has high link redundancy and large server resource, which makes the server load pressure increase, which leads to the decrease of the whole computing efficiency of the server, which leads to the increase of cost and the loss of performance.

Mixed Structure. Hybrid structure is the mixture of the above two structures, the structure of the switch as a server connection node, and installs multiple servers in the network card, so that the server is also involved in the packet forwarding function. The use of hybrid structure can be designed to be more suitable for specific scenarios of the network structure, more flexible, free, in the same performance conditions are usually lower cost.

The Data Fault-Tolerant Technology

One of the key technologies in distributed storage in cloud computing environment is data fault tolerance. Data fault-tolerant technology refers to the technology and method of guaranteeing uninterrupted data service when the system is activated with errors. The use of data fault-tolerant technology can not only improve the reliability and applicability of the system, but also enhance the reliability of data. Data access efficiency. The current data fault-tolerant technology is mainly used to achieve data redundancy, but also increases the consumption of resource storage, so the need to ensure system fault tolerance, but also to minimize the loss of resource storage to achieve cost reduction. Commonly used data fault-tolerant technology includes erasure code fault-tolerant technology and replication fault-tolerant technology of these two categories.

The replication-based fault-tolerant technology is simple, intuitive, easy to implement and deploy, but requires the creation of several copies of the same size for each data object. The storage space is very expensive; error-correcting techniques based on erasure codes can fuse information from multiple data blocks To less redundant information, so can effectively save storage space, but the data read and write operations to separate encoding and decoding operations, you need some computing costs. When the data fails, the replication-based fault-tolerant technology only from the need to other replicas to download the same size data can be repaired; based on erasure code technology you need to download the amount of data is generally much larger than the size of the failure data, repair costs higher.

Encoding-based fault-tolerant techniques use multiple data objects for encoding and generate encoded data objects to achieve a reduction in the amount of storage that can result from full replication. Error-correcting technique is a kind of coding technology based on channel transmission. It can tolerate the loss of multiple data blocks and introduce it into the distributed storage area. The fault-tolerant technology is the most commonly used, and can tolerate multiple fault-tolerant technology of data block failure. Erasure-tolerant technology can tolerate more than one data block information to less redundant information, can effectively reduce the storage space, but the amount of data downloaded is often greater than the size of the failure data, the calculation of large overhead, repair Of the high cost.

Replication fault-tolerant technology is in a data object to create a number of the same module, and distributed to different storage nodes. When a data object fails, can still use other effective data module to obtain the same data. Replication fault-tolerant technology mainly includes data replication strategy and data organization structure of these two aspects. Replication fault-tolerant technology is easy to implement and simple and intuitive features, can effectively improve the data read efficiency, so the actual cloud computing environment has been more widely used. However, because each data object needs to create a certain data module, so the storage space requirements are higher.

The Energy-Saving Technology

Data storage in the cloud computing environment is an important part of it, and it is also the basis of various computing and services. Data show that the energy consumption of the storage system usually accounts for about 40% of the energy consumption of the entire data center, thus enhancing

the research and application of distributed storage energy saving technology in the cloud computing environment has important practical significance to reduce operating costs and protect the ecological environment. For the loss of distributed storage energy in cloud computing environment, the current energy-saving technology can be divided into two categories: software energy saving and hardware energy saving. Software energy-saving technology mainly utilizes the regulation and control of software to realize efficient scheduling of storage resources to reduce energy consumption. Hardware energy-saving technology mainly reduces the energy consumption of storage hardware devices to achieve the reduction of distributed storage energy consumption.

The Software Energy-Saving Technology. Software energy-saving technology goal is to minimize the performance of the system under the premise of the storage data center energy consumption is reduced. The current software energy-saving technology mainly has the node management technology and the data management technology these two aspects emphases. Node management technology through the cloud computing environment in the distributed storage part of the node or disk selection application, so that the remaining nodes and disk into a low power or off state, in order to achieve energy consumption reduction; data management technology and node selection close, Including the management technology based on dynamic data, static data management.

The Hardware Energy-Saving Technology. Hardware energy-saving technology from the hardware level, can be divided into two technologies. One technique is to replace high-performance, high-energy hardware devices with low-performance and low-power devices in order to reduce the energy consumption of distributed storage from the data center. Another technology is to use the new- To achieve a reduction in the energy consumption of the computer nodes and the overall distributed storage.

Conclusion

As a new computing model, cloud computing is playing an important role in the IT industry and business and it changes our lives in many ways. The current cloud computing is still in the embryonic stage, this paper analysis the key technologies of distributed storage in current cloud computing environment from the data center network design, data storage organization and other aspects and we hope that the relevant technology to be better studied and promote the growing development of cloud computing.

References

[1]   Chen Quan, Deng Qianni: Journal of Computer Applications, Vol. 9 (2009) No 53, p.25-26

[2]   Hongli Zhang: Software Journal, Vol. 20 (2009) No 27, p.74-76

[3]   Qin Guo: Computer CD-ROM Software and Application, Vol. 1 (2006) No 33, p.11-14

[4]   Wang Yijie, Sun Weidong: Guangxi Normal University, Vol. 3 (2007) No33, p.121-124

[5]   Chen Zhao: Beijing University of Posts and Telecommunications, Vol. 12 (2011) No5, p.58-61