

Data Storage Technology and Its development based on Cloud Computing

Linna Huang, Fenghua Liu

Department of Computer Engineering
Cangzhou Normal College
Cangzhou, 061001, China
E-mail: Hln0322@163.com

Abstract—Cloud storage of high performance is the basic condition for cloud computing. This article introduces the concept and advantage of cloud storage, discusses the infrastructure of cloud storage system as well as the architecture of cloud data storage, researches the details about the design of Distributed File System within cloud data storage, at the same time, puts forward different developing strategies for the enterprises according to the different roles that the enterprises are acting as during the developing process of cloud computing.

Keywords- cloud storage; cloud computing; service; Distributed File System

I. INTRODUCTION

Recently, by the birth of cloud computing and Software -As A Service (SaaS), it has become a trend to, through the Internet, get storage space, computing abilities and every kind of software services. Cloud computing [1] is actually one kind of distributed computing, the aim of which is to make the ultra-large-scale computing ability integrate with the storage resources through the internet, and to provide the computing ability and storage resources to the users as reliable services by their needs. However, cloud storage of high performance is the basic condition for cloud computing service, that is to say, almost all of the services based on cloud computing need the high-performance cloud storage to satisfy the data processing needs, so the cloud storage has become a research hotspot in the information storage field.

II. CLOUD STORAGE AND ITS ADVANTAGES

Cloud storage [2] is a new concept [3] derived and developed from the concept of cloud computing. Cloud storage is a system which makes a number of storage devices of different types of work, and provide data storage and business access functions cooperatively through application software. The high-performance cloud storage is the basic condition for cloud computing services. Compared with the traditional storage devices, the cloud storage is not only some hardware, but also a system [4] that is composed of network devices, storage devices, servers, application software, public access interfaces, access network and client applications and so on.

Cloud storage is professionally providing online storage services based on internet to the users. The users don't need to consider the complicated basic techniques, such as storage

capacity, types of the storage devices, data storage positions, the data usability, reliability and security; they can get almost infinite storage space and enterprise-level service quality through paying money by their needs from the cloud storage service providers. Thus, it can get extensive attention and support immediately when the cloud storage concept is created, such as Amazon, Google, IBM, who are representing the cloud technology providers, have turned the cloud storage concept into some practices applications with huge business value.

Compared with the traditional storage device and software purchase, cloud Storage has the following advantages:

A. Low cost and quick effectiveness

The enterprises don't need extra investment to construct any platform but some necessary terminal devices for receiving the cloud storage services. They just need to rent the storage services, which is able to avoid the risk of one-time investment and lower the running cost. The desired service of the enterprises can be available immediately and conveniently.

B. Easy management

Traditionally, the enterprises requires professional IT staff to maintain the system, however, under the cloud storage system, the maintaining and system updating are managed by cloud storage providers, which can benefit the enterprises with the low cost but the most updated and professional services.

C. Flexible options

Generally the cloud storage services can be purchased by customer numbers, duration and service items. The enterprises can adjust their renting options anytime according to the changed requirement, employee turnover, financing pressure, which lives up to "use as needed". The traditional purchase or customization ways are incomparable with the cloud storage way.

III. THE INFRASTRUCTURE OF CLOUD STORAGE SYSTEM

The cloud storage system is catering to many kinds of different types network-based online storage services, which has to consider such indexes as data security, reliability, efficiency and so on, however, it is definitely facing a lot of technical challenges because of large-scale users, extensive scope of service and changeable network environment.

The cloud storage platform infrastructure includes 4 layers, which are, from bottom to top, data storage layer, data management layer, application interface layer and user access layer. See figure 1.

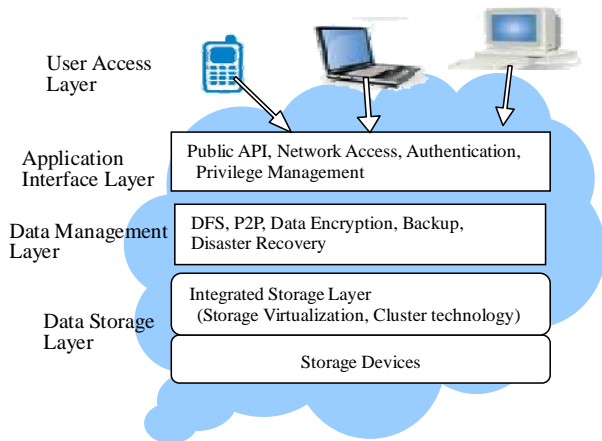


Figure 1. Cloud storage infrastructure

A. Data Storage Layer

Cloud storage system contains all kinds of data and provides a variety of storage services, which makes the cloud storage system a huge data pool. From the way most of the data storage infrastructure of the network service console is organizing, the multi-server-based data organization can better satisfy the online storage service requirements [5], when there are large-scale users, it can provide better quality services[6] for people in different places by constructing distributed data centers.

The data storage layer of the cloud storage connects all the different types of storage devices, which makes it possible for unified management of huge data, and at the same time, as well as for unified management of storage devices, state monitoring and dynamic capacity extension. It is actually one kind of distributed service-oriented storage system.

B. Data Management Layer

The data management layer provides the unified view of public management for the upper layers. By designing the unified management functions such as user management, security management, replica management, strategy management, etc., it can connect the underlying layer storage with the upper layer application seamlessly and make multiple storage devices work collaboratively, which can provide multiple services with better performance.

C. Application Interface Layer

The application interface layer is the flexibly extended, directly user-oriented part in cloud storage platform. It can provide different interfaces to provide corresponding service according to users' needs. The users can input and output data from the cloud storage through cloud service interface. The platform provider can carry out effective resource

allocation and access control through the application Interface layer. The collaborative design of the API of program and file system promotes the flexibility of the whole cloud storage.

D. User Access Layer

Through the user access layer, any authorized user can login the Cloud Storage platform and enjoy the cloud storage service anywhere with a network-connected terminal device in accordance with the standard public application interface. The cloud users include personal data storage users, enterprise data storage users, service integration providers and so on, among which, the enterprise storage users are the main users of the Cloud Storage service at present.

IV. DISTRIBUTED FILE SYSTEM

The foundation for cloud storage is the huge storage center which is composed of thousands of low-cost storage devices. These isomerized storage devices, through their respective Distributed File System, assemble these scattered and low reliable resources into a integration with high reliability and scalability, based on which is the user-oriented cloud storage service built. The typical cloud storage infrastructure is as shown in figure 2.

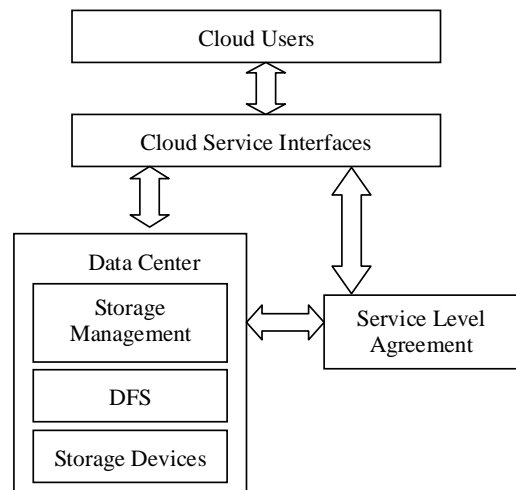


Figure 2. Data storage infrastructure in cloud computing environment

The data center, which includes all kinds of storage devices as well as the Distributed File System managing all kinds of isomerized storage devices, is the foundation to implement the cloud storage. DFS is the core of the cloud storage, which is used to make multiple devices in cloud storage work collaboratively. DFS makes the documents stored in scattered servers be accessed like in one position, providing the same service and more powerful data access performance.

According to the characteristics and requirements of data storage management in cloud computing environment, the corresponding DFS functions are as follows:

1) *Fault Tolerance management.* The cloud storage is a complex storage system constructed by a great number of

low-cost hardware resources, which always requires thousands of nodes to calculate together, so there must be some nodes not working at any time. It's common for some resource nodes not being working, so the reliability problem becomes particularly crucial. As a reliable file system, it has to consider usability, reliability, security and maintainability.

2) *Consistency and replication.* The storage model based on data block is commonly used in distributed file system which is adopted in cloud storage. In the Google GFS-like distributed file systems, they deal with the meta-data in centralized management, also they protect the data server from single-point failure by hot backing up the host computer.

3) *Naming System.* In cloud storage, the naming system realizes the mapping from the files to the data blocks as well as locating the data blocks to the data block servers. The name is used to share the resources, identifying utilities, pointing the positions and so on. Because of the complexity of the environment in isomerized thousands of nodes within cloud storage system, the naming system is required to consider the uniqueness of naming, the consistency of resolution, the dynamic link of names and objects, the expansibility of naming system and so on.

4) *Communication.* The communication between the processes is the core of distributed file system. The communications in distributed file system are all based on the underlying layer message transmission mechanism provided by the underlying network. The communication process has to follow certain protocols, and the network communication protocols are the foundation to construct the distributed file system. In the early cloud computing platforms, their storage mode may result in high storage I/O delay. The delay of local network is negligible when storing and backing up data in the local host, however, it becomes a crucial problem storing and backing up data in the cloud.

5) *Security.* The corresponding strategies, according to the security threat and security requirements, have to be made to carry out corresponding protection mechanism in order to solve the security problem in cloud storage, to promote the defending ability of cloud computing data centers. In cloud storage, the data security is based on trust which exists in management programs. When most virtual machines are sharing physical LUN, CPU and memory, the management programs have to protect the data from damaging or from fault access by the virtual machines. At this time, the design of file system has to consider these two aspects:

a) *Secure Channel Problem:* the communication security problem of the users and processes. The secure channel problems refer to data encryption, user authentication, message integrity & confidentiality as well as digital signature.

b) *Authorization Management Problem:* Based on secure channel, the users must access data and other resources under the authorization. Authorization management involves the access control problem of the system.

In Figure 2, the optional Service Level Agreement (SLA) is very important to a commercial cloud storage system. SLA defines standards for user storage service and the corresponding price. Service level agreement is the key element to decide whether the cloud storage service is successful or not. The scientific and reasonable SLA can attract more commercial users to use Cloud Storage service.

V. THE DEVELOPMENT TREND AND ENTERPRISE COUNTERMEASURES FOR CLOUD COMPUTING AND CLOUD STORAGE

As an emerging commercial mode, the cloud computing satisfies the requirements of cost-saving, resource-saving, environment protection and so on. It means a lot to multiple industries to develop cloud computing. For example, the scientific earthquake monitoring, ocean information monitoring, astronomical information calculation, DNS information analysis, the patience case storage & analysis, medical image process in medical science, the TV broadcasting and video on demand, virus database storage in network security, junk-email shielding, graphics and image processing, animation source material analysis, high simulation animation making, picture searching in huge amount of pictures, huge information searching and so on. So the cloud computing will be used in every aspect of our life in the future.

Following the gradually maturing concepts and technologies of cloud computing and cloud storage, the enterprise data center is experiencing the transformation from the traditional data center to cloud data center. In the development of cloud computing, every company or enterprise should make different developing strategies according to their different roles.

A. The Cloud Computing Provider constructs the Cloud Computing and Cloud Storage Center

Google, Amazon, IBM, Microsoft and Sun have joined the cloud market competition in succession, becoming cloud computing providers. For example, Amazon opened its "Elastic Compute Cloud" in 2007 [7], provider its compute ability to small companies by needs. Sun launched "Blackbox" based on cloud computing theory, which is called the future data center. Beginning with Gmail, Google has been trying to provide its compute ability and service to the users through the internet, in an open-source way to spread its cloud computing platform and to challenge Microsoft's authority [8]. However, Microsoft was trying to propel the cloud computing research by providing money and devices to the computer science researchers in 6 universities including Carnegie Mellon University, MIT and so on, what's more, it introduced Blue Cloud plan with high profile in August, 2007.

B. Small and medium-sized enterprises rent compute ability and data storage service from the Cloud Computing Providers

The cloud computing predicates huge commercial opportunities for the small and medium-sized enterprises or the entrepreneurs who don't want to construct or maintain their own basic infrastructure. By the help of cloud computing, they can compete with the large enterprises in a higher level. They don't need to buy the expensive hardware, but rent the compute ability and data storage services by paying the cloud computing providers. It avoids the risk of hardware investment and software maintenance, and it can save the cost to support innovation in order to enforce the enterprise competitive power.

C. The large enterprises who require more data confidentiality and exclusiveness construct private Cloud

Some large enterprises require strengthened confidentiality and exclusiveness of the applications and integrated data, so they don't want to risk putting its information on the public cloud. They are more likely to construct their own private cloud considering the security and confidentiality. The private cloud is constructed for one customer, so it can effectively control the data, security and service quality. The private cloud can be deployed under the Firewall of the enterprise or in a securely hosted position. It's easy to imagine the charm of the private cloud: It not only avoids the security and uncertainty problems related with the public cloud providers such as Amazon etc., but also provides the advantages brought by the cloud, such as easy deployment and easy expansibility, which enforces the security but is also good for innovation.

VI. CONCLUSION

The cloud computing maximizes the resource sharing. Its ultimate purpose lies in totally open large-scale distributed way, not only being used in computing, but also in storage. "Cloud Computing" brings such a revolution: Some professional network companies construct the cloud -- data storage and computing center, and the users can easily access the cloud which is used as a material storage and application service center through a wire and the browsers. By means of this practical commercial mode, the cloud computing and cloud storage are attracting more and more attention and application.

REFERENCES

- [1] Hayes B, "Cloud Computing," Communications of the ACM, vol. 51, pp.9-11, July 2008.
- [2] Wu Yongwei, Huang Xiaomeng, "Cloud Storage," China Computer Federation Communication, vol. 5, pp. 44-51, June 2009.
- [3] Shi Liping, "Analysis of Cloud Storage Technology Based on Web," Modern Computer, pp. 117-119, March 2010.
- [4] Wang Qingbo, Jin Zi and He Le, Virtualization and Cloud Computing. Beijing: Electronic Industry Press, 2009.
- [5] LILLIBRIDGE M, ELNIKETY S, BIRRELL A, et al. "A Cooperative Internet Backup Scheme," Proc. of the 2003 USENIX Annual Technical Conference (USENIX'03), San Antonio, TX, USA. Berkeley, CA, USA: USENIX Association, pp. 29-41, Jun 12-14, 2003.
- [6] PAMIES-JUAREZ L, GARCIA-LOPEZ P, SANCHEZ-ARTIGAS M. "Rewarding Stability in Peer-to-Peer Backup Systems," Proceedings of 16th IEEE International Conference on Networks (ICON'08), Dec 12-14, 2008, New Delhi, India. Piscataway, NJ, USA: IEEE, pp. 6.
- [7] Chen Geng, Zheng Wemin, "Cloud Computing: System Cases and Current Study," Software Journals, vol. 20, pp. 1337-1348, May 2009.
- [8] Wu Xinran, Yang Sirui and Wu Xiaoxi, "Infrastructure Facing Service of Cloud Computing," China Computer Federation Communication, vol. 5, pp. 32-43, June 2009.