Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 2, No. 2, 2022

Research on Cloud Storage Technology in Digital Campus Big Data Scenario

Chao Huang

Liuzhou City Vocational College, Liuzhou, Guangxi 545036, China

Abstract: Realizing campus intelligent management and service mode by combining cloud storage technology is the focus of current research in digital campus construction work. By analyzing the IT development trend of digital campus, the defective problems of traditional storage system, from the architecture design and platform construction of cloud storage system for digital campus, we propose a cloud storage construction scheme for campus big data scenario, and discuss the application of virtualization, streamlined configuration, storage optimization and other technologies in cloud storage. The cloud storage technology is applied in the construction of digital campus, which strongly accelerates the process of digital campus informatization.

Keywords: Digital campus, Cloud storage, Virtualization, Big data.

1. Introduction

The digital campus network has become an important resource sharing and collaboration platform for universities, and all kinds of informatization services are deployed on the campus network, which greatly improves the construction, management and services of the university. Big data in the campus in the 5G era comes from human users and the device side of computing devices, mobile devices, cameras, etc.[1] . Digital campus construction collects, cleans, processes, mines and analyzes data from various fields such as teaching, research and management, so that various resources can be fully utilized and shared, thus realizing campus networking, digitalization and intelligent services. Researching data storage in the big data scenario of digital campus is an important task in the current digital campus construction.

2. Digital Campus IT Trends.

2.1. Smart Device Era

IT devices are fully customizable and can be pre-configured so that when users use them, it takes minutes or hours to put them to work, rather than days. This greatly improves the deployment of applications. This greatly improves the efficiency of application deployment.

2.2. Infrastructure Optimization and Resilience

The closed and rigid nature of traditional infrastructure brings problems such as expensive and low utilization of IT infrastructure. Transforming traditional infrastructure into modular, developmental, scalable, and programmable infrastructure through consolidation, virtualization, automation, and sharing can effectively improve utilization as well as optimize infrastructure and bring more flexibility. With the rapid spread of big data analytics platforms and applications, data collection and processing and analysis are done quickly and cheaply, requiring an inexpensive infrastructure with super computing power as well as unlimited storage. The solution is to use cloud computing technology to build clusters that form dynamic, elastic and scalable resource pools to effectively meet the increasing and intermittent demand for computing power.

2.3. High-performance Cloud Environment

Given the sheer volume of processing as well as data storage, university big data applications need to be deployed in a cloud environment so that they have all the cloud characteristics such as flexibility, adaptability, and affordability. The real advantage of a cloud environment is automatic scaling. In addition to upward and downward scaling, outward and inward scaling is the key difference in the cloud. The ability to automatically add new resources or withdraw allocated resources to meet changing needs makes the cloud the most suitable choice for low-cost, high-yield HPC (high-performance computing). Any parallel workload can be efficiently solved in the cloud. With the convergence of multiple technologies, high-performance clouds are becoming the next generation of affordable supercomputers that comprehensively address the storage, processing, and analytics challenges posed by big data.

3. Defects of Traditional Storage System

Traditional storage such as server hard drives, DAS, NAS, etc. are more focused on hardware fault isolation and better stability and performance, but they also have some defects and are not suitable for the needs of high performance computing architectures and applications in big data scenarios. The main drawbacks are: (1) integration of technology stacks from multiple vendors. Storage systems use different components from different vendors' technology stacks and lack the ability to design and deliver end-to-end HPC solutions. This dramatically increases the complexity of the system and leads to I/0 bottlenecks, making testing, quality control, and long-term maintenance very difficult. (2) Limited automation in the configuration and tuning process. Multi-vendor HPC systems typically include a step-by-step installation process, which is often long and difficult due to the involvement of components from different vendors. This time-consuming installation and configuration process can impact project duration, making the system highly unsuitable for dynamically changing requirements. The lack of integration settings also affects the ability to interact with all the components of the system and performance tuning. (3) Lack of management software support. Management software is a very critical component of the storage system, because it plays a very important role in the smooth operation of the entire system. Users of traditional storage systems will encounter great difficulty in capturing and recording various parameters (such as performance) of the storage system. The lack of sophisticated diagnostic tools makes log analysis and troubleshooting extremely difficult.

4. Cloud Storage System Construction

4.1. Architecture Design of Cloud Storage System

Cloud computing has revolutionized the technology used to store information and run applications. Cloud storage is storage based on cloud computing technology. It does not refer to a specific storage device or technology, but rather to a large collection of storage devices and servers that are used to hold data in a cloud computing environment. Users of cloud storage do not use a specific storage device, but use the cloud storage system through some kind of access service. Storage resources in a cloud storage system are provided in the form of pools and are allocated in real time on demand; capacity is provided on demand and storage resources can be used according to the needs of big data applications, which are highly scalable and elastic; and cost-benefit, which provides significant economies of scale by paying for the use of resources. The above features of cloud storage make it the first choice for data storage in big data scenarios. The cloud-based storage architecture model is designed in Figure 1.

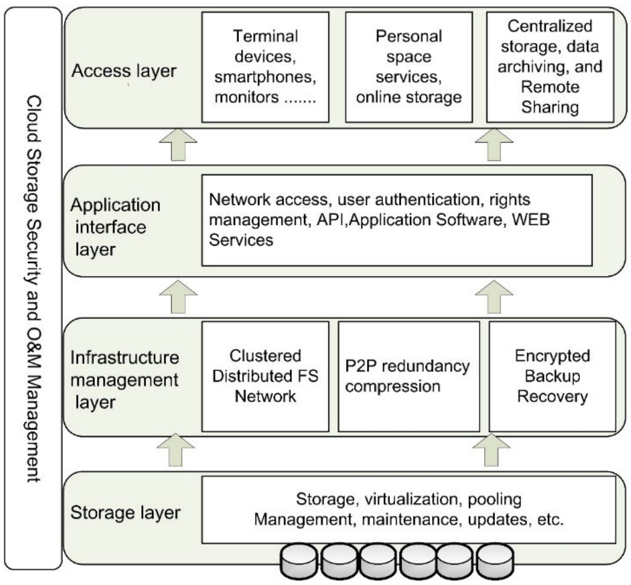

Figure 1. Architectural model of cloud storage system

(1) Storage layer: The bottom layer of the cloud storage architecture. It contains different types of storage devices. The storage devices included in this layer include Fibre Channel storage devices, IP storage devices, DAS storage devices, and so on. These storage devices may be geographically located in different regions and are connected together through the Internet or LAN. This layer has a unified storage management system that can manage all the heterogeneous types of devices located in the same storage pool and deliver them as a service on demand, which can be called Storage as a Service or Infrastructure as a Service[2] . The key technology used in this layer is storage virtualization.

(2) Infrastructure management layer: The underlying storage devices are managed in a unified manner and the necessary infrastructure is provided. This infrastructure is important because it provides various key functions such as security, space management, backup, storage consolidation, etc. using technologies such as clusters and grids. Other services provided by this layer are backup, disaster recovery, encryption, compression, etc.

(3) Application interface layer: The application interface layer provides for various interfaces or APIs to support or use cloud storage use cases. Some common cloud storage use cases are data archiving applications, backup applications, and so on. Different cloud storage service providers develop their own custom application interfaces based on the services they provide.

(4) Access layer: Any authorized user can log in to the cloud storage system through a standard public API to use the required cloud storage services. Different cloud storage service providers use different types of access mechanisms. The access layer has a catalog, which provides pricing and other usage details, and also provides details of the service agreement given by the particular service provider.

4.2. Cloud Storage Application Technology

4.2.1. Storage Virtualization

Storage virtualization is a mechanism to ensure that different heterogeneous types of storage devices are stored and managed as a single unit. This will support unified memory management, easier deployment, and integrated monitoring of the entire storage infrastructure[3] . Storage virtualization primarily involves partitioning available storage into virtual volumes. Virtual volumes can be created by combining different types of storage devices. These virtual volumes abstract away the details of the storage devices in the volume and then present them as storage devices to the operating system. Virtual volumes can be expanded, created, and deleted based on storage requirements without any downtime. Key benefits provided by storage virtualization include providing unified storage management capabilities; facilitating aggregation of heterogeneous storage devices; allowing storage resources to be allocated and released as storage requirements change; and providing scalability to the storage infrastructure.

4.2.2. Automatic Streamlined Configuration

The unused storage capacity allocated for various applications in the college campus network is costly to the college, including the energy waste it creates, as unused storage consumes energy and generates heat, increasing total energy consumption, and the space it takes up, as unused storage will take up unnecessary space that could be allocated to other useful components. To address such situations, thin provisioning is a necessary storage application technology. Thin provisioning refers to configuring storage according to actual demand. With this technique, logical storage is allocated to applications based on predicted demand. The actual allocated storage is based on the current demand of the application and is much smaller than the logical storage. Once the application's storage needs increase, more storage is allocated to the application from a pool of resources. With this approach, thin provisioning provides efficient use of storage and reduces waste of physical storage.

4.2.3. Cloud Storage Optimization

(1) Deduplication. This technique is used to ensure that duplicate data does not occur in the storage system, i.e., to ensure that duplicate copies of data are not kept in the system. This technique will reduce the storage space significantly. Deduplication is performed with the help of hashing method. The hashing method generates a unique hash value for each file based on its content. Each time a new file arrives in the storage system, the data deletion software generates a hash for the file and compares it with the existing hash. If the same hash already exists, it means that the same file is already stored in the system and no more storage will be done for it. If the new file makes only minor changes to an existing file on the system, only the changed part of the file will be saved, not the entire new file. The process of deduplication is shown in Figure 2 below.

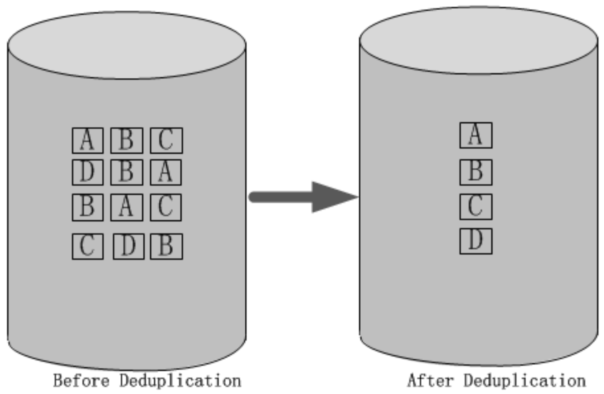

Figure 2. Deduplication

Deduplication can be performed at two levels: the file level and the block level. File-level deduplication is performed on files, and it ensures that only one copy of each file is retained in the system. In block-level deduplication, the files that need to be de-duplicated are split into blocks, and the software ensures that only one copy of the file is kept in each block. The uniqueness of a file or block is detected by comparing the generated hash value with the hash value of an already existing file or block.

(2) Compression. Compression techniques reduce the amount of data by removing the gaps that exist in the data. The main drawback of compression techniques is that it consumes computational cycles, which in turn may lead to performance concerns for users of storage devices, since the transfer of data to the cloud is a continuous process. However, the use of compression techniques does significantly reduce the storage space occupied by the data.

(3) Data backup. Phenomena such as data loss and damage are difficult to avoid in information data storage, which can affect the user's feeling of use and can cause serious economic losses[4] . By applying data backup technology, data deleted due to fault problems can be restored to ensure the normal operation of the system and the integrity of user data advantages, which is especially important for big data platforms.

5. Conclusion

Building a digital campus is a fundamental means to promote education informatization. Technologies such as cloud computing, big data and 5G bring new opportunities and challenges to the construction of digital campus. The scale, changeability and diversity of big data bring great challenges to IT. Big data requires high-quality IT infrastructure, platforms, databases, file storage systems, etc. Cloud storage provides high-performance and high-efficiency data storage infrastructure for campus construction, and strongly promotes campus informatization.

Acknowledgment

Fund Project: Scientific Research Project of Liuzhou City Vocational College (No. 2021KYB06, 2020KYB26).

References

[1] Roy Zhang. Research on information convergence cloud storage in big data environment [J]. Computer Programming Skills and Maintenance,2019(12):97-98.

[2] Zheng L, Ma W, Zhou ZB, et al. Exploration and reflection on the innovative development of digital campus information technology [J] China Modern Education Equipment, 2020 (1): 4-7.

[3] Zhang Huanghui. Exploring the development of cloud storage technology in the era of big data [J]. Information Recording Materials,2020(1):143-144.

[4] Liu Yuan. Research on cloud storage technology and its application under big data scenario [J]. Computer Engineering Application Technology, 2020(9):13-15.