IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12660

# A Review paper on Cloud Storage

**Aditya Pardhikar[1], Dr. Mrs. Pratibha Adkar[2]**

Student, MCA Department, PES Modern College of Engineering, Pune, India[1]

Professor, MCA Department, PES Modern College of Engineering, Pune, India[2]

**Abstract**: Cloud storage technology has become a critical component in modern-day data storage and management, providing a scalable and reliable solution for organizations of all sizes. This paper provides an extensive review of cloud storage technology, covering its history, types, models, core technologies, architecture, and applications.The paper traces the evolution of cloud storage and explains the different types and models of cloud storage solutions, including public, private, and hybrid models. It then provides an in-depth analysis of the core technologies of cloud storage, focusing on the Google File System (GFS) architecture and the Hadoop Distributed File System (HDFS). The paper also examines the key features and benefits of cloud storage solutions, along with an assessment of their advantages and disadvantages. Overall, this paper serves as a valuable resource for students, researchers, and practitioners seeking to understand the current state of cloud storage technology and its potential impact on the future of data storage and management.

**Keywords:** Cloud Storage Models, GFS, HDFS, Amazon S3, Amazon EC2.

## I. INTRODUCTION

[2]As we all known disk storage is one of the largest expenditure in IT projects. ComputerWorld estimates that 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 meet this need, cloud storage has started to become popular in recent years.

Cloud storage is a new concept come into being simultaneously with cloud computing, which generally contains two meanings: It's the storage part of the cloud computing, virtualized and high scalable storage resource pool. Cloud users access to cloud computing services based on the cloud storage resources pool, but not all storage part can be separated in cloud computing. Cloud storage means that storage can be provided as a service over the network to the user. User can use storage pass through a number of ways, and pay by the use of time, space or a combination of both. Obviously, such statement is not tightly defined this new concept of cloud storage. In addition, the relationship between the concepts of Cloud Storage, Storage Cloud, Storage as a Service, Cloud-Based Storage should be cleared.

Cloud storage is divided into public cloud storage, private cloud storage and hybrid cloud storage. Public cloud storage is designed specifically for large-scale, multi-user cloud storage. All components are built on a shared infrastructure, and public storage devices were logical partitioned through virtualization technology, data access, data management technology, according users need. Also known as internal cloud storage, private cloud storage is designed for a specific user. Unlike the public cloud storage, private cloud storage running on a dedicated storage devices in the data center so as to meet safety and performance requirements. However, it's obvious disadvantage is the relatively poor scalability. The hybrid cloud storage is the cloud storage to integrate public cloud storage and private cloud storage. Generally, hybrid cloud storage was case-based in private cloud storage, supplemented by public cloud storage.

[1]In the present digital transformation age, moving towards cloud computing has made a massive progress over the past few years, which transforms the traditional network communication architectures into a new one model, aiming to improve agility and reduce operation costs as cloud computing depends on sharing of resources which allow business to introduce new products and services in many sectors. Cloud computing can be described as an IT paradigm that enables ubiquitous access to shared pools of configurable system resources and higher-level services that can be rapidly provisioned with minimal management effort, often over the Internet. Due to that, there is a huge amount of data from storage, transactions and connecting devices that become new sources for competitive advantage to support telecom to sustain in highly competitive environment. This means telecom needs to be solution and service oriented and only connectivity provider.

For example, an MGA-MENA Company, a world-renowned telecommunications group, currently operating in 16 countries across the Middle East, Asia, and Africa, has deployed its cloud computing models offering significant improvements in its services and it aims to be the smart company making use of artificial intelligence in running its operation and managing customer experience.

  

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12660

## II. TYPES OF CLOUD STORAGE

These days, cloud offerings are segmented based on use requirements—personal, public, and private.

- **Personal cloud**

[8]With a personal cloud, all that is truly required is a willingness to house your documents online as opposed to putting away everything on your PC's hard drive. Contemporary personal cloud options let you do huge numbers of varying digital activities, pulling from the idea of conventional cloud storage, including utilizing programming to synchronize your gadgets, sharing records, and easily accessing content from your smartphone.

In the event that you'd rather not put resources into buying costly equipment, but still need your own personal cloud, you can partner with a cloud vendor. To make this work, you'll have to look for a solution that supports web server hosting. With a personal cloud, the greatest advantage is speed, provided that you are utilizing a virtual host. Syncing your devices and streaming move much faster with a personal cloud.

Fig.1 Personal Space

Another benefit of a personal cloud is better protection. Sites such as Dropbox and Google Drive take every precaution to ensure their customer's data is protected against potential cyber threats. Whether the idea of threat actors stealing your data keeps you sleeping with one eye open or you simply need quicker access to your records, setting up a personal cloud may be the right decision.

Cloud drives can be accessed via different models and implementation techniques, including the most prevalent Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). The basic foundational engineering can take on different structures and features, including virtualized or a software-defined model.

Cloud drives utilize the cloud computing service model, where IT administration is conveyed over the web. The administration might be free, freemium, or a monthly membership-based offering depending on the size of storage needed. Some of the functionalities may vary based on provider and include email, applications, and storage to manage advanced software development and testing.

- **Public Cloud**

[8]The cloud vendor is liable for creating, overseeing, and keeping up with the pool of digital assets shared between numerous subscribers over a network. A public cloud configuration incorporates high flexibility and versatility for digitally-based organizations with little-to-no effort required on the backend.

[8]As the most well-known model of cloud computing, the public cloud offers varied options for computing resources and storage to address the developing needs of organizations. An example of public cloud usage would be an individual utilizing a PC connected to a cloud drive to transfer vacation photographs or to open a work document remotely. In an example like this, the end user may use a public cloud app such as Dropbox, iCloud, or Google Drive.

As a result, public clouds are often called customer clouds. For business use, cloud drives offer administration, storage, and applications to the end client through the web as opposed to an application/programming interface downloaded by

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12660

means of a server. Numerous organizations appreciate utilizing their "cut" of an open cloud interface since it's quick, efficient, and they can scale here and there depending on evolving needs.

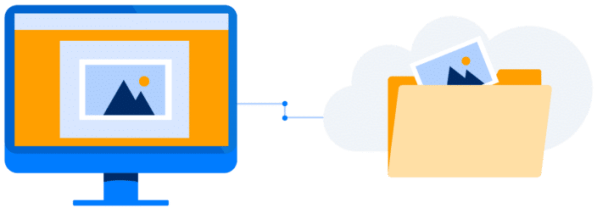

Fig.2 Public Cloud

Many companies also feel they're getting a good deal since they're only paying for the space they're using. Further, the transition of storage from physical to virtual has implied an adjustment in environments and design where applications, programming, and information are concerned. As a public cloud client, you aren't responsible for dealing with cloud management—that's left to your cloud vendor.

Thus, you don't have to suffer through a long procurement, installation, or deployment process. The public cloud also facilitates more agility, with easy connectivity significantly improving both productivity and efficiency. Plus, all updates are automated through your solution vendor.

- **Private cloud**

[8]As your organization chooses which type of cloud format to adopt, there are myriad questions to ask. What is a private cloud? Could a cloud truly be private? Furthermore, if it's private, can it truly be a cloud? While each of these inquiries may spur further discussion, "private clouds" are characterized by Gartner as "types of cloud computing that are utilized by just a single association, or that guarantee that an association is totally segregated from others."

It is by and large offered as a service with month-to-month rent. Since it is only available for a single institution, a private cloud can be designed to meet that association's particular needs. These attributes are fundamental for companies who manage sensitive consumer information and whose handling might be subject to government regulations or other mandates. This incorporates financial institutions, for example, banks or charge card businesses, and medical organizations.

In fact, most servers are underutilized, and virtualization gives private cloud clients improved asset use, as overwhelming workloads can be conveyed to an alternate physical server when digital asset storage needs change.

A key facet of private cloud computing is that the intended organization doesn't need to buy and keep up the framework, but instead use it, and pay for it as required, with a pay-more option only as costs rise. Obviously, for an in-house private cloud, the organization will need to pay for all the framework essential for the cloud, regardless of whether it is being utilized, or not. While any cloud offering requires substantial cybersecurity firewall protocols, a private cloud runs on certain physical machines, which makes its physical security easier to guarantee.

Access is also more secure since it is linked to and accessed through secure and private networks as opposed to the public web. Having a remotely-facilitated private cloud spares organizations from ownership, set up and overall maintenance of the framework fundamental for their cloud infrastructure. This can provide positive effects since private cloud vendors already have the server framework and are better situated to maintain awareness around advancing cloud technologies and any required maintenance.

A cloud supplier may also offer enhanced benefits around security, since they have the resources and expert security staff who are highly trained in quick and critical response to cyber threats.

[8]If you want to combine the varying cloud types, a hybrid cloud is a concoction of on-premises and off-premises IT assets. Hybrid clouds are made up of a blend of on-premises private cloud assets and cloud drive assets. The upside of the hybrid cloud model is that it facilitates remaining workloads and information to move among private and cloud drives in a scalable manner as requests, needs, and costs change, giving organizations more flexibility and more alternatives for information networking. The essential advantage of a hybrid is agility. The need to adjust and alter course rapidly is a now-common standard of the digital era.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12660

## III. GOOGLE FILE SYSTEM ARCHITECTURE

Storage technology has gone through the development course from the tape, disk, RAID to storage networking system. n recent years, the application demand for massive data storage is increasing, which directly contributed to the emergence and development of highperformance storage technology, and have produced some typical storage technology being fully applied in Cloud computing, such as the Google File System (GFS), Hadoop Distributed File System (HDFS), S3, SAN.

Traditional storage means a specific storage device, or the assembly constituted by the large number of the same storage device. In using traditional storage, you need a very clear understanding of some of the basic information of the storage device, such as device type, capacity, supported protocols, transmission speed. In addition, regular maintenance of the equipment, hardware and software updates and upgrades need to be considered separately. Although cloud storage is composed by a large number of storage devices, storage devices can be heterogeneous and cloud storage users do not care about the basic information of the storage devices and its location. In cloud storage, issues such as equipment failures, equipment updates and upgrades were also be full considered, and can provide more reliable service. The core of cloud storage is the combining of application software with storage device, to achieve the changes from storage device to storage service by application software. It's not directly to use storage devices but the Data Access Service provided by the entire cloud storage system for users. So in the strict sense, cloud storage is not storage but a service. Cloud storage provides directly data storage services for end users and indirect data access in application system, and other forms of service, with many service forms of network hard drive, online storage, online backup and online archive storage service.

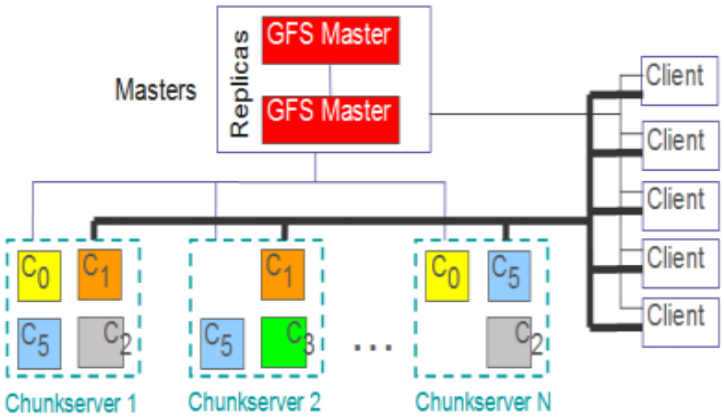


Fig.3 Google File System Architecture

Operating system, service procedures, user application or the vast amounts of data are stored in storage systems, the basis role and status of storage is widely recognized by the industry in supporting the popular cloud computing. To achieve massive data storage, cloud computing service providers must build a huge database of globalization and storage center. For example, Google's great success in the field of cloud computing considerable extent thanks to its advanced cloud storage platform based on GFS. GFS is a distributed file system to handle large-scale distributed data, Fig3 . Show its architecture . A GFS cluster consists of a Master server and multiple block server (Chunkserver), and access by multiple clients. The Master server is responsible for the management of metadata, the name space of the stored files and blocks, the mapping relationship between the file to blocks and the storage location of each block copy. The file is split into blocks of fixed size (64M) and be storage in the Chunkserver. Blocks were processed as Linux files and stored on the local hard disk. To ensure reliability, each block was saved with 3 backups in default. The Chunkserver obtain the data directly back to the client after the client transmitting a data request.

## IV. HADOOP DISTRIBUTION FILE SYSTEM

After the analysis of Google's GFS massive data storage system, we will discuss the popular open source cloud storage system Hadoop's HDFS. Hadoop is an Apache open source organizational design of a distributed computing framework, its core technology HDFS, MapReduce, HBase were the open source implementation of GFS, MapReduce, Bigtable in Google cloud platforms. It is worth mentioning that Hadoop can run on a large number of cheap machinery and equipment. Having many similarities with other distributed file system, the features of HDFS are also very obvious

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12660

because of its targets and assumptions of Hardware Failure based design, Streaming Data Access, Large Data Sets, Simple Coherency Model, Moving Computation is Cheaper than Moving Data, Portability Across Heterogeneous Hardware and Software Platforms. Although HDFS running on cheap commodity hardware, it can meet the data access requirements of high reliability, high throughput, large data sets.

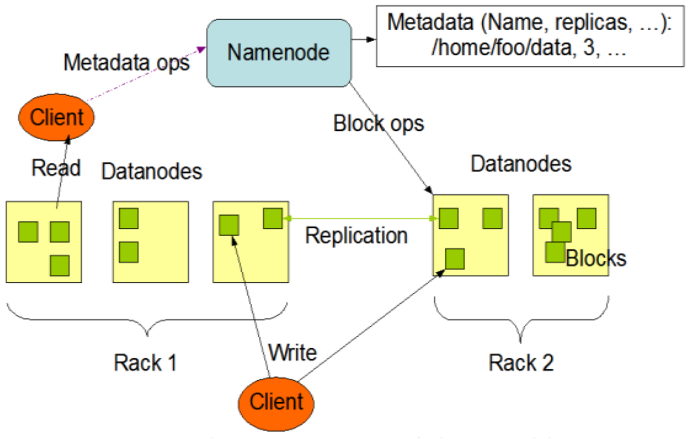


Fig.4 The HDFS Master/Slave architecture

As shown in Fig.4, in HDFS Master / Slave architecture, every Cluster is composed of a NameNode and multiple DataNode and multiple Clients. The NameNode is mainly responsible for managing the file system namespace, cluster configuration information and stored block copying, storing Metadata of the file system to memory, and this information includes the file information, the file block information for each file, DataNode information of each file Blocks. The NameNode execute file operations, including open, close, rename, catalog maintenance, it also determines the mapping between the Block and DataNode. Internally, a file is divided into one or more Blocks. DataNode is responsible for the read and write requests from Clients, and executing instructions of Block establish, delete, copy and other, issued by the NameNode. DataNode stored Blocks and their Metadata in the local file system, and periodically sending information of existing Blocks to the NameNode at the same time. Clients are applications to get a file from the file system.

## V. CLOUD STORAGE ARCHITECTURE

[9]Cloud storage is based on virtualized infrastructure and is like cloud computing in terms of accessible interfaces, scalability and metered resources. A cloud-storage service is utilized from an off-premises service (Amazon S3). It refers to a hosted object storage service, but the term has broadened to include other types of data storage that are now available as a service, like block storage. Object storage services like Amazon S3, Oracle Cloud-Storage and Microsoft Azure Storage, Object Storage Software like Open Stack Swift are all examples of storage that can be hosted and deployed with cloud-storage characteristics.

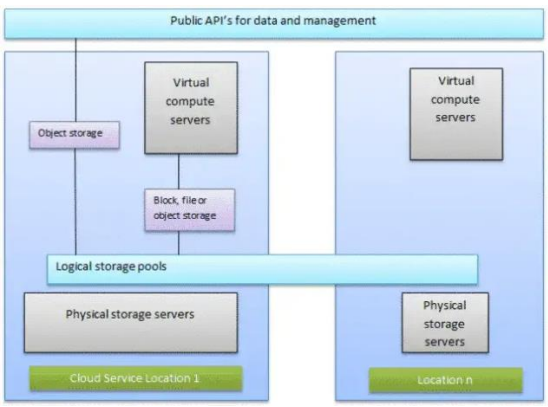


Fig.5 High Level Cloud Storage Architecture

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12660

[10]Cloud storage architecture refers to the design and structure of the underlying infrastructure that supports cloud storage services. In general, cloud storage is a type of online storage where data is stored remotely and can be accessed over the internet.

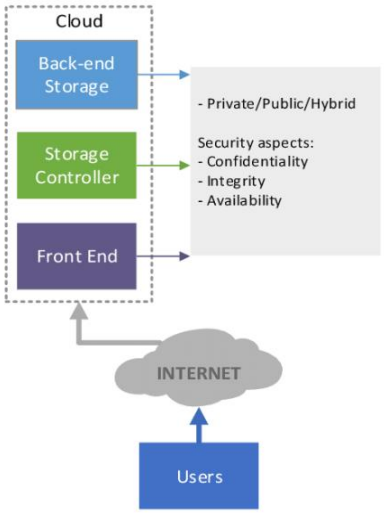


Fig.6 Generic Cloud Storage Architecture

Here's a brief description of the architecture and application of cloud storage:

- **Data Center:**

The data center is the physical location where the cloud storage service stores and manages data. It is typically a large facility that houses servers, storage devices, and networking equipment. These data centers are designed to be highly secure, reliable, and scalable to accommodate the storage needs of many users.

- **Virtualization:**

Virtualization allows multiple users to access the same storage resources while maintaining the security and privacy of each user's data. This is achieved by creating virtual machines or containers that separate each user's data from others. Virtualization also allows for greater flexibility and scalability in the allocation of storage resources.

- **Storage Management:**

Storage management is responsible for managing the allocation of storage resources, data backup and recovery, and ensuring the data is available to users at all times. This includes monitoring storage capacity, performance, and availability, as well as implementing data backup and recovery strategies to prevent data loss.

- **Cloud Storage Services:**

Cloud storage services provide users with access to cloud storage resources and typically include features like file sharing, collaboration, and synchronization across multiple devices. These services are typically offered by cloud storage providers and can be accessed through a web browser or mobile application

## CONCLUSION

Cloud storage is a rapidly growing technology that is revolutionizing the way individuals and businesses store and access their data. It offers many benefits over traditional storage solutions, including scalability, flexibility, accessibility, and reliability. Cloud computing and cloud storage are two critical technologies that work together to enable the delivery of computing services over the internet, including storage, processing, and applications.

Several key technologies have made it possible to store and manage massive amounts of data in the cloud. Google File System (GFS) is a distributed file system that stores and manages large datasets across multiple servers. Hadoop HDFS is an open-source distributed file system used in big data analytics and processing. Future research in cloud storage will focus on performance analysis and optimization, storage virtualization and management, data compression, backup and recovery, deduplication, encryption, and platform architecture.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023
DOI: 10.17148/IJARCCE.2023.12660

Although there are concerns about the mainstream adoption of cloud computing, new open source systems are emerging that can be installed and run on local clusters. This provides users with greater control over their data and reduces potential risks associated with cloud-based storage. Overall, cloud storage is a critical technology area that will continue to drive innovation and shape the future of how we store, access, and manage our data.

## REFERENCES

[1]. 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.
[2]. Brandon Rich, Douglas Thain. DataLab: Transactional data-parallel computing on an active storage cloud. In: Proc. of the 17th International Symposium on High Performance Distributed Computing,2008, 233-234.
[3]. Amazon, http://aws.amazon.com/s3/ ,2011.
[4]. Amazon, http://aws.amazon.com/ec2/ ,2011.
[5]. "A History of Cloud Computing". ComputerWeekly.
[6]. Louden, Bill (September 1983). "Increase Your 100's Storage with 128K from Compuserve". Portable 100. New England Publications Inc. 1 (1): 22. ISSN 0738-7016.
[7]. *Daniela Hernandez (May 23, 2014).* "Tech Time Warp of the Week". *Wired.*
[8]. https://capacity.com/cloud-storage/history-of-cloud-storage/
[9]. https://electricalfundablog.com/cloud-storagearchitecture-types/
[10]. https://google.com
[11]. OpenStack Swift. https://wiki.openstack.org/wiki/Swift (2015)
[12]. https://hcis-journal.springeropen.com/articles/10.1186/s13673-019-0173-x#Sec3