2010 International Conference on Intelligent Computing and Cognitive Informatics

Cloud Storage as the Infrastructure of Cloud Computing

Jiyi WU[1,2], Lingdi PING[1], Xiaoping GE[3],Ya Wang[4], Jianqing FU[1]
1.School of Computer Science and Technology, Zhejiang University
2.Alibaba Business College, Key Lab of E-Business and Information Security,Hangzhou Normal University
3.Department of Electronics Information,Hangzhou Academy of Scientific Professional Technology
4.Zhejiang Second Institute of Surveying and Mapping
Hangzhou, P.R.China
e-mail: Dr_PMP@yahoo.com.cn

***Abstract*—As an emerging technology and business paradigm, Cloud Computing has taken commercial computing by storm. Cloud computing platforms provide easy access to a company's high-performance computing and storage infrastructure through web services. With cloud computing, the aim is to hide the complexity of IT infrastructure management from its users. At the same time, cloud computing platforms provide massive scalability, 99.999% reliability, high performance, and specifiable configurability. These capabilities are provided at relatively low costs comparedto dedicated infrastructures.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.**

Keywords-Cloud Computing; Cloud Storage; architecture; service; virtualization

I. INTRODUCTION

Cloud computing portends a major change in how to store information and run applications. Instead of running programs and data on an individual desktop computer, everything is hosted in the "cloud"—a nebulous assemblage of computers and servers accessed via the Internet. Cloud computing lets you access all your applications and documents from anywhere in the world, freeing you from the confines of the desktop and making it easier for group members in different locations to collaborate.

Providers such as Amazon, Google, Salesforce, IBM, Microsoft, and Sun Microsystems have begun to establish new data centers for hosting Cloud computing applications in various locations around the world to provide redundancy and ensure reliability in case of site failures. Since user requirements for cloud services are varied, service providers have to ensure that they can be flexible in their service delivery while keeping the users isolated from the underlying infrastructure. Recent advances in microprocessor technology and software have led to the increasing ability of commodity hardware to run applications within Virtual Machines (VMs) efficiently. VMs allow both the isolation of applications from the underlying hardware and other VMs, and the customization of the platform to suit the needs of the end-user. Providers can expose applications running within VMs, or provide access to VMs themselves as a service (e.g. Amazon Elastic Compute Cloud) thereby allowing consumers to install their own applications.While convenient, the use of VMs gives rise to further challenges such as the intelligent allocation of physical resources for managing competing resource demands of the users.

One of the primary uses of cloud computing is for data storage. With cloud storage, data is stored on multiple third-party servers, rather than on the dedicated servers used in traditional networked data storage.When storing data, the user sees a virtual server—that is, it appears as if the data is stored in a particular place with a specific name. But that place doesn't exist in reality. It's just a pseudonym used to reference virtual space carved out of the cloud. In reality, the user's data could be stored on any one or more of the computers used to create the cloud. The actual storage location may even differ from day to day or even minute to minute, as the cloud dynamically manages available storage space. But even though the location is virtual, the user sees a "static" location for his data—and can actually manage his storage space as if it were connected to his own PC.

Cloud storage has both financial and security-associated advantages. Financially, virtual resources in the cloud are typically cheaper than dedicated physical resources connected to a personal computer or network. As for security, data stored in the cloud is secure from accidental erasure or hardware crashes, because it is duplicated across multiple physical machines; since multiple copies of the data are kept continually, the cloud continues to function as normal even if one or more machines go offline. If one machine crashes, the data is duplicated on other machines in the cloud.

II. KEY TECHNOLOGIES

A. Cloud Computing

Cloud computing is about moving services, computation or data—for cost and business advantage—off-site to an internal or external, location-transparent, centralized facility or contractor. By making data available in the cloud, it can be more easily and ubiquitously accessed, often at much lower cost, increasing its value by enabling opportunities for enhanced collaboration, integration, and analysis on a shared common platform.

Depending on the type of provided capability, there are four scenarios where Clouds are used as showed in Fig.1:

1) Infrastructure as a Service

IPs manage a large set of computing resources, such asstoring and processing capacity. Through virtualization, they are able to split, assign and dynamically resize these re-sources to build ad-hoc systems as demanded by customers,the SPs. They deploy the software stacks that

978-0-7695-4014-6/10 $26.00 © 2010 IEEE
DOI 10.1109/ICICCI.2010.119

run their ser-vices. This is the Infrastructure as a Service (IaaS) scenario.

2) Platform as a Service

Cloud systems can offer an additional abstraction level:instead of supplying a virtualized infrastructure, they canprovide the software platform where systems run on. Thesizing of the hardware resources demanded by the execu-tion of the services is made in a transparent manner. Thisis denoted as Platform as a Service (PaaS). A well-known example is the Google Apps Engine.

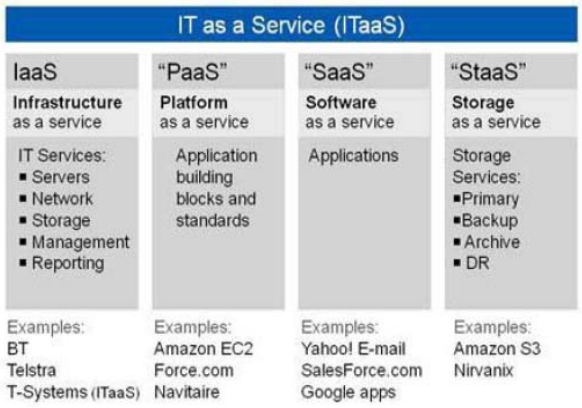

Figure 1. Cloud Computing service types with examples

3) Storage as a Service

Commonly known as Storage as a Service (StaaS),it facilitates cloud applications to scale beyond their limited servers. StaaS allows users to store their data at remote disks and access them anytime from any place.Cloud storage systems are expected to meet several rigorous requirements for maintaining users' data and information, including high availability, reliability, performance, replication and data consistency;but because of the conflicting nature of these requirements, no one system implements all of them together.

4) Software as a Service

Finally, there are services of potential interest to a widevariety of users hosted in Cloud systems. This is an alter-native to locally run applications. An example of this isthe online alternatives of typical office applications such asword processors. This scenario is called Software as a Ser-vice (SaaS).

B. Cloud Storage

Cloud storage is amorphous today, with neither a clearly defined set of capabilities nor any single architecture. Choices abound, with many traditional hosted or managed service providers (MSP) offering block or file storage, usually alongside traditional remote access protocols or virtual or physical server hosting. Other solutions have emerged, typified by the Amazon S3 service, that resembles flat databases designed to store large objects.

The Taneja Group defines cloud storage as a specific category within the larger field of "storage in the clou" solutions. Storage in the cloud encompasses traditional hosted storage, including offerings accessed by FTP, WebDAV, NFS/CIFS, or block protocols either remotely or from within a hosted environment. Cloud storage is an evolution of this hosted storage technology that wraps more sophisticated APIs, namespaces, file or data location virtualization, and management tools, around storage. Fig 2. shows the evolution of Cloud Storage based on traditional network srorage and hosted storage.

There are hundreds of different cloud storage systems. Some have a very specific focus, such as storing Web e-mail messages or digital pictures. Others are available to store all forms of digital data. Some cloud storage systems are small operations, while others are so large that the physical equipment can fill up an entire warehouse. The facilities that house cloud storage systems are called data centers.

At its most basic level, a cloud storage system needs just one data server connected to the Internet. A client (e.g., a computer user subscribing to a cloud storage service) sends copies of files over the Internet to the data server, which then records the information. When the client wishes to retrieve the information, he or she accesses the data server through a Web-based interface. The server then either sends the files back to the client or allows the client to access and manipulate the files on the server itself.

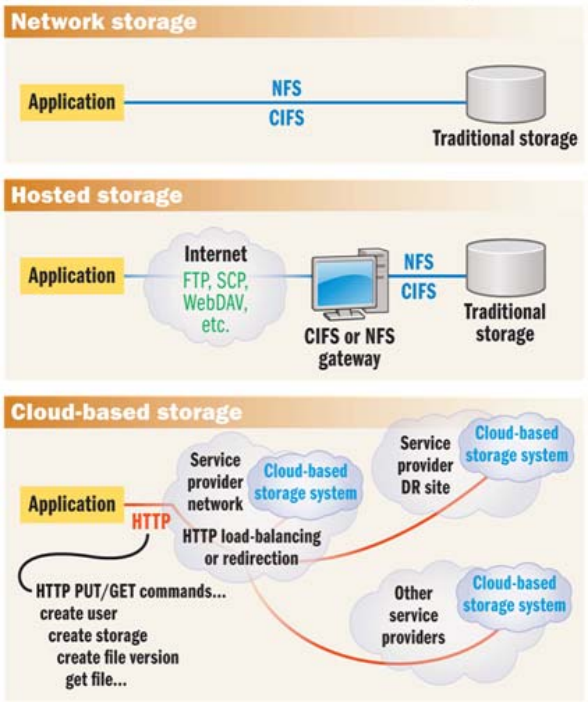

Figure 2.Evolution of Cloud Storage

Comedian George Carlin has a routine in which he talks about how humans seem to spend their lives accumulating "stuff". Once they've gathered enough stuff, they have to find places to store all of it. If Carlin were to update that routine today, he could make the same observation about computer information. It seems that everyone with a computer spends a lot of time acquiring data and then trying to find a way to store it.

A typical cloud storage system architecture includes a master control server and several storage servers,as shown in Fig 3.

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.

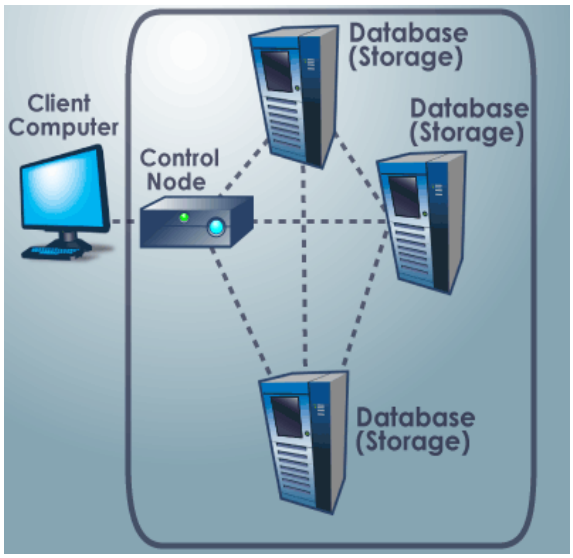

Figure 3. A typical Cloud Storage system architecture

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.

III. 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 4, 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.

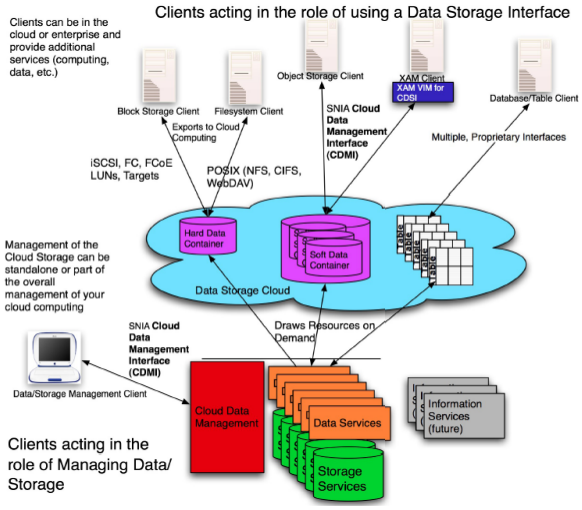

Figure 4.Cloud Storage reference model

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 4.

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.

IV. ADVANTAGES AND CHALLENGES

With everything, as has been said before, the devil is in the details. Certainly, there are more examples of the

growing popularity of cloud storage and valid business reasons for its popularity. Here are five key benefits of using cloud storage and of applications that take advantage of storage in the cloud.

- **Ease of management:** The maintenance of the software, hardware and general infrastructure to support storage is drastically simplified by an application in the cloud. Applications that take advantage of storage in the cloud are often far easier to set up and maintain than deploying an equivalent service on premise. At the customer site, often all that is needed to manage your storage implementation is a simple web browser leaving the headaches to the service provider.
- **Cost effectiveness:**For total cost of ownership, cloud storage is a clear winner. Elimination of the costly systems and the people required to maintain them typically provides organizations with significant cost savings that more than offset the fees for cloud storage. The costs of being able to provide high levels of availability and the scalability an organization needs are also unmatched. The economies of scale achieved by data centers simply can't be matched by all but the very largest of organizations.
- **Lower impact outages and upgrades:**Typically cloud computing provides cost effective redundancies in storage hardware. This translates into uninterrupted service during a planned or unplanned outage. This is also true for hardware upgrades which for the end user will no longer be visible.
- **Disaster preparedness:** Off site storage isn't new. Keeping important data backed up off site has been the foundation of disaster recovery since the inception of the tape drive. Cloud storage services not only keep your data off premise, but they also make their living at ensuring that they have redundancy and systems in place for disaster recovery.

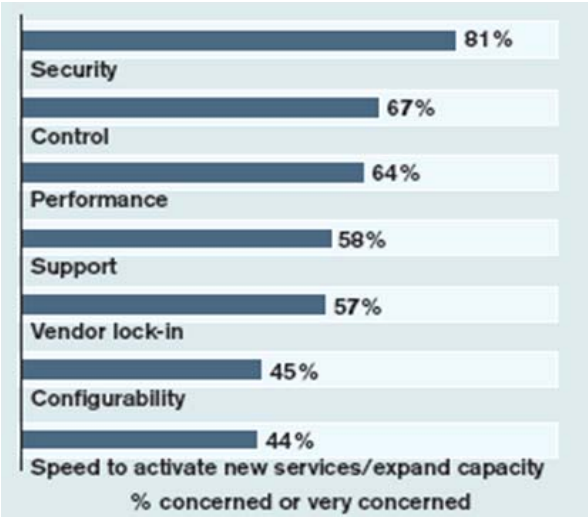

Figure 5. Survey on concerns with cloud storge services

- **Simplified planning:** Cloud storage solutions free the IT manager from detailed capacity planning. Cloud-based solutions are flexible and provide storage as needed. This eliminates the need to over provision for storage that may be needed to meet

However, with every type of cloud storage, there are challenges in the implementation (i.e. the devil is in the details). These challenges include:

- Security (always an issue and not necessarily a cloud storage specific issue)
- Data integrity (making sure the stored data is "correct")
- Power (since you have copies you will have extra storage which adds power)
- Replication time and costs (how fast can you replicate data since this can be important to data resiliency)
- Cost (how much extra money do you have to pay to buy the extra storage for copies)
- Reliability

As show in Fig 5, the problem include security,control, performance,support,vendor lock-in, are concerned by users with cloud storge services.

V. 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] Wikipedia. Cloud computing .Retrieved from http://en.wikipedia.org/wiki/Cloud_computing,2010.

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

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

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

[6] Jeffrey B.Layton.Cloud Storage Concepts and Challenges[OL], http://www.linux-mag.com/cache/7617/1.html,2010.

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