ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 151-154

Recent Advances in Cloud Storage

Jiyi Wu[1,2], Jianlin Zhang[1],Zhijie Lin[2,3], Jiehui Ju[3]
1. Key Lab of E-Business and Information Security, Hangzhou Normal University,Hangzhou,China
Email:Dr_PMP@yahoo.com.cn
2.School of Computer Science and Technology, Zhejiang University,Hangzhou,China
3.School of Information and Electronic Engineering, Zhejiang University of Science and Technology,Hangzhou,China
Email: zhangjohn@vip.sina.com, jianqing_fu@zju.edu.cn, lin@zju.edu.cn

***Abstract*—As the latest development of the distributed storage technology, Cloud storage is the product of the integration of distributed storage and virtualization technologies. Cloud storage is a method that allows you to use storage facilities available on the Internet. There are ten critical common denominators that must be considered to make cloud storage valuable. A typical cloud storage system architecture includes a master control server and several storage servers. With the advent of cloud computing, multi-tenancy has simply been extended to include any cloud architecture—that supports multiple tenants.**

***Index Terms*—Recent Advances, Cloud Computing, Cloud Storage, Multi-Tenancy, reference model**

I. Introduction

One of IT's biggest expenses is disk storage. ComputerWorld estimates that in many enterprises storage is responsible for almost 30% of capital expenditures as the average growth of data approaches close to 50% annually in most enterprise. Amid this milieu, there's strong concern that enterprise will drown in the expense of storing data, especially unstructured data.

To address this need, Cloud storage services have started to become popular. Ranging from Cloud storage focused at the enterprise to that focused on end users, Cloud storage providers offer huge capacity cost reductions, the elimination of labor required for storage management and maintenance, and immediate provisioning of capacity at a very low cost per terabyte.

Cloud storage, though, is not a brand new concept. The central ideas for Cloud storage are related to past service bureau computing paradigms and to those of application service providers and storage service providers of the late 90's.

This time, however, the economic situation and the advent of new technologies have sparked strong interest in the Cloud storage provider model. With on-premises storage costs already high and rising in many IT departments, Cloud storage providers can lower cost by off-loading the burden of storage management and shielding enterprises from other costs as well, such as storage and network hardware changes. Cloud storage providers deliver economies of scale by using the same storage capacity to meet the needs of many organizations, passing the cost savings to their customer base.

Corresponding Author: Jiyi WU

Cloud Storage is part of a wider definition called Cloud Computing which, according to the National Institute of Standards and Technology, is "a model for enabling convenient, on demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".

The service models are divided in Cloud Software as a Service (SaaS), Cloud Platform as a Service (PaaS) and Cloud Infrastructure as a Service (IaaS).

Computing resources like servers and network can be replaced, but the core of most of the organizations is the information, usually stored in data centers. For this reason security and availability are the first issues when companies are deciding to migrate part of their data to the cloud, generally by the internet.

This kind of precaution is not so different from the one when data is stored in private data centers, but there are some analysis concerned to this migration to public cloud that need to done by corporations and service providers.

II. Cloud Storage Infrastructure Requirements

When you combine the technology trends such as virtualization with the increased economic pressures, exploding growth of unstructured data and regulatory environments that are requiring enterprises to keep data for longer periods of time, it is easy to see the need for a trustworthy and appropriate storage infrastructure. Whether a cloud is public or private, the key to success is creating a storage infrastructure in which all resources can be efficiently utilized and shared.

Because all data resides on the storage systems, data storage becomes even more crucial in a shared infrastructure model. There are ten critical common denominators that must be considered to make cloud storage valuable. These include:

A. Elasticity

Cloud storage must be elastic to rapidly adjust the underlying infrastructure to changing subscriber demands and comply with Service Level Agreements (SLAs).

B. Automatic

Cloud storage must have the ability to be automated so that policies can be leveraged to make underlying infrastructure changes such as placing user and content management in different storage tiers and geographic locations quickly and without human intervention.

C. Scalability

Cloud storage needs to scale quickly and to tremendous capacities. This translates into scalability across objects, performance, users, clients, and capacity with a single name space across all storage capacity being critical for low Opex reasons.

D. Data Security

For private clouds, security is assumed to be tightly controlled. For public clouds, data should either be stored on a partition of a shared storage system, or cloud storage providers must establish multi-tenancy policies to allow multiple business units or separate companies to securely share the same storage hardware.

E. Performance

A proven storage infrastructure providing fast, robust data recovery is an essential element of a cloud service.

F. Reliability

Enterprise users also want to make sure that their data is reliably backed up for disaster recovery purposes and that it meets pertinent compliance guidelines.

G. Ease of Management

The need for improved manageability in the face of exploring storage capability and costs is a major benefit enterprises are expecting from cloud storage deployment.

H. Ease of Data Access

Ease of access to data in the cloud is critical in enabling seamless integration of cloud storage into existing enterprise workflows and to minimize the learning curve for cloud storage adoption.

I. Energy Efficiency

IT datacenters are growing bottlenecks and approaching ceilings on available power, cooling and flooring space. Green storage technology is the technology that enables energy efficiency and waste reduction in storage solutions leading to an overall lower carbon footprint.

J. Latency

Not all applications are suitable for a Cloud storage model. It is important to measure and test network latency before committing to a migration. Virtual machines can introduce additional latency through the time-sharing nature of the underlying hardware and unanticipated sharing and reallocation of machines can significantly affect run times.

III. MULTI-TENANCY CLOUD STORAGE REFERENCE MODEL

A. Typical cloud storage system architecture

A typical cloud storage system architecture includes a master control server and several storage servers, as shown in Fig 1.

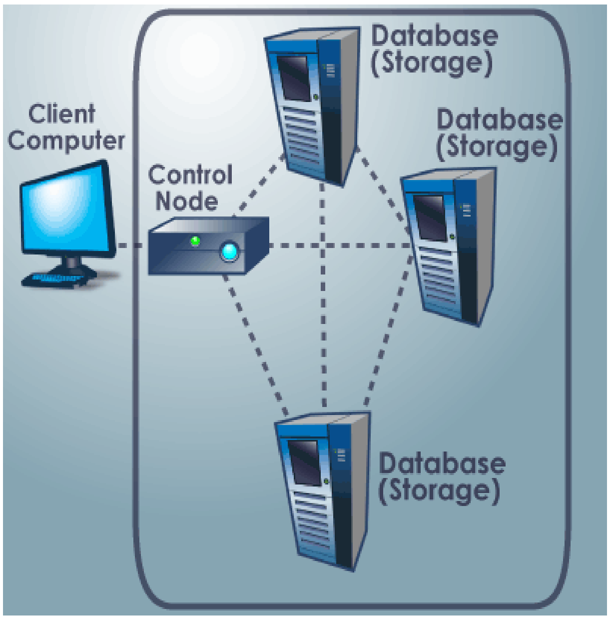

Figure 1. A typical Cloud Storage system architecture

For some computer owners, finding enough storage space to hold all the data they've acquired is a real challenge. Some people invest in larger hard drives. Others prefer external storage devices like thumb drives or compact discs. Desperate computer owners might delete entire folders worth of old files in order to make space for new information. But some are choosing to rely on a growing trend: cloud storage.

While cloud storage sounds like it has something to do with weather fronts and storm systems, it really refers to saving data to an off-site storage system maintained by a third party. Instead of storing information to your computer's hard drive or other local storage device, you save it to a remote database. The Internet provides the connection between your computer and the database.

On the surface, cloud storage has several advantages over traditional data storage. For example, if you store your data on a cloud storage system, you'll be able to get to that data from any location that has Internet access. You wouldn't need to carry around a physical storage device or use the same computer to save and retrieve your information. With the right storage system, you could even allow other people to access the data, turning a personal project into a collaborative effort.

So cloud storage is convenient and offers more flexibility, but how does it work? Find out in the next section.

B. Cloud Storage reference model

The appeal of cloud storage is due to some of the same attributes that define other cloud services: pay as you go, the illusion of infinite capacity (elasticity), and the simplicity of use/management. It is therefore important that any interface for cloud storage support these attributes, while allowing for a multitude of business cases and offerings, long into the future.

The model created and published by the Storage Networking Industry Association™ ,shows multiple types of cloud data storage interfaces able to support both legacy and new applications. All of the interfaces allow storage to be provided on demand, drawn from a pool of resources. The capacity is drawn from a pool of storage capacity provided by storage services. The data services are applied to individual data elements as determined by the data system metadata. Metadata specifies the data requirements on the basis of individual data elements or on groups of data elements (containers).

As shown in Fig 2, the SNIA Cloud Data Management Interface (CDMI) is the functional interface that applications will use to create, retrieve, update and delete data elements from the cloud. As part of this interface the client will be able to discover the capabilities of the cloud storage offering and use this interface to manage containers and the data that is placed in them. In addition, metadata can be set on containers and their contained data elements through this interface.

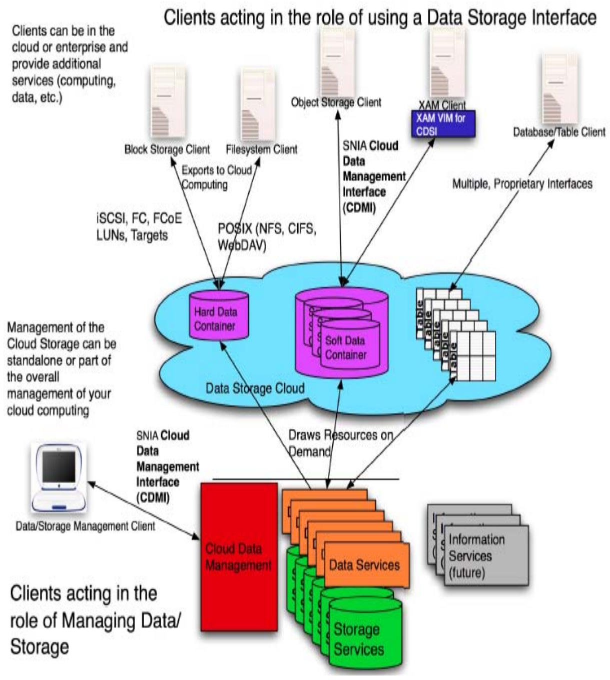

Figure 2.Cloud Storage reference model

It is expected that the interface will be able to be implemented by the majority of existing cloud storage offerings today. This can be done with an adapter to their existing proprietary interface, or by implementing the interface directly. In addition, existing client libraries such as XAM can be adapted to this interface as show in Figure 2.

This interface is also used by administrative and management applications to manage containers, accounts, security access and monitoring/billing information, even for storage that is accessible by other protocols. The capabilities of the underlying storage and data services are exposed so that clients can understand the offering.

Conformant cloud offerings may offer a subset of either interface as long as they expose the limitations in the capabilities part of the interface.

C. Multi-Tenancy Cloud Storage

The terms multi-tenant and multi-tenancy are not new; both have been used to describe application architectures designed to support multiple users or "tenants" for many years. With the advent of cloud computing, this terminology has simply been extended to include any cloud architecture—or infrastructure element within that architecture (application, server, network, storage)—that supports multiple tenants. Tenants could be separate companies, or departments within a company, or even just different applications.

To provide "secure" multi-tenancy and address the concerns of cloud skeptics, a mechanism to enforce separation at one or more layers within the infrastructure is required:

- **Application layer.** A specially written, multi-tenant application or multiple, separate instances of the same application can provide multi-tenancy at this level.
- **Server layer.** Server virtualization and operating systems provide a means of separating tenants and application instances on servers and controlling utilization of and access to server resources.
- **Network Layer.** Various mechanisms, including zoning and VLANs, can be used to enforce network separation. IP security (IPsec) also provides network encryption at the IP layer (application independent) for additional security.
- **Storage Layer.** Mechanisms such as LUN masking and SAN zoning can be used to control storage access. Physical storage partitions segregate and assign resources (CPU, memory, disks, interfaces, etc.) into fixed containers.

Achieving secure multi-tenancy may require the use of one or more mechanisms at each infrastructure layer.

While mechanisms to support multi-tenancy and enforce separation exist at every infrastructure layer, this paper is primarily concerned with storage and the requirements for secure and effective storage multi-tenancy in a cloud environment. To understand the full set of storage requirements, it is necessary to consider cloud storage from both the perspective of the tenant (user) and the provider of cloud services.

Cloud computing services can be broken down into a variety of types, ranging from Software as a Service (SaaS)—in which the provider delivers specific application services to each tenant—to Data storage as a Service (DaaS) —which is virtualized storage on demand over a network. Regardless of the type of cloud service, from a tenant perspective there will be specific requirements that apply directly or indirectly to data storage.

Tenant requirements are typically defined in terms of service level agreements (SLAs), which cover a variety of capabilities including:

- Security
- Performance
- Data protection and availability
- Data management

From the provider's perspective, multi-tenant storage should provide convenient mechanisms for satisfying these and other tenant SLAs as well as supporting additional capabilities such as:

- **Accounting.** The ability to monitor usage by each tenant for billing or other purposes.
- **Self service.** The ability to allow a tenant to perform a defined set of management tasks on their data and the storage they use, thereby offloading these functions from the provider.
- **Non-disruptive upgrades and repairs.** Downtime in multi-tenant environments may be difficult or impossible to schedule, so maintenance activities must be possible without incurring downtime from the point of view of the tenant.
- **Performance management.** The ability to balance cost and performance as the lifecycle requirements of data changes over time.

Designed to enable multi-tenant storage offerings, the SNIA's Cloud Data Management Interface (CDMI) for cloud storage and data management integrates and is interoperable with various types of client applications. CDMI offers a standard approach to data portability, compliance and security, as well as the ability to connect one cloud provider to another, enabling compatibility between cloud vendors.

Using this approach, a client will be able to discover the capabilities of cloud storage and use this interface to manage data containers and the data elements that are placed in them. CDMI makes extensive use of metadata to simplify application access and enable multiple levels of service as required by a diverse set of users.

In the storage layer, the CDMI interface can simplify management since data system metadata can be applied to container hierarchies. For the functional data path interface for data storage, CDMI assigns each data object a separate URI (Uniform Resource Identifier). Since objects can be fetched using the standard HTTP protocol employing RESTful (REpresentational State Transfer) operations, each data element can be managed as a separate resource. In this way, it is possible to separate and classify data elements and containers for secure access as well as service levels. The result is a level of isolation suitable to tenant based, on-demand data access.

VI. CONCLUSIONS AND FUTURE WORK

Cloud Storage with a great deal of promise, aren't designed to be high performing file systems but rather extremely scalable, easy to manage storage systems. They use a different approach to data resiliency, Redundant array of inexpensive nodes, coupled with object based or object-like file systems and data replication (multiple copies of the data), to create a very scalable storage system.

This article gives a quick introduction to cloud storage. It covers the key technologies in Cloud Computing and Cloud Storage,several different types of clouds services, and describes the advantages and challenges of Cloud Storage after the introduction of the Cloud Storage reference model.

ACKNOWLEDGMENT

Funding for this research was provided in part by the Scientific Research Program of Zhejiang Educational Department under Grant No.20071371. We like to thank anonymous reviewers for their valuable comments.

REFERENCES

[1] Luis M.Vaquero,Luis Rodero-Merino, Juan Caceres,Maik Lindner. A Break in the Clouds: Toward a Cloud Definition. ACM SIGCOMM Computer Communication Review,2009,39(1):50-55.

[2] Wu Jiyi,Ping Lingdi,Pan Xuezeng.Cloud Computing: Concept and Platform, Telecommunications Science, 12:23-30, 2009.

[3] Jonathan Strickland.How Cloud Storage Works[OL], http://communication.howstuffworks.com/cloud-storage.htm, 2010.

[4] Storage Networking Industry Association.Cloud Storage Reference Model,Jun.2009.

[5] Storage Networking Industry Association.Cloud Storage for Cloud Computing,Jun.2009.

[6] Luiz Andre Barroso,Jeffrey Dean,Urs Holzle.Web search for a planet: The Google cluster architecture. IEEE Micro, 2003,23(2):22−28.

[7] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung. The Google file system. In: Proc. of the 19th ACM SOSP. New York: ACM Press, 2003. 29−43.

[8] Robert L.Grossman, Yunhong Gu, Michael Sabala,Wanzhi Zhang. Compute and storage clouds using wide area high performance networks. Future Generation Computer Systems, 2009,25(2):179-183.

[9] Yunhong Gu and Robert L.Grossman. Sector and Sphere: the design and implementation of a high-performance data cloud. Philosophical Transactions of the Royal Society. A(2009)367:2429-2445.

[10] Robert L Grossman,Yunhong Gu. Data Mining Using High Performance Data Clouds: Experimental Studies Using Sector and Sphere. In Proc. of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining,2008, 920-927.

[11] Daniel J. Abadi. Data Management in the Cloud: Limitations and Opportunities. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2009,32(1):3-12.

[12] Peter Mell and Tim Grande. NIST. Retrieved from http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc,2010.

[13] S Lesem. Cloud Storage Strategy Retrieved from http://cloudstoragestrategy.com/2009/07/security-and-cloud-storage-everybody-talksabout-it-but-is-it-really-all-that-different.html,2010.