A Survey on Cloud Storage

Swathi.C, Vishnupriya. S, Rajeshkumar.R

Sri Eshwar college of Engineering, Kinathukadavu, Coimbatore, India[1, 2, 3]

ABSTRACT— As technology in the cloud increases, there has been a lot of improvements in the maturity and firmness of cloud storage technologies. Many end-users and IT managers are getting very excited about the potential benefits of cloud storage, such as being able to store and retrieve data in the cloud and capitalizing on the promise of higher-performance, more scalable and cut-price storage. In this thesis, we present a typical Cloud Storage system architecture, a referral Cloud Storage model and Multi-Tenancy Cloud Storage model, value the past and the state-of-the-art of Cloud Storage, and examine the Edge and problems that must be addressed to implement Cloud Storage. Use cases in diverse Cloud Storage offerings were also abridged.

KEYWORDS—Cloud Storage, Cloud Computing, referral model, Multi-Tenancy, survey

I. INTRODUCTION

One of IT's biggest expenses is disk storage. Computer World estimates that in many enterprises storage is responsible for almost 30% of capital investment as the average growth of data approaches close to 50% annually in most enterprise. Among this background, there's strong concern that enterprise will drown in the expense of storing data, mainly in 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 reduction 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.

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 offering convenient, on demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, and other services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".

The service models are splitted up 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, generally 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.

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 span of time, it is easy to see the need for a trustworthy and appropriate storage infrastructure. In case 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 pivotal in a shared infrastructure model. There are ten key common character that must be considered to make cloud storage valuable. These include:

A. Elasticity

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

B. Automatic

Cloud storage must have the aptness to be programmed 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 huge capacities. This translates into scalability across objects, performance, 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 highly 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 evince storage infrastructure providing fast, robust data recovery is an essential element of a cloud service.

F. Reliability

Enterprise clients 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 highly 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 to access of 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 data centers 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-dividing nature of the underlying hardware and unanticipated sharing and reallocation of machines can significantly affect run times.

III. MULTI-TENANCY CLOUD STORAGE REFERRAL 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.
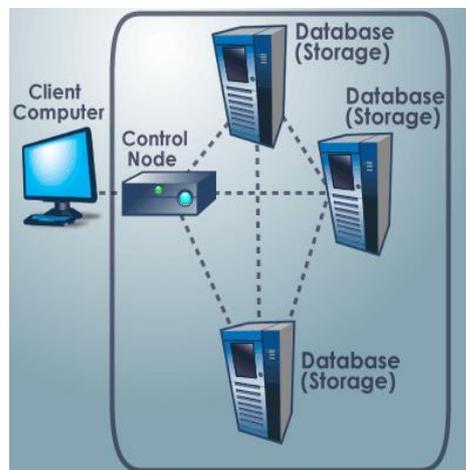
Figure 1. A typical Cloud Storage system architecture

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

While cloud storage sounds like it has something to do with weather fronts and storm systems, it refers to saving data to an off-site storage system maintained by a third party. Inspite of storing information to user computer's hard drive or other local storage device, user save it to a remote databases. The Internet provides the connection between user computer and the database.
So cloud storage is more convenient and offers more flexibility.
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 user go, the illusion of infinite capacity (elasticity), and the easy of management. It is therefore important that any interface for cloud storage support these attributes, while allowing for a multiple no of business cases and offerings, large 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 give storage to be provided on demand, obtained from a pool of resources. The limit is drawn from a pool of storage capacity provided by storage services. The data services are appeal to individual data elements as determined by the data system metadata. Metadata mention 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, update , retrieve and delete data elements from the cloud. As a 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 secondary, metadata can be set on containers and their contained data elements through this interface.
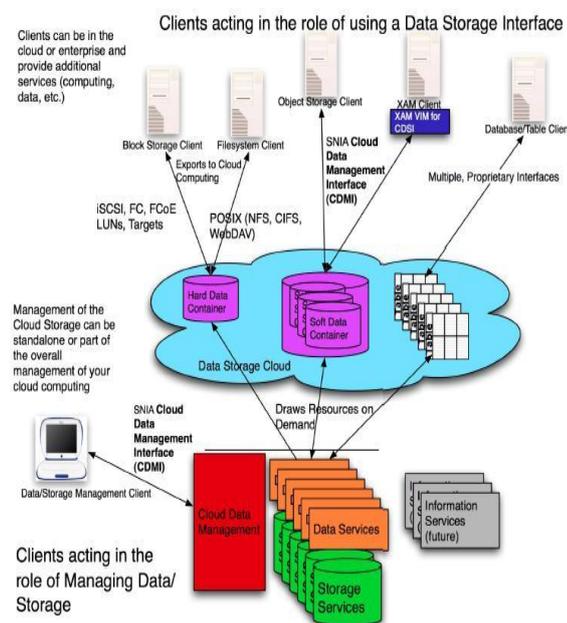
Figure 2.Cloud Storage reference model

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

Accord 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 long years. With the advent of cloud computing, this has simply been extended to include any cloud architecture—or infrastructure element within that architecture (application, network, server, storage)—that supports multiple users. Users 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, divided 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.

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.

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.

IV. CLOUD STORAGE USE CASES

In this part we will summarizes the use cases in various Cloud Storage offerings.
A. Web Facing Applications

Web facing applications will typically use a Cloud Storage offering that provides the data directly to the user's browser using a URL. The data is typed (MIME) and the browser invokes the appropriate application to view the data.

Media (audio, video) files are served as a stream of data, allowing use of parts of the data within the file without requiring all data in the file to have been received by the client.
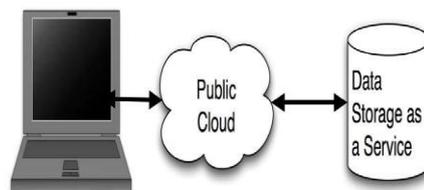
Figure 3. Data Storage as a Service

Social media sites include Myspace, Facebook, Twitter, Blogs, etc. Cloud Storage is used as a auxiliary storage space augmenting the web facing social application.

Pictures and content are stored in Cloud Storage (URL based typically). A content management system is used to keep track of additional metadata associated with the data. Smugmug is an example of this.

B. Unstructured Data Storage

This is a pre-allocated storage space (LUN, Filesystem) that is exported via standard client protocols (ex: WebDAV, NFS, CIFS), and "mounted" on a local machine. Normal POSIX semantics are available at that point for creating/reading/writing/deleting the files.

A number of vendors have offerings in this space. A sub case is cloud desktop. Examples include iCloud, ThinkGrid etc.

This is the ability to synchronize local client data, from multiple clients, with a Cloud Storage version. Changes are detected and then synchronization is done asynchronously and opportunistically. Access may or may not be through standard file protocols and URIs. Clients and servers have a way of sharing state describing what has/needs to be synced.
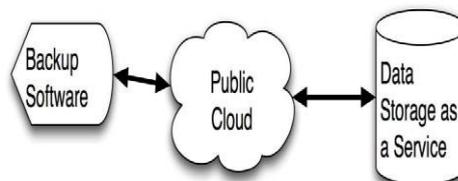
Figure 4. Backup to the Cloud

C. Backup to the Cloud

1)   Backup Software running on, some, local machines

– destination Cloud Storage.

This is local backup software or backup server using Cloud Storage as the destination of backup data.

Backup Software on each machine (i.e.☐ TimeFinder)

Backup Server based for multiple local machines

File Server Appliance locally with embedded☐ backup to Cloud

A common technique used by some local servers is to have the client computers turn on data sharing (i.e. becoming a CIFS server). Then having the local backup server become a CIFS client of the backup client and then backing up the data in that manner. This is an elegant way to circumvent having to install 3rd party backup software on the backup clients.

2) Backup of Cloud Computing Data

Backup of the data used in Cloud Computing (IaaS). Example(s): vSphere (includes de-dup as well)

3) Backup from one cloud provider to the other
This is the case of using a second cloud provider as the target of backup data from the first cloud provider.

4) Restore (i.e. Give me back all my laptop data)

D. Archive/Retention to the Cloud

This is the use case of using the Cloud Storage for archiving of data. Theoretically XAM should be an ideal interface for this. Iron Mountain - VFS (not XAM based) is an example of this.
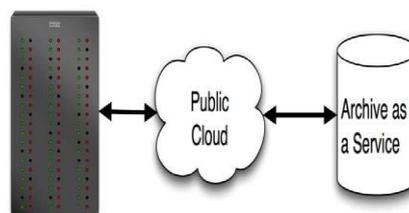
Figure 5. Archive as a Service

Considerations for the use case:

•   Does the user maintain a local copy?

- Retention Period: "Keep my files for X amount of time"

- Secure Deletion: "When it's gone, it's REALLY gone"

- eDiscovery : "Satisfy my subpoenas"

E. Preservation in the Cloud

Preservation is distinguished from Archive/Retention in that the goal of preservation is to actively maintain the upkeep of information, most likely for long periods of time.

F. Databases in the Cloud

Cloud Table Storage falls into the following categories:
1)   Horizontally Scalable, Object-Relational:Examples include Microsoft Azure Tables, Google BigTable, (hyperTable) , SimpleDB.

2)   Vertically Scalable, Traditional Relational: SQL services is an example.

3)   Document Model: Example(s): CouchDB is an example.

G. Storage for Cloud Computing

1)   Storage for IaaS

Traditional data storage which is accessible as part of the computing infrastructure. EC2 leverages S3 as if it were a private cloud.

- Image Storage

- Guest Auxiliary Storage

- Application Image Storage
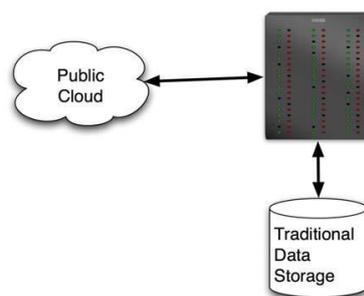
Figure 6. Storage for Cloud Computing

2) Storage for PaaS

This type of storage is not usually surfaced or manageable.
3) Storage for SaaS (Software as a Service)

This type of storage is not usually surfaced or manageable. A typical examples is Salesforce.com uses storage service from Aamzon S3.

H. Content Distribution

Distributing data globally for the purposes of decreasing latency and increasing scalability. Here list the following examples:

- "Hot" media serving – move to point of presence, replicate out to caches, then recover the resources when unused.
- Data transformation in route (i.e. localization,
NTSC->PAL)

This use case as a layering on top of the standard interfaces.

I. Cloud Storage Peering (i.e. "Intercloud" Storage)

This is the concept of having the Storage Clouds of different Cloud Storage Providers being able to interoperate between each other (in other words doing for Storage Clouds what the Internet did for separate, proprietary networks). Possible Characteristics:

- Shared storage and replication between cloud storage offerings.

- Distribute the data across cloud storage providers (possibly via a storage broker that provides a blended rate).
- Data may be erasure encoded as well as replicated.
- Caching and distribution between the client and "dumb storage" provider, geographic staging and replication.
- Activation and de-activation relative to some trust model – activation requires assembly from the erasure coded blocks, decoding and decryption, de-activation involve encryption, encoding and distribution.

- Topology of the network is more nuanced than the typical two tier processing model and more

    dynamic as well.

Examples include "Federated" Cloud Storage , "Cloud Exchange",Cloud "Bursting", offloading, Hybrid Internal/External Clouds .

V. ADVANTAGES AND CHALLENGES

A. Advantage of cloud storage

o   Ease of management

o   Cost effectiveness

o   Lower impact outages and upgrades

o   Disaster preparedness

B. Challenges in the implementation

However, with every type of cloud storage, there are challenges in the implementation (i.e. the devil is in the details).
1) Physical Security

2) Data encryption

3) Access Controls
    4) Service Level Agreements (SLA)

    5) Trusted Service Provider

Figure 7. Survey on concerns with cloud storge services

As show in Fig 7, the problem include security,control, performance,support,vendor lock-in, are concerned by users with cloud storge services.
These challenges include:

o   Security (always an issue and not necessarily a cloud storage specific issue)

o   Data integrity (making sure the stored data is "correct")
o   Power (since you have copies you will have extra storage which adds power)
o   Replication time and costs (how fast can you replicate data since this can be important to data resiliency)
o   Cost (how much extra money do you have to pay to buy the extra storage for copies)
o   Reliability

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.

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.