A Study on Cloud Storage: Amazon S3

Tanushree B S¹, Ms. Kokila²

Computer Engineering, Presidency University, Bengaluru, Karnataka
Assistant Professor - CSE, Presidency University, Bengaluru, Karnataka

---

ABSTRACT: Cloud computing has revolutionized the way businesses operate by providing on-demand access to computing resources and services over the Internet. Cloud storage is a key component of cloud computing, enabling users to store and access data from anywhere, anytime. In this survey paper, we provide an introduction to cloud computing and cloud storage, highlighting the need for cloud storage in today's data-driven world. We also discuss the architecture of cloud storage, consisting of a top, middle, and last layer, and key requirements in cloud storage, such as data durability, availability, and security. We then explore the advantages of cloud storage, including cost savings, scalability, and ease of use. One of the most popular cloud storage solutions is Amazon S3, which provides a highly durable, scalable, and secure storage infrastructure for businesses of all sizes. We delve into the concepts and architecture of Amazon S3, including bucket and object storage, access control, and versioning. We also explain why Amazon S3 is a compelling choice for businesses, and highlight its functionalities, such as Reduced Redundancy Storage and AWS Import/Export.

In summary, this survey paper provides a comprehensive overview of cloud computing and cloud storage, with a focus on Amazon S3. We hope that this paper will be a useful resource for researchers, practitioners, and businesses looking to leverage the benefits of cloud storage for their data management needs.

KEYWORDS: Cloud computing, Cloud Storage, Amazon S3, etc.

1. INTRODUCTION

In recent times, the use of data in computers has increased significantly, not just among individuals but also among organizations. This has led to the need for efficient and cost-effective methods of storing and sharing important data, while also ensuring global accessibility and continuous availability. Cloud computing has emerged as a solution to these challenges, allowing users to access and share data from anywhere, while enabling collaboration among different groups and individuals. Major companies like Google, Amazon, Microsoft, and others have established data centers around the world to ensure reliability and redundancy in case of site failures.

Cloud computing services are designed to meet the diverse needs of different users, with service providers offering flexibility while keeping users isolated from the backend infrastructure. One of the most significant uses of cloud computing is data storage, where data is stored on multiple third-party servers rather than on a single dedicated server. This approach offers several advantages, such as enhanced security, protection, and cost-effectiveness. In terms of security, cloud storage is more secure from hardware crashes or accidental loss, as data is stored on multiple machines, and even if one of them goes offline, others keep running, ensuring that data is not lost. Additionally, the virtual resources of cloud computing are generally more affordable and cheaper than physical resources, making it an attractive option for businesses and organizations.

Cloud computing is becoming an increasingly popular option for businesses and organizations, thanks to a range of benefits it offers. These benefits include reduced maintenance costs, as businesses can eliminate licensed software fees and reduce hardware and software purchases. Additionally, cloud computing allows for anytime, anywhere access to applications and data, which can improve productivity and flexibility. Cloud services are also scalable, enabling businesses to adjust their resources as needed, and offer greater flexibility in managing resources and data. Other benefits of cloud computing include disaster recovery capabilities, reduced capital expenditures through pay-per-use pricing models, user-friendly environmen6+ts, quick deployment, and reduced energy consumption. Overall, cloud computing can help businesses to streamline their operations, reduce costs, and improve efficiency.

This paper provides an overview of cloud storage principles, with a particular focus on Amazon S3 as a storage solution. The paper is structured as follows: The introduction section gives an overview of cloud storage, while section 3 explores the need for cloud storage. Section 4 examines the cloud storage architecture, while Section 5 outlines the advantages of cloud storage. Section 6 reviews the key requirements of a cloud storage system. The paper then turns to a discussion of Amazon S3 in section 7, providing an introduction to the service. Section 8 presents the concepts and architecture of Amazon S3. Section 9 explores why Amazon S3 should be used while section 10 discusses the Amazon S3 functionalities. Finally, section 10 concludes the paper. By the end of the

paper, readers will have a solid understanding of cloud storage principles and the Amazon S3 service in particular.

2. CLOUD STORAGE: AN OVERVIEW

The conventional method of storing data locally on a user's hard drive is no longer sufficient to meet the evolving needs of users who routinely handle large volumes of digital data. Such a method lacks the required scalability, high availability, and optimized resource allocation that users demand. In fact, it appears to have limitations in handling massive data volumes and modern workloads. Fortunately, cloud storage offers all these capabilities and more.

Cloud computing presents an opportunity to store data in a cloud storage system, instead of relying on a computer's local hard drive. Users can avoid the need to maintain large storage infrastructures by storing their data in remote data centres, which are controlled and managed by big companies like Apple, Microsoft, Google, and Amazon. Files saved in cloud storage can be accessed from anywhere using any device with an Internet connection. Cloud storage is a crucial component of cloud computing and is available as an online storage service provided by third-party vendors. Data is stored on virtualized pools of storage and is delivered on demand in a scalable and multi-tenant way.

Cloud storage is a service model that involves transmitting and storing data remotely on remote storage systems. The data is backed up and maintained on these remote storage systems and can be accessed by users over the network, such as the Internet. With cloud storage, users typically pay a monthly fee based on their consumption rates to the cloud storage providers.

There are alternative solutions to traditional storage, such as Amazon S3, which functions as a flat database designed for storing large objects. Cloud storage is an evolution of traditional hosted storage technologies, including offerings that can be accessed remotely by FTP, WebDAV, NFS/CIFS, or block protocols from within a hosted environment. Cloud storage incorporates more sophisticated APIs, namespaces, file or data location virtualization, and management tools into storage solutions. As shown in Figure 1, cloud storage has evolved from traditional network storage and hosted storage. There are numerous cloud storage systems available, with some focusing on specific applications like storing web emails or digital pictures, while others offer storage for all types of digital data. Some cloud storage systems are small operations, while others are so large that they require an entire warehouse to house the physical equipment. The facilities that house cloud storage systems are known as data centres.
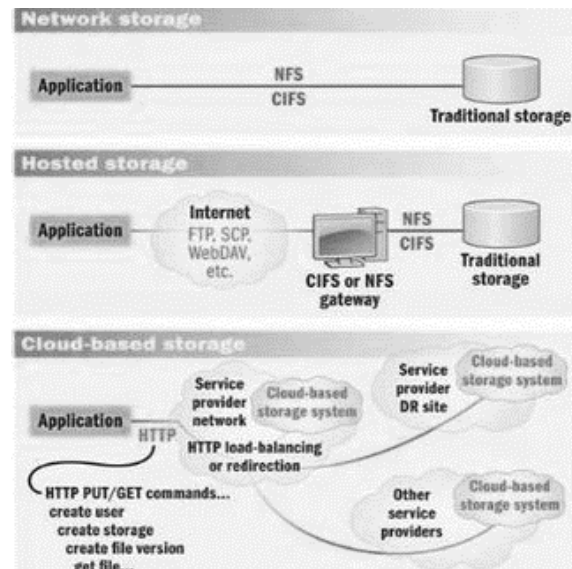

Figure 1.0 – Evolution of Cloud Storage

At its most basic level, a cloud storage system requires only one data server connected to the internet. A client 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, they access 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.

3. NEED FOR CLOUD STORAGE

Users deal with massive amounts of digital data daily and need the availability of data from any device, anywhere. This has led to the need for cloud storage, which can provide such availability.

- A growing dependency on Information and Communications Technology (ICT) is crucial for the success of businesses. ICT helps businesses respond to changes in the competitive market and gain flexibility.
- The explosion of digital data has made it easier to capture, alter, and store data. Companies generate massive amounts of data daily, and this data is essential for making smarter decisions.
- Users are actively creating and sharing different types of data, such as text, video, and photo postings, along with comments, tags, and ratings through blogs. This has created a need for new data sets and applications to deal with such data.
- The consumerization of IT has led users to work faster and easier using their own devices such as smartphones, laptops, and tablets. This has created a huge demand for massive and efficient storage that is accessible from any device, anywhere.
- Organizations need to store large quantities of digital data, and storage professionals are required to design, manage, and maintain the changing storage requirements. However, companies are struggling to find skilled storage professionals due to a lack of storage technology education.

- The availability of limited funds, economic slowdowns, and cuts in grants and subsidies have made people consider cost-effective alternatives for storing data. Cloud storage eliminates system costs and the need for people to maintain them while providing high levels of scalability and availability for organizations.
- Virtualization is the primary accelerator of cloud computing and cloud storage. It makes it possible to run multiple applications on virtual machines within the same physical server, delivering greater availability, scalability, and resource optimization for storage and servers.

4. CLOUD STORAGE ARCHITECTURE

The architecture of cloud storage can vary across different platforms, but it generally involves a large number of storage devices that are clustered by a distributed file system, network, and storage middleware. Cloud storage is designed to provide on-demand storage services to users in a scalable way, and it typically includes a front end that exports an API for storage access. While SCSI protocol is commonly used in traditional storage systems, cloud storage APIs are expanding to include file service front ends, web service front ends, and traditional front ends like iSCSI and internet SCSI. In cloud storage architecture, the front end is supported by a middleware layer called storage logic, that enables the implementation of data reduction and replication features. The back end, on the other hand, handles the physical storage of data, which can be a protocol that supports specific features or a conventional physical disk back end. Cloud storage typically includes a storage resource pool, Service Level Agreements (SLAs), distributed file systems, and service interfaces as key components.
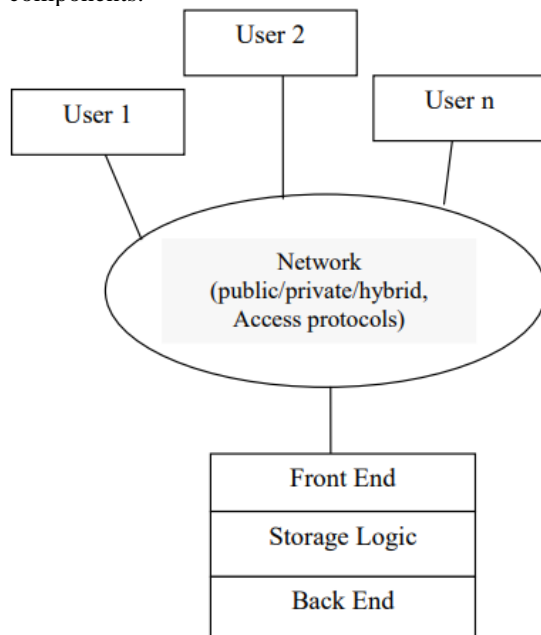

Figure 2.0 – Generic Cloud Storage Architecture

5. ADVANTAGES OF CLOUD STORAGE

The benefits of utilizing cloud storage are numerous and can be summarized as follows:

- Remote Backup: Cloud storage allows organizations to create off-site backups of their data at a relatively low cost.
- Universal Access: Data stored on the cloud can be accessed from anywhere in the world, as long as there is an internet connection.
- Data Security: Cloud storage services typically offer advanced security features, such as encryption and authentication, which help protect against cyber threats.
- Efficient Bandwidth Usage: Cloud storage allows for efficient use of bandwidth, as files can be shared through web links instead of being sent as email attachments.
- Simplified Management: Cloud storage providers handle the maintenance and management of the underlying infrastructure, hardware, and software. This allows users to easily manage their data through a web browser.
- Cost-Effective Solution: Cloud storage is often a more cost-effective solution compared to traditional storage methods. The availability and scalability of cloud services can be financially advantageous for organizations, as it eliminates the need for costly systems and personnel to maintain them.

6. KEY REQUIREMENTS IN CLOUD STORAGE

The popularity of cloud storage stems from its resemblance to other cloud services, such as pay-as-you-go pricing, the appearance of infinite capacity, and user-friendly management. As such, any interface for cloud storage must support these characteristics.

Cloud storage must meet various design requirements, including scalability, availability, security, multi-tenancy, reliability, speed, control, cost, and simplicity. Business requirements determine storage needs, which can scale up or down as necessary. Therefore, cloud storage must be scalable enough to accommodate requests from multiple concurrent users without impacting performance and speed.

Cloud storage services must be available 24/7, and fault-tolerant, decentralized techniques like replication must be used to enhance availability. To prevent a single point of failure, data must be replicated on different servers located in different locations, and multiple nodes must provide the same services. Backup nodes take over if the primary node fails.

Multi-tenancy implies that multiple users (tenants) use the same storage. Tenants must access their data without any interruptions. Therefore, data must be stored in a way that guarantees constant accessibility without any downtime. De-duplication and compression services can reduce storage space by eliminating redundant data, reducing the amount of data sent over the network, lowering storage consumption, and ultimately reducing user bills.

7. AMAZON S3

Amazon S3 is an internet-based storage service designed to simplify web-scale computing for developers. It offers a straightforward web services interface for storing and retrieving any amount of data from anywhere on the web at any time. This service provides developers access to the same reliable, secure, fast, and cost-effective infrastructure that Amazon uses to run its own web network globally. The goal is to leverage the benefits of scale and pass them on to developers.

However, before signing up with a cloud storage provider like Amazon, it's essential to understand a few things such as the concepts of objects and buckets in S3, S3 architecture, costs, and the best services for SMBs.

The fundamental building blocks of the Amazon Web Services (AWS) cloud storage service, Amazon S3, are objects and buckets.

- An object is a file uploaded to the Amazon cloud with several properties such as name, size, and date modified. The metadata is created automatically by the cloud service when a user creates a file, and each object can be retrieved by a URL, provided it's signed with the account owner's keys and a signing method used by Amazon.
- A bucket acts as a container for holding objects, and there is a limit on the number of buckets a user can create. Bucket names are unique to avoid naming clashes, and the buckets are backed up in different regions to provide a high level of availability.

This paper aims to provide you with all the necessary information to answer your most important questions about Amazon S3, including its concepts, architecture, costs, and best services for SMBs.

8. CONCEPTS AND ARCHITECTURE OF AMAZON S3

S3 organizes data using a two-level namespace where buckets, which are similar to folders, serve as the top level. Buckets have a unique global name and allow users to organize their data, identify the user to be charged for storage and data transfers, and serve as the unit of aggregation for audit reports. Each AWS account can have up to 100 S3 buckets, and each bucket can store an unlimited number of data objects. In the Amazon cloud, objects are characterized by a unique name, a data blob with a maximum size of 5GB, and metadata containing predefined entries and user-specified name/value pairs of up to 4KB. Users have the ability to create, edit, and access objects within buckets, with access being restricted according to access control policies. However, renaming an object or moving it to a different bucket requires downloading the entire object and writing it back with the new name. Search is limited to queries based on the object's name and a single bucket, and no metadata or content-based search capabilities are provided. S3 charges are based on storage volume, data transfer activity, and a per-transaction charge. All charges, regardless of the object owner or the user accessing the object, are directed to the owner of the bucket that stores the object generating the charges.

9. AMAZON S3: WHAT SETS IT APART

9.1 Affordable Pricing

Amazon AWS now offers a free usage tier that provides cost-effective services compared to other cloud systems.

9.2 Access Control Mechanisms

Amazon S3 provides various mechanisms for controlling access to your account, including:

9.2.1 Identity and Access Management (IAM USERS)

With AWS IAM, you can create multiple users and manage their permissions.

9.2.2 Access Control Lists (ACL)

Amazon ACL is applicable to every object and bucket and determines which groups or users can access certain objects and buckets.

9.2.3 Bucket Policies

Bucket Policies enable bucket owners to determine which permissions to allow or deny. Only the bucket owner has the authority to write bucket policies.

9.2.4 Query String Authentication

Query String authentication authenticates browsers or HTTP requests by signing the URL with the Owner's credentials and appending a signature that matches the signature computed by Amazon S3.

9.3 Multiple Storage Types

Amazon S3 offers various storage types with differences in availability, reliability, and storage regions. These types include Standard Storage, Reduced Redundancy, and Glacier Storage.

10. AMAZON S3 FUNCTIONALITY

Amazon S3 was intentionally designed with a basic set of features:

- This functionality enables users to create, retrieve, and remove objects that can range in size from 1 byte to 5 terabytes.
- There is no limit to the amount of items that can be stored, and each individual item is saved in a bucket and can be accessed through a distinct key assigned by the developer.
- Users have the option to choose a Region to optimize for latency, minimize costs, or address regulatory requirements. Amazon S3 is available in several Regions including US Standard, US West (Oregon), US West (Northern California), EU (Ireland), Asia Pacific (Singapore), Asia Pacific (Tokyo), South America (Sao Paulo), and GovCloud (US) Regions.
- Objects stored in a Region remain within the Region unless manually transferred. To prevent unauthorized access, various authentication mechanisms are available to safeguard the data's security. Users can choose to make objects private or public, and grant specific rights to individual users.
- Additionally, options for secure data upload/download and encryption of data at rest are available to provide additional data protection.

- Amazon S3 employs REST and SOAP interfaces that are based on industry standards and are compatible with any Internet development toolkit. The service is designed to be flexible, allowing for the easy addition of protocol or functional layers. By default, the download protocol is HTTP, but a Bit Torrent™ protocol interface is provided to reduce costs for high-scale distribution.
- Amazon S3 also includes options for recurring and high-volume deletions. For recurring deletions, rules can be defined to remove sets of objects after a pre-defined time period, and for efficient one-time deletions, up to 1,000 objects can be deleted with a single request.

10.1 DATA PROTECTION

Amazon S3 provides default security for data storage, allowing only bucket and object owners access to the resources created. The service offers multiple access control mechanisms and encryption options for secure data transit and storage, protecting data from logical and physical failures. These features guard against unintended user actions, application errors, and infrastructure failures, making them suitable for regulatory compliance standards like PCI and HIPAA. Amazon S3's various data security and reliability features work together to provide a comprehensive solution for safeguarding data.

10.2 DATA SECURITY DETAILS

Amazon S3 offers four access control mechanisms, including IAM policies, ACLs, bucket policies, and query string authentication, providing flexibility in controlling who can access your data and how they can access it. IAM enables multiple users to be managed under a single AWS account, and fine-grained control to be granted to Amazon S3 buckets or objects. ACLs allow the selective addition of certain permissions to individual objects. Bucket Policies can add or deny permissions across some or all objects within a single bucket, and Query string authentication allows sharing of Amazon S3 objects through URLs with predefined expiration times. Amazon S3 supports SSL-encrypted endpoints for secure data upload/download, multiple options for data encryption at rest, and logging of requests made against your Amazon S3 resources.

10.3 ENSURING DATA AVAILABILITY AND CONSISTENCY

Amazon S3 is a storage infrastructure designed for the primary storage of critical data. It provides a high level of durability through the redundancy of objects stored on multiple devices across various facilities in an Amazon S3 Region. To maintain durability, data is stored across multiple facilities through synchronous PUT and COPY operations before returning SUCCESS. Amazon S3 frequently verifies data integrity using checksums and repairs lost redundancy quickly.

Amazon S3's standard storage is designed to sustain concurrent data loss in two facilities and to provide 99.999999999% durability and 99.99% availability of objects over a year. The Amazon S3 service is supported by a Service Level Agreement (SLA) to guarantee its reliability and performance. Versioning is another feature that Amazon S3 offers to preserve, retrieve, and restore every version of every object stored in an Amazon S3 bucket. This feature allows for easy recovery from unintended user actions or application failures. By default, requests will retrieve the most recent version, but storage rates will apply for every version stored.

10.4 REDUCED REDUNDANCY STORAGE (RRS)

Amazon S3 provides Reduced Redundancy Storage (RRS) as a cost-effective solution for customers to store non-critical, reproducible data at lower levels of redundancy compared to Amazon S3's standard storage. RRS is suitable for sharing or distributing content that is already durably stored elsewhere or storing thumbnails, transcoded media, or other easily reproducible data. It is designed to offer high availability and provides a durability level of 99.99%, which is still high but less than Amazon S3's standard storage. RRS is not recommended for critical data or data that cannot be easily reproduced, making it a suitable option for user-generated content, backup data, and logs.

10.5 CLOUD CONTENT STORAGE AND DISTRIBUTION

Amazon S3 is a highly durable and available storage solution for various types of content, such as web applications and media files. By offloading your entire storage infrastructure onto the cloud, you can leverage Amazon S3's scalability and pay-as-you-go pricing to meet your growing storage needs. Amazon S3 also provides a distribution platform, allowing you to share your content directly from Amazon S3 or use it as an origin store for pushing content to Amazon Cloud Front edge locations. Furthermore, Amazon S3's Reduced Redundancy Storage (RRS) feature provides a cost-effective solution for sharing easily reproducible content or storing original copies elsewhere.

10.6 STORAGE FOR DATA ANALYSIS

Amazon S3 provides an ideal location to store various types of original content, including pharmaceutical data for analysis, financial data for computation and pricing, or photo images for resizing. By using Amazon S3, you can transfer this content to Amazon EC2 for processing, resizing, or other large-scale analytics without incurring data transfer fees for transferring the data between the two services. Once you have generated reproducible content, you can use Amazon S3's Reduced Redundancy Storage feature to store it cost-effectively. This feature proves especially beneficial when sharing content that can be easily replicated or when you have a primary copy of the content stored elsewhere.

10.7 BACKUP, ARCHIVING, AND DISASTER RECOVERY

Amazon S3 presents a secure, scalable, and incredibly resilient option for archiving and backing up crucial data. Amazon S3's Versioning capability can offer further protection for your stored data. For data sets of considerable size, AWS Import/Export can be used to move large amounts of data into and out of AWS using physical storage devices. This method

is particularly useful for periodic backups or quick data retrieval in disaster recovery scenarios.

11. CONCLUSION

This survey paper has provided a comprehensive overview of cloud computing and cloud storage, focusing on the importance of cloud storage in today's data-driven world. We examined the architecture of cloud storage, its benefits, and its role in providing cost-effective solutions for businesses of all sizes. Specifically, we explored the Amazon S3 solution, which is a highly durable, scalable, and secure storage infrastructure. We discussed the basic concepts and architecture of S3,

highlighting its ability to store and distribute a wide range of content, including web applications and media files. We also explored why Amazon S3 is a preferred option for many businesses, including its pay-as-you-go pricing, ease of use, and ability to handle large-scale analytics. Additionally, we delved into some of the functionalities of Amazon S3, such as the Reduced Redundancy Storage (RRS) and Versioning capabilities, which provide cost-effective solutions for sharing reproducible data and ensuring data protection. With these features, it is clear that Amazon S3 is a reliable and efficient option for businesses looking to store, back up, and archive their critical data in the cloud. In conclusion, cloud storage and Amazon S3 offer a powerful combination of flexibility, scalability, and cost-effectiveness that can help businesses stay competitive in today's digital landscape.

12. REFERENCES

[1.] M. Lakshmi Neelima1, M. Padma2, "A STUDY ON CLOUD STORAGE" IJCSMC, Vol. 3, Issue. 5, May 2014

[2.] Amira Elzeiny1, Ahmed Abo Elfetouh2, and Alaa Riad "Cloud Storage: A Survey" IJETTCS Volume 2, Issue 4, July – August 201

[3.] Yan Han University Libraries, The University of Arizona, Tucson, Arizona, USA "Cloud storage for digital preservation: optimal uses of Amazon S3 and Glacier"

[4.] Ameya Kanitkar1, Shreyas Joshi2, Sumit Chitnis3, Mayur Gite IJRET "ONLINE BACKUP AND RECOVERY USING AMAZON S3" Volume: 03 Issue: 04 | Apr-2014

[5.] Allan Liu, Ting Yu "Overview of Cloud Storage and Architecture"

[6.] Gurudatt Kulkarni1, Ramesh Sutar2, Jayant Gambhir (IJERA) Vol. 2, Issue 1, Jan-Feb 2012, "Cloud Computing-Storage as Service"

[7.] Mayur Palankar, Matei Ripeanu, Simson Garfinkel "Amazon S3 for Science Grids: a Viable Solution?"