e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:07/July-2022 Impact Factor- 6.752 www.irjmets.com

A CASE FOR CLOUD STORAGE AND DIVERSITY

Deepali Singh[*1], Menka Kumari[*2], Yashwant Soni[*3]

[*1,2,3]Student, School Of Computing Science And Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India.

[*4]Associate Professor, School Of Computing Science And Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India.

ABSTRACT

As an alternative to the on-site datacenter, cloud computing is available. We have to manage everything with an on-premises datacenter, including ordering and installing hardware, implementing virtualization and installing the operating system. In this first, we will show all the types of and services of cloud computing. Computer scientist Dr. Joseph Carl Robnett Licklider is credited with creating cloud storage in the 1960s. About 20 years later, CompuServe started providing its users with limited amounts of disc space so they could keep some of their files. Digital images and music files are only two examples of the types of data that a cloud storage system can store. It can also offer general storage for all kinds of data, including spreadsheets, presentations, text documents, and photos. More and more data are being transferred to cloud storage systems as a result of the cloud infrastructure's rising use and popularity. This makes cloud storage service availability crucial, especially in because outages of cloud storage systems have occurred occasionally. Customers, however, run a serious risk if cloud storage providers succeed since switching. storage providers cost a lot of money. Because of the weakened service availability, simply relying on one cloud storage provider for storage services runs the risk of breaking the service-level agreement (SLA). In this paper we will introduce the concept of "Cloud-of-Clouds", Cloud Computing, Cloud Storage as well as the architecture of Cloud Storage.

Keywords: Cloud Computing Types And Their Services, Cloud Storage, Cloud Of Clouds, Cost Efficiency, Data Redundancy.

I. INTRODUCTION

The idea of cloud computing has gained popularity over the past few years. Distributed processing, parallel processing, and grid computing all contributed to the development of cloud computing as a new business model. Currently, Google Amazon, IBM, and Microsoft are all seeking to develop cloud computing technologies and products. Many businesses and organizations have migrated data out of their Own data centers and into the cloud as a result of cloud storage's rising popularity and low cost. Examples of typical applications include online digital media and backup data storage.

Cloud system:

Cloud computing is the supply of on-demand IT resources over the internet, to put it simply.[10] These businesses provide cloud computing services and are known as cloud service providers (CSPs).Cloud computing is of four types- Public cloud, Private Cloud, and Hybrid Cloud . Service level agreement- An SLA typically describes the type of services that will be offered, the objectives of both parties (the supplier and the client company), any necessary preconditions, and the points of contact. Additionally, it describes in detail what will happen if SLA objectives are not realized.[2].

II. RELATED WORK

More and more businesses and individual users will shift their data to the cloud as it becomes more popular and affordable. The idea of a "Cloud-of-Clouds" is an efficient strategy for resolving the availability issue brought on by single-cloud storage providers' service interruptions.

Several systems proposed for cloud of clouds- RACS Redundant array of cloud storage, Erasure coding is used by RACS to reduce the issue of vendor lock-in that arises when a user switches cloud vendor. Disks and file systems use RAID-like strategies. HAIL guarantees the availability and integrity of stored data. In order to increase the availability of cloud storage services, NCCloud successfully fixes a persistent single-cloud provider failure at a reasonable price. Cloud computing have four types- Public, Private, Hybrid, and Community Cloud.

It consists of three services- IaaS, PaaS, SaaS.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:07/July-2022 Impact Factor- 6.752 www.irjmets.com

III. CLOUD COMPUTING AND THEIR SERVICES

Different services are delivered via the Internet through cloud computing. These tools a, programs comprise software, servers, databases, networking, and data storage, among other things. Cloud computing consists of the following types- Public cloud, Private cloud, and Hybrid cloud. Users can acquire services from the cloud - based on their specific needs and utilize the cloud's tremendous processing and computing capabilities.[9]

Fig. 1. Cloud computing types[1]

Public Cloud- The term "public cloud" refers to computer services that are made available to users or purchasers via the open internet by third-party vendors. Customers can pay for the cycles, storage, or bandwidth they use on a per-user basis for free or on-demand. Examples- Sun Cloud, AWS, and Microsoft Azure.[4]

Private Cloud- A private cloud is a cloud computing system where IT services are provided over private IT facilities for the specific use of one enterprise. Only one company manages the cloud infrastructure. Examples- AWS and VMware.[7]

Hybrid Cloud- To exchange data and apps, a hybrid cloud is a computing environment that combines public andprivate clouds. Examples- Federal agencies choose personal clouds when dealing with sensitive data.[15]

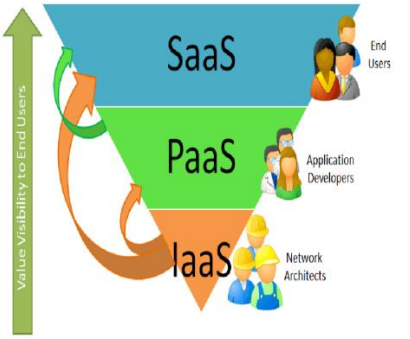

Fig.2. Cloud computing services[13]

Software as a Service (SaaS)- SaaS is defined as the software distribution model deployed on the internet in which a cloud service provider provides applications. It is also known as "on-demand software" or "pay-as-you-go application". Example- Amazon web services, salesforce, etc.

Platform as a Service (PaaS)- PaaS is a platform for programming developers and brings benefits. It is the computer platform that provides the facility to use web applications quickly. Examples- Microsoft azure, SAP cloud, etc.[5]

Infrastructure as a Service (IaaS)- The term "IaaS" refers to a method of providing cloud computing infrastructures, such as virtual machines, storage devices, servers, operating systems, and networks. Examples- Microsoft Azure and Digital Ocean.[14]

1. Cloud storage:

Data is stored in logical pools using several servers as part of the cloud storage technology. To store user, organization, or application data, the organizations purchase storage capacity from the suppliers. Due mostly to the advantages it offers, cloud storage has gained popularity over the past few years and is now directly

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:07/July-2022 Impact Factor- 6.752 www.irjmets.com

competing with local storage. Security: Because the backups are spread across several servers, they are more guarded against data loss and hackers.

Accessibility: No matter where you are, you may access the stored data online.

1.1. History of cloud storage:

In order to connect people and data from everywhere at any time, Joseph Carl Robnett Licklider is thought to have invented cloud computing in the 1960s while working on the ARPANET.

In 1983, CompuServe provided a meagre amount of disc space to its customer users, which could be utilized to store any files. Aplatform for online business and personal communication called Personal Link Services was introduced by AT&T in 1994. One of the first to be entirely web-based, the storage launched its cloud storage service AWS S3 in 2006. Box unveiled apersonal cloud information management and online file sharing solution for companies in 2005.

1.2. Architecture:

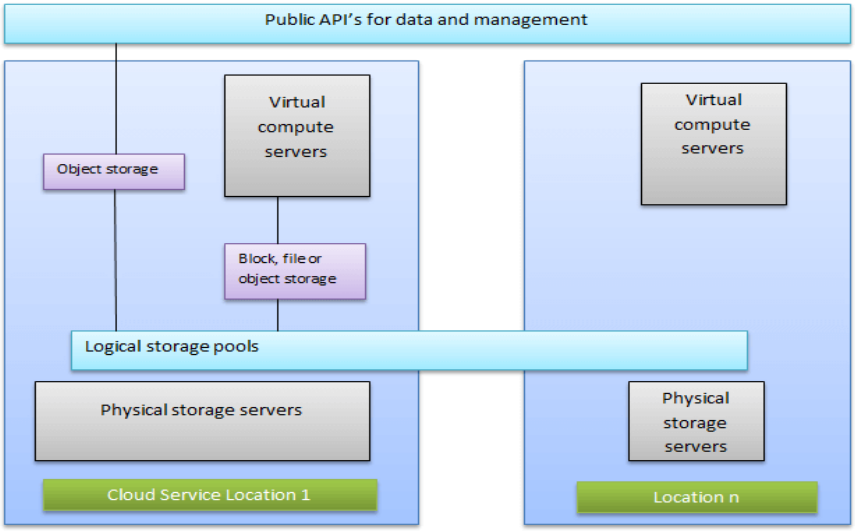

Fig.3. Cloud architecture [6]

The main goal of cloud storage designs is to supply storage on demand in a highly scalable, multi-tenant manner. An API is typically exported by the front end of cloud storage designs so that users can access the storage. This API represents the SCSI standard in conventional storage systems, but in the cloud, these protocols are changing. Storage logic layer of middleware is located behind the front end. Replication and data minimization are just a couple of the features that this layer puts into practice. Data storage is implemented physically on the back end. This could be a traditional back end to the physical discs or an internal interface that implements particular functionality.

1.3. Advantages of Cloud Storage:

File Accessibility - As long as you have Internet connectivity, you can access the files from anywhere at any time.

Offsite Backup - Cloud Storage gives businesses access to offsite (remote) data backups, which lowers costs.

Effective Bandwidth Use - Cloud storage effectively utilizes bandwidth by allowing receivers to receive web links via email in place of files.

Data security - As data is secured and requires proper authentication to access, it aids in safeguarding data from malware or ransomware.

2. Background and Motivation:

In this part, we first discuss some key findings gleaned from earlier reports on single Cloud service disruptions, cloud storage diversity, and workload characteristics. Then we give the characteristics of the replication- and erasure-code-based redundant data distribution in Cloud-of-Clouds.

2.1. The Outages of Single Cloud Storage Services:

The term "cloud outage" simply describes the time that the cloud infrastructure service cannot be used. You give up control when you store your data in the cloud. Additionally, there are serious security concerns. The

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:07/July-2022 Impact Factor- 6.752 www.irjmets.com

reason is that even for major providers like Amazon S3, Microsoft Azure, Google, Rackspace, and other IaaS vendors, cloud storage failures do occur occasionally, from a five-minute failure that costs $500,000 to a week-long disruption that causes incalculable brand damage till 2022, Drop Box has experienced many times of service outages. The availability of cloud storage services is directly impacted by outages of a particular cloud storage provider. There is a chance that data will be lost or not available due to the cloud storage outage. Even though the cloud provider and the user have rigorous Service Level Agreements (SLAs), service faults and outages do happen and are essentially inevitable. Users are able to preserve their mobility while being protected against disruptions caused by a certain cloud provider as a result.

2.2. Workload Characteristics:

Understanding the workload characteristics is essential for avoiding storage system design inefficiencies. findings from earlier research on the workload characteristics have been made. First, fewer than 20% of the total storage space is used by files larger than 4KB. Second, the most common types of access are too tiny files and metadata. Even though they make up a very small portion (10% to 20%) of the total number of files, huge files occupy a very large portion (80%) of storage space. While making the cloud of clouds it is necessary to understand two workload characteristics.

2.3. Growing Diversity in Cloud storage:

Growing Diversity of Cloud Storage Solutions by Arthur Cole April 26, 2017, Storage remains the primary application for the enterprise cloud, but its days as simply a low-cost warehouse for bulk data are coming to an end. According to 451 Research, overall pricing for cloud storage will continue to fall as providers shift the competitive playing field from virtual machines to object storage. Over the past year, object pricing has dropped by14 percent compared to a 5percent cut in VM costs. The organization can now more easily set up integrated storage solutions across various third-party clouds thanks to recent improvements. A blockchain solution from Sia was recently integrated into the open-source storage management platform NextCloud, enabling businesses to create encrypted, distributed storage pools with practically any provider.

Issues- Downtime history can be used to gauge a company's previous reliability and, by extension, predict its potential future reliability. Data accessibility, comprising the amount of bandwidth available within, between, and to the Internet from each data centers their costing schemes. Neglecting to determine how quickly you can scale your requirements up and down is a common issue with cloud storage.

2.4. Data distribution:

Cloud-of- Compared to their individual cloud storage provider, clouds' storage system is more dependable. The distribution of data is a challenging problem in cloud computing because all data is accessible to anyone with cloud access. A helpful concept called cloud computing provides servers, networks, and data storage as a service.

Fig. 4. Future of cloud storage [11]

This image shows how the data is distributed in stock.

2.5. Replication vs Erasure codes:

Replication-based and erasure-code-based schemes are two typical redundant data distribution techniques used in Cloud-of-Clouds to achieve high availability of data. To achieve high availability for large systems, system architects must first increase the number of copies. In comparison to tight replication, an erasure code offers redundancy with much less storage overhead. To allow for the recovery of the original pieces from a

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:07/July-2022 Impact Factor- 6.752 www.irjmets.com

subset of the n fragments, erasure codes split an item into m fragments and recode them into larger n fragments.

The only redundant data in erasure coding is the parity code, but in replication all the data is duplicated, requiring more storage space. Erasure coding, which is frequently used in RAID parity stripes, involves computing parity and striping the generated data across many storage subsystems to achieve high availability (or disks). Simply said, replication, sometimes referred to as mirroring, involves writing multiple copies of the data to several storage subsystems or discs.

IV. FUTURE WORK

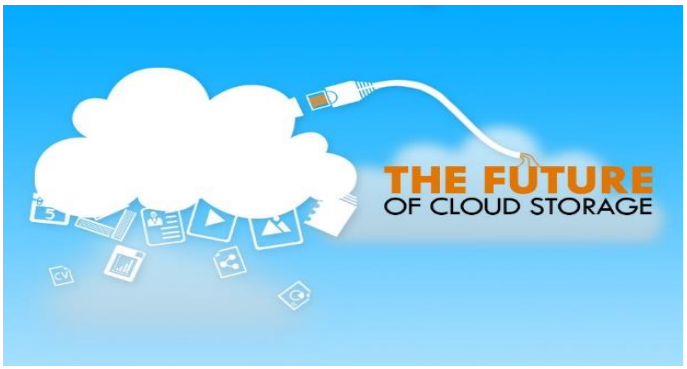

Fig. 5. Future of cloud storage [11]

We have considered cloud storage as a stand-alone product in this paper, but the cloud is an ecosystem. Cloud storage is the use of the internet to outsource tasks that might be performed on a computer. It is mainly used for Storing data online.

In the coming 2 years, 83 percent of the business enterprise will use the public cloud, 20 percent will use the private cloud and 22 percent will use the hybrid cloud. This is the future of cloud storage.

However, cloud computing will continue to be crucial in the case of large data transfers even if data analysis and storage are relocated to the edge. In order to improve data management and analysis, the best coherent practices that combine the finest features of both technologies will be developed in the future.

V. CONCLUSION

When determining whether or not to move their data to the cloud, users must thoroughly evaluate a number of important considerations, one of which is the accessibility of cloud storage services. The concept of a cloud of clouds, cloud storage, and cloud computing have all been thoroughly explored in this study. To increase storage availability in a cloud-of-clouds, we have taken use of workload characteristics and the variety of cloud storage providers. Both the benefits of erasure coding and replication are utilized by taking advantage of workload characteristics and cloud provider heterogeneity, while their drawbacks are reduced.

VI. REFERENCES

[1] RACS: A Case for Cloud Storage Diversity Hussam Abu-Libdeh Cornell University Ithaca, NY 14853 hussam@cs.cornell.edu Lonnie Princehouse Cornell University Ithaca, NY 14853 lonnie@cs.cornell.edu Hakim Weatherspoon Cornell University Ithaca, NY 1485

[2] Exploiting Workload Characteristics and Service Diversity to Improve the Availability of Cloud Storage Systems Bo Mao, Member, IEEE, Suzhen Wu, Member, IEEE and Hong Jiang Fellow, IEEE.

[3] Yunhong Gu, Robert L. 2009. Grossman. Sector: A high performance wide area community data storage and sharing system. Future Generation Computer Systems, 20 May 2009.

[4] James Broberg, Rajkumar Buyya, Zahir Tari. 2009. MetaCDN: Harnessing 'Storage Clouds' for high performance content delivery. Journal of Network and Computer Applications 32 (2009), 1012--1022.

[5] Takahiro Hirofuchi, Hidemoto Nakada, Hirotaka Ogawa, Satoshi Itoh, Satoshi Sekiguchi. 2009. A live storage migration mechanism over wan and its performance evaluation. Proceedings of the 3rd international workshop on Virtualization technologies indistributed computing, Barcelona, Spain, 2009, 67--74.

[6] Wenying Zeng, Yuelong Zhao, Junwei Zeng. 2009. Cloud service and service selection algorithm research. GEC '09: Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation, Shanghai, China, June 2009, 1045--1048.

[7] Ying Zhan, Yong Sun. 2009. Cloud Storage Management Technology. Second International Conference on Information and Computing Science. Manchester, England, UK, May 21-May 22, 2009, icic, vol. 1, 309--311.

[8] Henry Newman. 2009. Why people don't like to use cloud st orage? http://www.cnw.com.cn/storage-Technology/htm2009/20091013_183980_2.shtml, 2009-10-13.

[9] FalconStor Software, Inc. 2009. Demystifying Data Reduplic ation: Choosing the Best Solution. http://www.ipexpo.co.uk/c ontent/download/20646/353747/file/DemystifyingDataDedu pe_WP.pdf, White Paper, 2009-10-14, 1--4.

[10] Mark W. Storer Kevin Greenan Darrell D. E. Long Ethan L. Miller. 2008. Secure Data Deduplication. StorageSS'08, October 31, 2008, Fairfax, Virginia, USA. 2008, 1.

[11] Albert Greenberg, James Hamilton, David A. Maltz, Parveen Patel. 2009. The Cost of a Cloud: Research Problems in Data Center Networks. ACM SIGCOMM Computer Communication Review, Volume 39, Number 1, January 2009:68--73.

[12] K. Rashmi, N. Shah, D. Gu, H. Kuang, D. Borthakur and K. Ramchandran, "A “Hitchhiker's” guide to fast and efficient data reconstruction in erasure-coded data centers", Proc. ACM SIGCOMM Conf. Appl. Technol. Archit. Protocols Comput. Commun., pp. 331-342, Aug. 2014.

[13] K. Rashmi, N. Shah, D. Gu, H. Kuang, D. Borthakur and K. Ramchandran, "A solution to the network challenges of data recovery in erasure-coded distributed storage systems: A study on the facebook warehouse cluster", Proc. 5th USENIX Workshop Hot Topics File Storage Technol., pp. 1-5, Jun. 2013.

[14] R. Villars, C. Olofson and M. Eastwood, "Big data: What it is and why you should care white paper IDC", 2011.

[15] N. Agrawal, W. J. Bolosky, J. R. Douceur and J. R. Lorch, "A five-year study of file-system metadata", Proc. 5th USENIX Conf. File Storage Technol., pp. 31-45, Feb. 2007.

[16] Traeger, E. Zadok, N. Joukov and C. Wright, "A nine year study of file system and storage benchmarking", ACM Trans. Storage, vol. 48, no. 2, pp. 1-56, 2008.