Cloud Storage Technology and Its Applications

Zhou Ke, Wang Hua, Li Chunhua

(School of Computer Science & Technology of Huazhong University of Science and Technology, Wuhan 430074, P. R. China)

Abstract: Cloud storage employs software that interconnects and facilitates collaboration between different types of storage devices. Compared with traditional storage methods, cloud storage poses new challenges in data security, reliability, and management. This paper introduces four layers of cloud storage architecture: data storage layer (connecting multiple storage components), data management layer (providing common support technology for multiple services), data service layer (sustaining multiple storage applications), and user access layer. A typical cloud storage application—Backup Cloud (B-Cloud)—is examined and its software architecture, characteristics, and main research areas are discussed.

Key Words: cloud storage; service; backup cloud

With the rise of cloud computing[1-2] and Software as a Service (SaaS)[3-5], cloud storage has become the focus of attention in information storage. It, unlike traditional storage, comprises not only hardware devices, but is a system of network and storage equipment, server, application software, public access interface, access network, and client programs[6]. Since its introduction, cloud storage has attracted great interest from service providers.

A user's local data can be stored in online spaces provided by a Storage Service Provider (SSP). A user need not build their own data centers but can apply for services from the SSP. In this way, repeated construction of storage platforms can be avoided and expensive investments in hardware and software infrastructure be saved.

1 Cloud Storage

Cloud storage differs from traditional

storage in many aspects. In terms of functionality, it is designed to deliver many online storage services, whereas traditional storage systems are primarily designed for high performance computing and transaction processing. In terms of performance, cloud storage places great importance on data security, reliability, and efficiency. With a larger number of users, a wider service range, and a complex and ever-changing network environment, cloud storage systems face greater technical challenges than traditional systems when delivering high-quality services. In terms of data management, cloud storage systems not only offer access to traditional files such as Portable Operating System Interface for Unix (POSIX), but also support mass data management for providing public service support functions, and maintaining data in the background.

A cloud storage platform can be classified into four layers: data storage layer, data management layer, data service layer, and user access layer. Figure 1 shows the architecture of a cloud storage platform.

(1) Data Storage Layer

A cloud storage system offers diverse storage services, and all data stored in the system forms a massive pool. For efficient storage, this data should be properly organized. Traditional data organization uses a single server and cannot meet throughput and storage capacity requirements of multiple users in a Wide Area Network (WAN)[7]. A Peer-to-Peer (P2P) architecture based organization method requires a large number of nodes and complicated coding algorithm to ensure data reliability[8]. In contrast, using multiple storage servers to organize data better satisfies the requirements of online storage services[9]. Distributed data centers can provide good Quality of Service (QoS) for a large number of users in different geographical regions[10].

By interconnecting different types of storage devices, the data storage layer can manage massive amounts of data in a unified way, and can employ centralized management, status monitoring, and dynamic capacity expansion of storage devices. A cloud storage system is essentially a
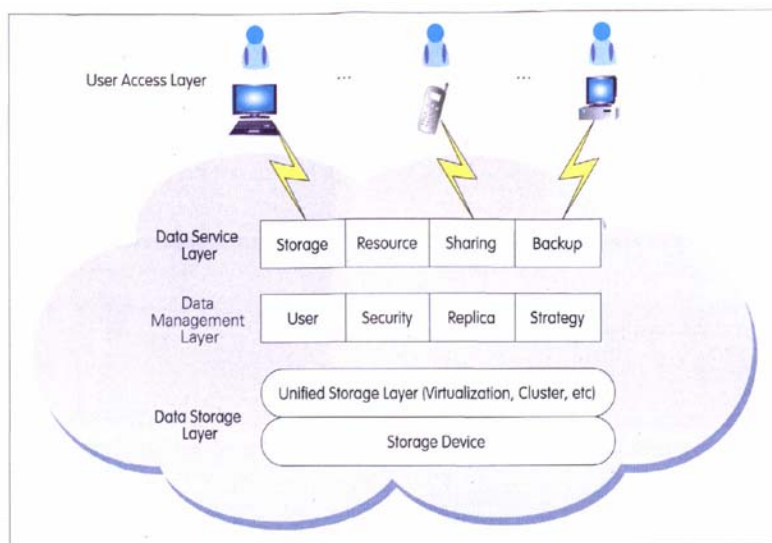
▲Figure 1. Architecture of cloud storage platform.

service-oriented distributed storage system.

(2) Data Management Layer

The data management layer provides the upper layer with a unified public management interface for different services. With functions such as user management, security management, replica management, and strategy management, this layer seamlessly associates upper-layer applications with lower-layer storage services. It also promotes cooperation between storage devices, enabling them to offer diverse and optimized services.

(3) Data Service Layer

The data service layer deals directly with users and can be flexibly expanded. Depending on user demands, different application interfaces can be developed to provide services such as data storage, space leasing, public resource, multi-user data sharing, or data backup.

(4) User Access Layer

In the user access layer, an authorized user can log into the cloud storage platform from any location via a standard public application interface and access cloud storage.

As an alternative to purchasing storage devices and deploying storage software, cloud storage has the following advantages:

(1) Low Cost and Quick Return

In building a storage platform that meets information management demands, purchasing storage devices and deploying software requires heavy initial investment. Software development often involves a long process of feasibility analysis, requirement analysis, software design, coding, and testing. By the time the software is developed, the demands might have been changed so that the software has to be redeveloped. This reduces the Quality of Service (QoS), increases cost, and delays the progress of information management. Enterprises repeatedly invest in traditional low-tech storage approaches; and for an individual enterprise, this means cyclic, high-cost technical upgrades.

Taking the cloud storage approach, terminal devices need only be configured to receive storage services so that heavy investment in platform building is not necessary. Storage services can be purchased according to the number of users and the usage time span, thereby avoiding the risk of heavy initial investment and reducing usage cost. Services can be used immediately and conveniently.

(2) Ease of Management

Traditional storage systems require maintenance to be performed by dedicated IT staff, and this incurs

additional cost. Maintenance and upgrade of a cloud storage system, however, is performed by the service provider, so that professional services are provided at the lowest cost.

(3) Flexibility

Traditionally, once an investment has been made into purchasing devices or deploying software, the storage system cannot be dynamically adjusted during its lifetime. As devices are renewed, disposal of the existing outdated hardware platform becomes difficult. Ever-changing business needs may require the software to be constantly updated, upgraded, or even redeveloped. So high maintenance costs are incurred as a necessity and are to some extent beyond control. Cloud storage services are generally charged according to the number of users, usage time, and service items. On-demand services released by an enterprise can be changed anytime according to business needs, personnel changes, or financial status.

2 Application of Cloud Backup

Cloud storage has a variety of applications, including data backup, data sharing, and resource service. It can also provide standardized interfaces for other network services. The following self-developed B-Cloud system is an example of cloud storage technologies and applications.

The deployment structure of B-Cloud is illustrated in Figure 2. It consists of three cloud levels.

The top level is called wide area (public) cloud. It covers all areas that backup clients can access via the WAN. Wide area cloud servers include wide area manager and wide area cloud storage nodes.

The middle level is called regional cloud and usually divided by geographic region (province or prefecture). Similar to wide area cloud, it has service nodes such as regional cloud manager and regional cloud storage servers.

The lowest level is called local (private) cloud. This is divided either by small geographical region or by an
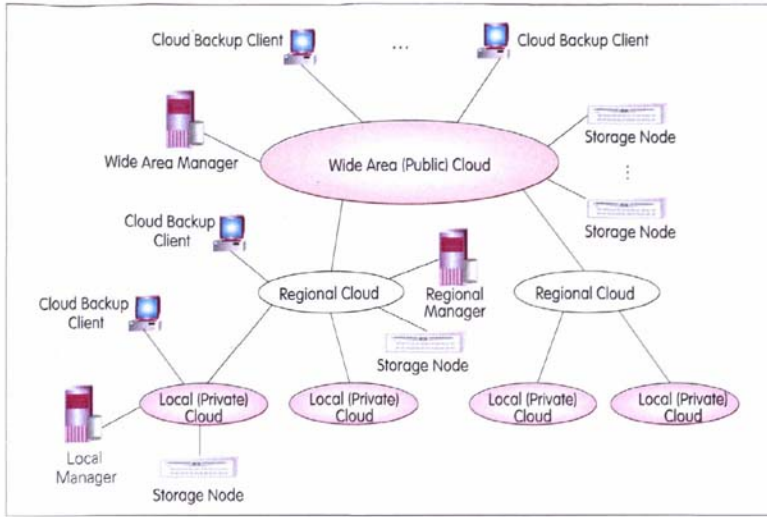
▲Figure 2. Topological structure of B-cloud.

entity such as enterprise, institute, or campus. Local cloud can run on a WAN or Local Area Network (LAN), and its users are only those within the cloud. It has service nodes such as local manager and private cloud storage servers.

Like wide area cloud, regional and private clouds have multiple local storage nodes that serve multiple backup clients.

The topology of the B-Cloud system looks like a tree; the wide area cloud acts as the root node, and regional and local clouds act as branch nodes. Each node has its own manager and storage nodes that perform backup task scheduling and backup data access. All nodes—including wide area cloud, regional clouds, and local clouds—are physically connected. Nodes at any two adjacent levels have a parent-child relationship in which the child node can be viewed as a special user of the parent node. The topological structure is very scalable. Even though only three levels are currently defined in the system, any node can be split into more levels when the number of users grows or the service area expands.

When a new user (or backup client) registers in the system, it first visits the super director server of the system—which is responsible for global user management. The server then assigns the user a backup cloud node

according to predefined assignment strategies and user information (such as the segment or region of the user's IP address, organization of the user's email address, or the user's geographic location). User information is maintained by the system's super director server. After registration, the backup client can log onto the system and communicate with the backup manager and storage nodes of the specific cloud to receive a service.

The principle of proximity access dictates that the nearer the client, the higher the data transmission efficiency and the lower the cost. An hierarchical topological structure creates an orderly relationship between multiple scheduling servers and multiple

storage servers of the backup system, enabling the system to better serve backup clients in different regions.

The characteristics of cloud backup determine application requirements, which in turn drive the development of three key backup technologies: parallel task scheduling, data organization and compression, and backup security. These technologies are research subjects of the B-Cloud system, and involve several aspects of cloud backup service architecture. The relationship among them is illustrated in Figure 3.

Cloud backup differs from traditional backup software in the following ways:

(1) Number of Users

Backup software is traditionally used in a LAN or WAN by specific user groups. Because the number of users is small, only a small number of storage servers are usually configured—for ease of deployment, ease of maintenance, and to reduce cost. User access paths to the servers are fixed, and dynamic assignment or adjustment is not required to meet different scenarios.

In contrast, cloud backup is designed for a large number of users in the WAN. As this number grows, the system has to be configured with multiple storage servers to meet scalability requirements. The system should also be capable of processing concurrent access requests from a large number of users, and assigning proper target storage servers to these users by means of efficient parallel
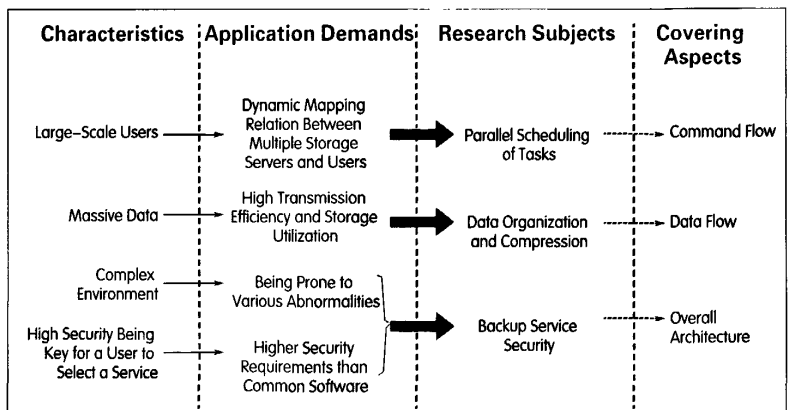
▲Figure 3. Relationship between application requirements and research subjects.

scheduling policy. Load balance and high storage utilization can thus be achieved among all storage servers. The process is completely transparent to users.

(2) Amount of Data

The difference in user numbers between cloud backup and traditional backup software represents a great difference in the amount of data to be processed. Backup data generated by large-scale users in the WAN may easily reach one TB or even one PB. Thus, proper data organization methods and compression algorithms are of great importance to improve the transmission and storage efficiency of massive amounts of data. The ultimate objective of these methods and algorithms is to improve system performance, reduce hardware costs and save energy.

(3) Service Security

Cloud backup must be compatible with heterogeneous data platforms of different backup clients; must ensure data integrity at the block, file and application levels; must adapt to a complex and changing WAN environment; and must guarantee data security.

Cloud backup systems have higher requirements on reliability than common backup software. However, people unconsciously feel it safer to backup critical data on visible devices. Doubts may arise about the security of backing up private data in faraway data centers.

On the one hand, cloud backup is prone to various abnormalities; on the other, users subjectively impose higher security requirements on cloud backup than on backup software. Security is therefore a pressing area of study for cloud backup.

According to the characteristics of cloud backup, the study of cloud backup focuses on:

(1) Command Flow

The B-Cloud system consists of backup clients, manager, and storage servers. The manager is the administrative center of the entire system, responsible for task scheduling, operation management, and status monitoring in the service process.

After receiving a service request from a backup client, the three parts of the system implement bidirectional security certification; and importantly, the manager completes job scheduling to establish a connection between the backup client and the storage server. At this point, the system begins to deliver the backup or recovery service.

(2) Data Flow

The backup or recovery data flow is transmitted between the backup client and the storage server without passing the manager. This method, whereby data goes directly from the source to the destination and does not pass any intermediate node, improves efficiency and also balances the overall load of the system. For data transmission and storage, backup data organization and compression are key.

(3) Service Security

Security of cloud backup involves the security within service platform, modules, and coordination and communication between modules. This issue requires in-depth discussion which is not intended here.

3 Conclusion

This paper introduces cloud storage technology and takes cloud backup as an example to discuss issues to be addressed in cloud storage application.

A trend of technology development is to satisfy requirements with services. Cloud storage conforms to this trend. However, in-depth study is needed into the implementation of cloud storage and its widespread applications.

References

[1] HAYES B. Cloud Computing [J]. Communications of the ACM, 2008, 51(7), 9–11.

[2] LIN G, DASMALCHI G, ZHU J. Cloud Computing and IT as a Service: Opportunities and Challenges [C]//Proceedings of the IEEE 6th International Conference on Web Services (ICWS'08), Sep 23–26, 2008, Beijing, China. Los Alamitos, CA, USA: IEEE Computer Society, 2008: 5

[3] NAMJOSHI J, GUPTE A. Service Oriented Architecture for Cloud Based Travel Reservation Software as a Service [C]//Proceedings of the 2009 IEEE International Conference on Cloud Computing (CLOUD'09), Sep 21–25,2009, Bangalore, India. Los Alamitos, CA, USA: IEEE Computer Society, 2009. 147–150

[4] LAPLANTE P A, ZHANG J, VOAS J. What's in a Name? Distinguishing Between SaaS and SOA [J]. IT Professional, 2008, 10(3): 46–50.

[5] CAMPBELL-KELLY M. Historical Reflections on the Rise, Fall, and Resurrection of Software as a Service [J] Communications of the ACM, 2009, 52(5): 28–30.

[6] 土庆波, 金涂, 何乐, 等. 虚拟化与云计算 [M]. 北京: 电子工业出版社, 2009
WANG Qingbo, JIN Xing, HE Le, et al Virtualization and Cloud Computing [M]. Beijing. Publishing House of Electronics Industry, 2009

[7] ZHU B, LI K, PATTERSON H. Avoiding the Disk Bottleneck in the Data Domain Deduplication File System [C]//Proceedings of the 6th USENIX Conference on File and Storage Technologies (FAST'08), Feb 26–29, 2008, San Jose, CA, USA. Berkeley, CA, USA: USENIX Association, 2008: 269–282

[8] 田敏, 代亚非. P2P持久存储研究综述 [J]. 软件学报, 2007, 18(6). 1379–1399.
TIAN Jing, DAI Yafei Study on Durable Peer-to-Peer Storage Techniques [J]. Journal of Software, 2007, 18(6): 1379–1399

[9] LILLIBRIDGE M, ELNIKETY S, BIRRELL A, et al. A Cooperative Internet Backup Scheme [C]// Proceedings of the 2003 USENIX Annual Technical Conference (USENIX '03), Jun 12–14, 2003, San Antonio, TX, USA Berkeley, CA, USA USENIX Association, 2003: 29–41.

[10] PAMIES-JUAREZ L, GARCIA-LOPEZ P, SANCHEZ-ARTIGAS M Rewarding Stability in Peer-to-Peer Backup Systems [C]//Proceedings of 16th IEEE International Conference on Networks (ICON'08), Dec 12–14, 2008, New Delhi, India. Piscataway, NJ, USA: IEEE, 2008: 6p.

Biographies

Zhou Ke

Zhou Ke received his doctoral degree from Huazhong University of Science and Technology (HUST). He is now a professor and a doctoral advisor at HUST, mainly engaged in researching information storage technologies. He has published over 50 papers.

Wang Hua

Wang Hua received her doctoral degree from Huazhong University of Science and Technology (HUST). She is now a lecturer at HUST, and is engaged in researching network storage and its application, and software engineering.

Li Chunhua

Li Chunhua received her doctoral degree from Huazhong University of Science and Technology (HUST). She is now an associate professor at HUST, and is mainly engaged in researching data security.