

Design and Implementation of Storage System Based on Big Data

Biao Wan

Wuhan Polytechnic, Wuhan City, Hubei Province, 430070 China

Abstract. In the era of big data, the storage of massive data has become an important issue that enterprises need to solve. However, the existing storage model hinders the pace of the times, and new storage technologies and storage models that adapt to the development of the times must be studied. This paper first summarizes and summarizes the problems faced by Big Data, then proposes and analyzes popular cloud storage and cloud computing for storage problems, discusses its structure and models, introduces such technologies, and will lead the development of the times.

1 The Introduction

In the information age, the Internet of things has changed the way of life, and the scale of data [1] is also growing rapidly. With the increasing awareness of the value of data [2], people begin to explore how to use these data better. At the same time, the sources of big data [3] are also increasingly diverse, with complex structures such as pictures, text, video, voice and so on. Sometimes big data is produced in real time, which requires the analysis tool to analysis it immediately. Some large enterprises also need data storage [4] and analysis work. They need to analyze the data in real time, make strategic decisions that are conducive to the development of enterprises, and improve the competitiveness of enterprises.

As the user grows, the computer world becomes larger and more complex. While people's demands for the Internet are becoming more and more complex, they also require corresponding advances in technology. Cloud computing has evolved into a more popular computing model because of its excellent performance. In general, cloud computing selects cluster commercial computers for massive data processing. This technology makes it more convenient for users to process data [5], attracting more users. At present, the amount of data that users need to process is larger and larger, and the structure types of data become more diverse. How to safely and efficiently store and backup big data becomes a problem. This paper mainly studies the related problems in big data storage.

2 The big data

Big data is of great value because of its huge amount of information. It is important to find and use the data we need from big data. Traditional data management model cannot

meet the requirements of big data, so it is hot to study efficient data storage.

The basic processing flow of big data consists of three parts: data analysis, data extraction and integration and data interpretation. The value of big data lies in the improvement of decision making after analyzing the data, so data analysis is of great significance in the process of big data processing. Traditional data analysis techniques generally include data mining, statistical analysis, machine learning, etc., and relevant technologies face some challenges in big data:

(1) Big data has many application features, such as real-time, accuracy, etc., but we often need to deal with the balance between them. To deal with big data and parallel computation, we usually need to improve the scalability of the algorithm.

(2) The increase in data volume will eventually increase the data noise. Therefore, before data analysis, we need to conduct pre-processing, such as data cleaning. However, when we need to preprocess massive data, the hardware performance and related algorithms of the machine will also be severely tested.

(3) Because of the large data type and large amount of data, the distribution characteristics of the whole data are often difficult to grasp when analyzing. This leads directly to the problem of designing indicators when measuring the results of data analysis.

Cloud computing integrates new research ideas based on traditional technology, which leads to the development direction of distributed computing. Cloud computing is a popular way to process large amounts of data. It not only uses the distributed resources more efficiently, but also has a high throughput. What's more, it also applies to large-scale computing problems, and figure 1 represents the relevant technologies for implementing cloud computing.

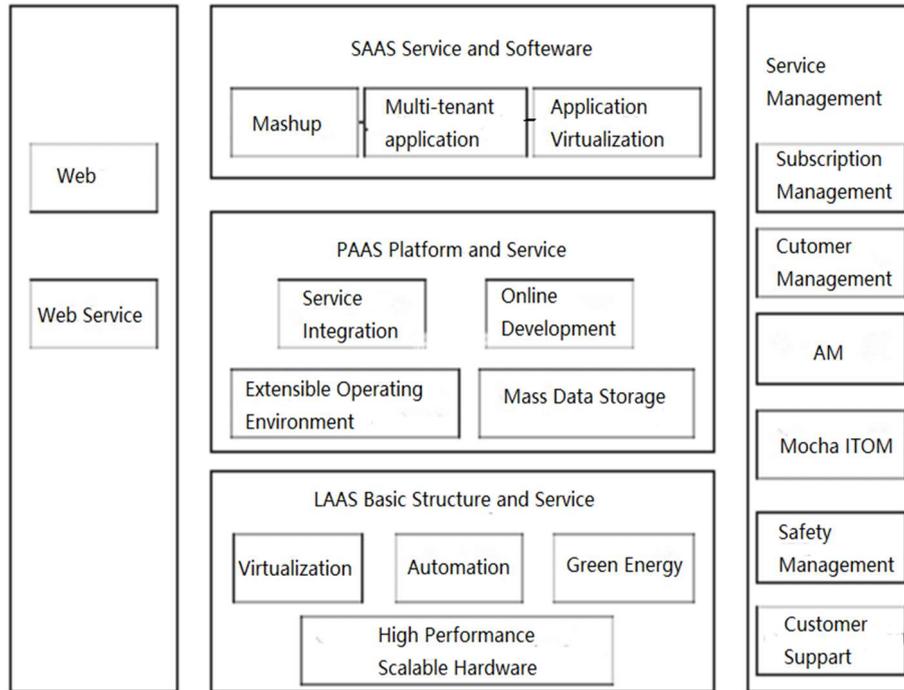


Figure 1. Related technologies for cloud computing.

3 Cloud Storage

To make business access and storage services more widely available to all users, cloud computing typically uses software to mobilize all kinds of storage devices. If we use cloud storage, we don't need to know anything like disk type and quantity, storage device model, interface and transport protocol, capacity size, and so on. The users can guarantee the continuity of the business and the security of the data, so we often don't need to create a disaster recovery system or

even a data backup system. There is no need for us to worry about upgrades or updates to some of the software/hardware systems in the storage device or even the maintenance of storage devices.

Compared with traditional storage, cloud storage is usually a complex system composed of many parts, not just a hardware device. There are many components, such as storage devices, client programs, public access interfaces, and servers. Figure 2 represents the cloud storage model.

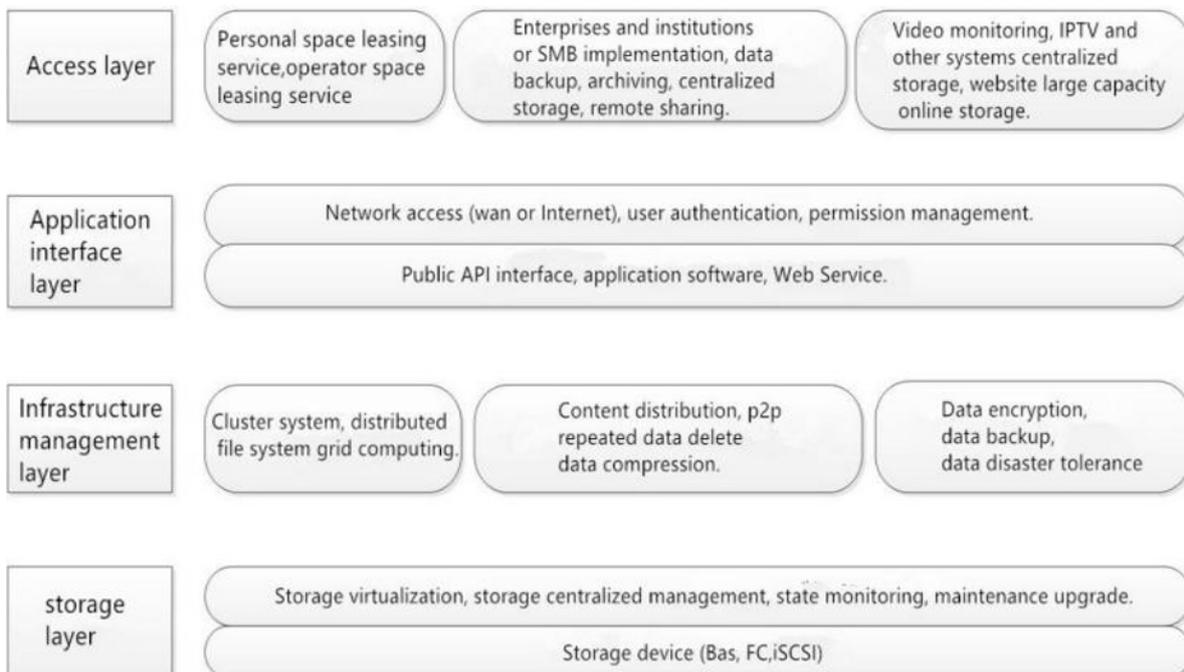


Figure 2. Cloud storage model.

Cloud storage reduces the cost of investment and enables enterprises to have a set of cloud storage services. The current mainstream approach to cloud storage selection is centralized architecture. The purpose of the data backup is to restore the data, first to collate the data and the application collection into a backup copy, and then to store it in the

remote space. In that case, if the data is easy to recover after the data is lost, this method guarantees the security of the data, as shown in figure 3, which has four layers from top to bottom.

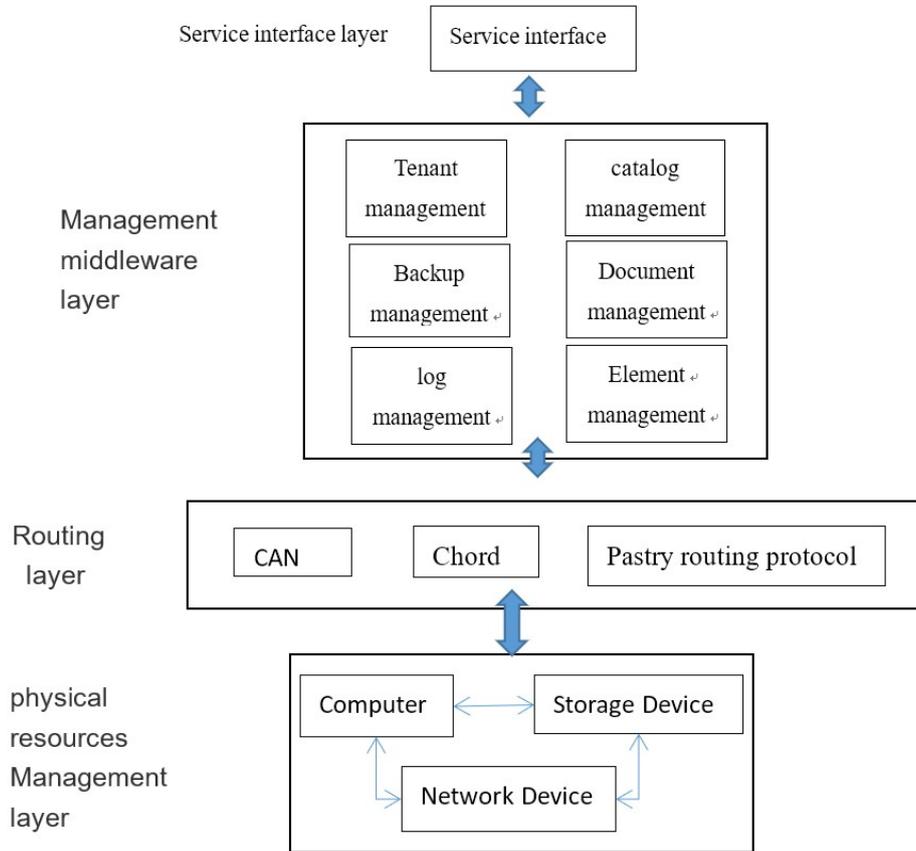


Figure 3. Cloud storage backup structure.

What shown in figure 4 is the Connections between nodes in the cloud storage and backup system.

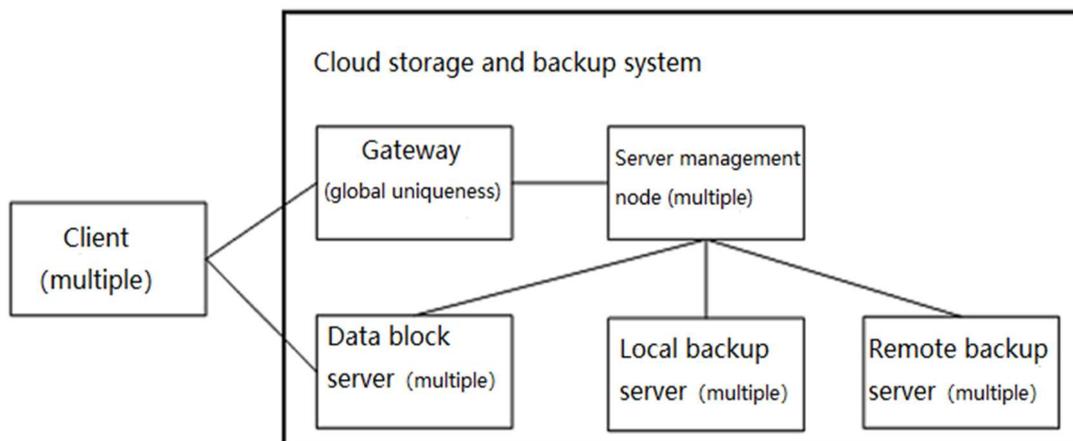


Figure 4. Connections between nodes.

The client sends requests through the interface while forwarding the routing request, but if the file is deleted, it needs to recover the file, shown as figure 5.

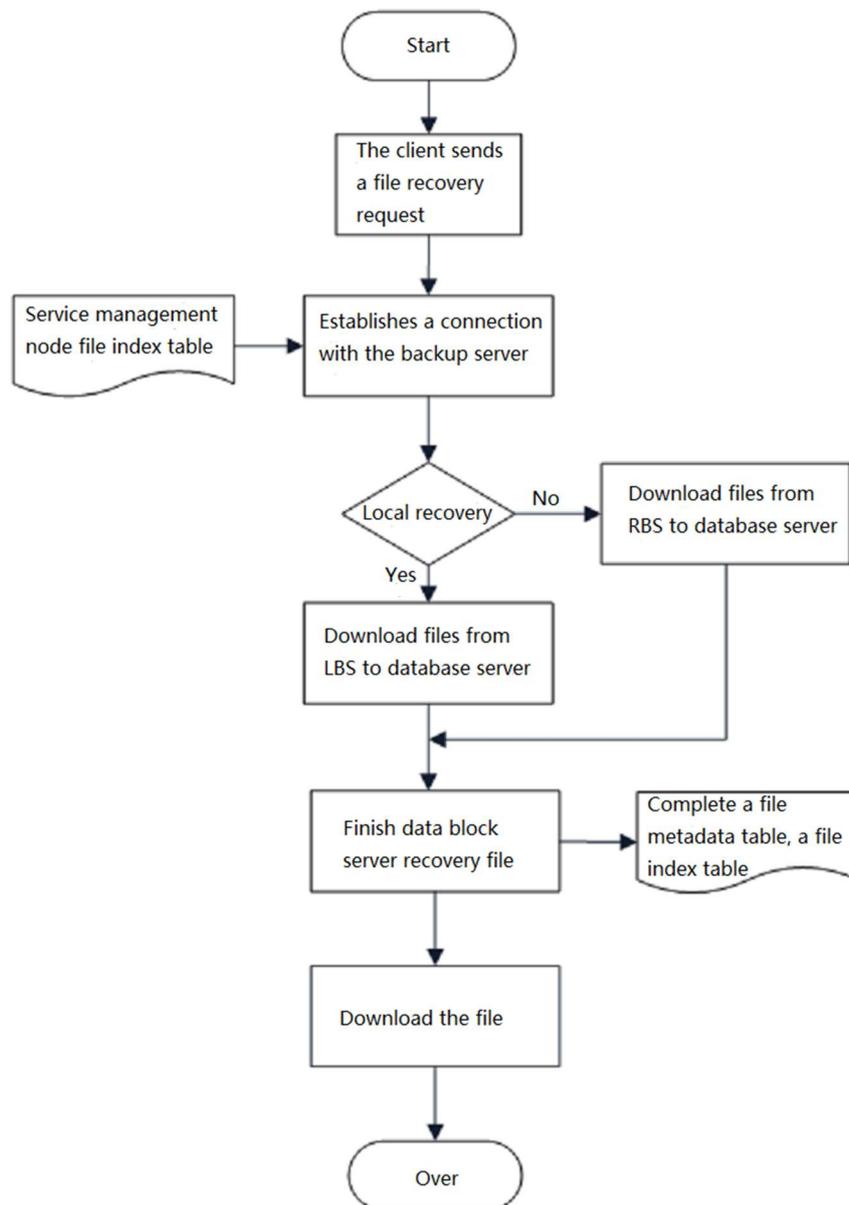


Figure 5. Recovery flow chart.

4 Application and prospect

With the development of network science and technology, more and more people use the Internet, resulting in big data explosion. This paper studies the existing and related models of big data storage systems to meet the needs of big data processing. Cloud computing and storage are new directions in the field of storage and computing services, and the generation of cloud storage indicates a new development direction for mass data storage. At the same time, it is an extension of cloud computing, they have market space between each other, and can meet the differentiated needs of different users. To better adapt to the storage needs of The Times, the technology of data management and storage should be further improved and developed.

References

1. Guo Kai, Pen Keyin, Lei Lei. Design and implementation of data storage and management system for seismic waveforms [J]. China science and technology resource guide, 2014,49 (06): 76-80.
2. Li Rongli. Design and implementation of large-scale network data storage system [J]. Modern electronic technology, 2014, 40 (02): 118-121.
3. Wang Wei, Liao Zhengyu, Zhang Hui, Guo Dong. Design and implementation of data storage and analysis system for railway signal system based on big data [J]. Information network security, 2017 (01): 29-37.
4. Zhao Yan, Huang Weijian. Design and implementation of big data cloud storage system for smart home [J]. Modern electronic technology, 2016, 39 (10): 21-24.
5. Shi Kang, Zhu Yuanping. Research on the storage technology of intelligent power grid in smart grid based on big data technology [J]. Automation and instrumentation, 2018 (02): 65-67.