

A STUDY to CHALLENGES of CLOUD STORAGE ARCHITECTURE: PERFORMANCE and SCALABILITY

HATEM Taha*, ALI. Manea, EI KADIRI. Kamal Eddine

Abstract— Cloud storage is one of the cloud computing-based services, which provides remote storages and management operations for resources. The potential benefits of a cloud storage system depend on a number of factors, such as being able to store and manipulate data in the cloud with higher performance, more scalability, and cheaper storage. While there are many cloud storage systems, such as Amazon s3, cloud drive, DropBox, Microsoft Sky Drive, Google drive and SugarSync being developed, they are using different technology for storage data. In this paper, we will review the cloud storage history. In this paper, we will review the cloud storage history. We will discuss the progress made; the challenges that still need to be addressed to implement an efficient architecture of cloud storage. Also, we will present a methodology to improve the performance and the scalability to save data in the cloud using geographic cloud storage station, provide an overview of the trend they each.

Keywords —Architecture, Cloud Storage, Challenge, Performance, Scalability

1 INTRODUCTION

At the moment, Cloud storage is a model of networked online storage where data is stored on multiple virtual servers, rather than being hosted on dedicated server. Therefore, Cloud storage services have started to become popular and accessible via the Internet. The economic situation and the advent of new technologies have sparked strong interest in the cloud storage provider model. It offers the huge capacity to cost reductions, the elimination of labor required for data storage management, maintenance of resources, and immediate provisioning of capacity at a very low cost, that is in keeping with the fast world of dynamically changing in data storage service.

Cloud Storage is part of a wider definition called Cloud Computing which, according to the National Institute of Standards and Technology-NIST[1], is “a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction”.

Cloud Storage as part of a cloud computing which offers an opportunity to meet today's demands according to some fundamental models: Cloud Software as a Service (SaaS), Cloud Platform as a Service (PaaS) and Cloud Infrastructure as a Service (IaaS). The Three levels of cloud offering are shown in figure 1. Wherefore, it brings along all technologies (Web services, virtualization, Service-Oriented Architecture (SOA), grid computing, ...) with the service request, scalable and elastic[2].

In addition to some services, such as support asynchronous resource sharing among various platforms, online backup and archiving, allows the customer to store their data at remote storages and access them anytime from any place. These

services are offered by the infrastructure provider through tailored service contract: Service Level Agreements (SLA). It also uses billing mechanisms, allowing on-demand model: pay-per-use[3]. For example; Amazon S3 is widely used as a cloud storage service, other cloud service providers such as Google, IBM, and Eucalyptus. Today, all those offer cloud storage solutions, especially Amazon EC2, Microsoft Azure, Google Apps and IBM blue cloud.



Fig. 1 Cloud Computing

This paper provides a summary review of cloud storage; it also analyzes and gives opinions on the solution submitted by the recent researches for cloud storage system. It recalls the principal challenges, obstacles in our article[4], proposes a framework for the analysis of the challenges for this proposal, identifies enablers that tackle each of them and discusses future directions and trends in this area. The challenges of the architecture of Cloud storage service is a major topic in this paper that mainly concentrated on performance, cost issues and scalability options.

2 OVERVIEW CLOUD STORAGE

This section describes the cloud storage data methodologies that are converging to offer a cost-effective delivery model for cloud storage. Generally, public cloud storage is best used for unstructured data, whereas the private cloud storage is used by customers with more need for customization and more control over their data and also the hybrid cloud may be a solution for the customers those have a need for both types of cloud storage, it might be for one user, tiny company or massive company.

We see all users would need to know what the needs exactly in order to assess the best for cloud storage, for example, benefits and drawbacks of these services, application domains, and many services. The aim of development cloud storage system for exploit the broadband network, Web 2.0, virtualization, storage network, application storage for servers and storage devices, cluster technology, grid computing, distributed file system, content delivery network, data compression, data encryption[5].

According to Storage Networking Industry Association (SNIA), that is an association of producers and consumers of storage networking products, "When discussing cloud storage and standards, it is important to distinguish the various resources that are being offered as services".

2 NEED FOR CLOUD STORAGE

The objective is to store a large amount of long-term information. In fact, the future of cloud storage is closely associated with the internet, due to the importance of the Internet to store data in the cloud. In general, the digital environment experiencing nowadays of pose new requirements and challenges in the Cloud Storage, especially with respect to the explosion of personal and organizational digital data. Although there are other alternatives, such as traditional storage methods, Cloud Storage (Online, Offline, Nearline) and nanotechnologies suggest important perspectives on IT storage capacity in the coming years.

We can say that cloud storage at its basic level is simply backing up data enough times in order that the prospect of losing that data becomes nearly digressive by using multiple servers to store data. It's a key component in the cloud computing services which provide large amounts of data storage and processing, and availability of data, easy to share with others,

as well as reduced infrastructure costs at the simplest level. That these motivations or essential points that indeed allowed and pushed towards the establishment of the cloud storage term, there are many providers such as Amazon [6], Google [7] and Yahoo [8], are typical examples of cloud storage systems, hence providing means of data storage a user. This data stored in a cloud is safe; depend on various technologies to resolve the problem of data storage.

Cloud Storage-As-A-Service (STaaS): In the study and the methodology by Cisco [9], "STaaS is one of biggest expenses; in many enterprises, the storage is responsible for almost 30% of capital expenditures, as the average growth of data approaches close to 50% annually in most enterprise".

3 DATA GROWTH INTERPRETATION IN STAAS

Our first task in this study is to understand the need for cloud storage. What are its objectives and main services? Who are its customers and their need? In some cases, it is clear that technological advances and data growth which generate by the customers or the enterprise are rapidly outpacing the ability to process, analyze and store the data, also the enterprise needs for data storage are increasing by up to 40% per year[9]. That means, that the customers need reliable and affordable storage solutions their ever, to data storage and backup. For that, we believe that the presence a good architecture of cloud storage is a solution to this growth and this need.

4 CLOUD STORAGE INFRASTRUCTURE REQUIREMENTS

This section presents general challenges for the architecture of cloud storage including Performance, scalability, cost, reliability and efficient. Many businesses, governments, non-profit organizations and some consumers are facing growing challenges in storing, managing and protecting of data being generated in an increasingly digital world. The fastest-growing data is the archive data.

But not all cloud storage providers has all these characteristics, for example, one provider may focus primarily on cost or security, while another focuses on availability or performance. However, there are some common denominators that must be considered to make the architecture of cloud storage valuable.

We can see different cloud storage architectures, but these architectures are usually incompatible and complex. We focus on the challenge of performance and scalability in our architecture of cloud storage.

4.1 High Performance

Cloud storage performance is an important metric within this technology and in each layer of the cloud storage being viewed as the technology of these days and also the future. In the user application layer, apps hosting platform layer, storage management layer, and storage resource layer. Performance varies based on a number of factors related to the physical connection by service architecture cloud storage [10].

However, we see the good performance of the service is a shared responsibility between service providers and customers, which covers several aspects: the performance of the storage media as hardware (hard disk, optical disk, SSD, etc.), the organization of disks, the technology and software

solutions for accessing the data itself (data compression, deduplication, hierarchical, thin provisioning, etc.). Performances from the view of the end-user are the upload and download speeds for files; we must distinguish the performance problems to cloud provider's service and to personal devices of customers.

Cloud storage performance can be enhanced by using multiple technologies of storage; for example disk-to-disk, snapshot and data deduplication to increase performance in terms of storage capacity, reliability, scalability, and availability. There are many comparisons between different cloud storage services, but most of them related to function and price. Performance is an important feature for the storage system. But it is difficult to evaluate the performance because of the complexity of system and network conditions[9]. So the challenge for many cloud storage providers is how to tap into the cloud's scalability with low cost while maintaining a modicum of performance.

Other metrics can be estimated of the architecture of cloud storage. But our view the scalability, performance, and cost are becoming major issues in cloud storage and which determines cloud storage architecture services capacity when accessing, retrieving and saving data. There are other factors that must be considered to augment performance of cloud storage, for example in Morocco and Yemen, some factors that affect cloud storage performance and scalability include:

- **The Coverage:** The explosive growth of IT and Smartphones and App usage bring new challenges to mobile operators worldwide.
- **Network Bandwidth:** Network bandwidth efficiency is the most significant, also extremely challenging issue in cellular networks.
- **The distance between Cloud Providers and Cloud Customers:** Additionally, the performance of these approaches is affected by long WAN latency due to the long distance between cloud consumer and cloud storage providers service (CSPs).
- **Transfer Speeds:** The upload and download performance is limited by the Internet bandwidth. Even unlimited bandwidth, distance and coverage without solving the latency problem will not improve the performance of cloud storage services.

When we talk about the cloud storage architecture, we know that very little work has been done to create valid and standard cloud storage architecture, but this can only be done if we have solutions that address the various challenges in this area, according to International Data Corporation[11].

In this study, we see that the cloud storage is a hybrid approach combines new and traditional technology that used to reduce burdens for the enterprises, especially in the start-up phase. Also to protecting the customer's data so that it could be extensively used for storing and sharing data. For that, we

need to cloud storage accessible to all, safe and scalable, with offsite backup service for any data or any system anytime, anywhere, which means access to cheap store services for the customer. Furthermore, we see that these requirements may be realized in the technology Disk-to-Disk and snapshot which is gaining in popularity in both large enterprises and small businesses.

4.2 Scalability

First of all, we should study the relationship between performance and scalability of cloud storage; they are closely linked to the building an efficient cloud storage architecture. In addition, that they have some of the common characteristics of this technology where data is measured in terabytes and petabytes. These data have become multipoint as well as multi-directional. So, we see that building efficient and scalable storage architecture is a significant challenge.

According to[12], he said "We have to consider that today. The scalability is becoming a major issue in cloud storage. If we are looking at the dozens, hundreds or thousands of organizations, as well as the capacity of a single organization that needs to scale its storage into the petabytes of cloud storage".

Consequently, one of the major challenges of cloud-based storage is the scalability and elasticity. We can measure this scalability in various dimensions, in the administration, functionality, and location of cloud storage. Many providers this service those providing different types of services in cloud storage, they vary in terms of infrastructure, platforms, APIs, facilities provided and their challenges. Such variations cause interoperability, portability, data migration. Thus, scalability becomes very challenging. All of these factors contribute to what is called the scalability of storage. We see the scalability is the availability of additional storage services in any geographical location and multiple cloud storage providers.

Similarly, one of the difficult challenges in the scalability of cloud storage is Vendor lock-in or customer lock-in who has a technology, solution or service develops. In this challenge where cloud storage customers cannot easily change service providers with there is another provider has different technology because of legal constraint, price increases, failure the company, substantial costs for data transfer, or technical incompatibilities [13].

We can look at a number of factors in determining a performance and scalability of cloud storage not only the storage capacity, or the internet to provide communicate between cloud storage providers and cloud storage customer for data storage. But, principal factor is the difficulty and importance of the geographic distribution of data that connect users to the nearest site of cloud storage. For example, users cloud can find the nearest site of cloud storage according to their geographical location. The aim of the study is to provide the integrated view for our architecture[4]. This paper reviews

some challenges of Cloud storage and presents our views on Cloud storage. At the core of all these problems, we see that the solutions for challenges should meet requirements, such as expand the database, make full use of system resources, technical function, user numbers and patterns access to these services, etc.

4 CONCLUSION

Cloud storage service provides high scalability, cost-effective performance and data services for different applications. But also overall reliance on the internet creates some challenges in taking full advantage of its services, such as storage, backup online, and also in improved performance and scalability, especially those in developing countries.

The objective of this work is to knowledge the challenges of cloud storage based on two factors the cloud's performance and the scalability. In this paper, it has been discussed and understands the scalability and performance in the cloud storage, which has shown that there is an urgent need for avoiding some challenge. In this paper, it has been discussed the scalability and performance in the cloud storage that has shown that there is an urgent need for avoiding some challenge. In our future work, we propose the following strategic approach to address the issue: (i) assess technology such as Snapshot and Disk-to-Disk in Backup technology; and (ii) select simulate, platform, or service for implementation; (iii) the importance of storage and backup data according to their geographical location.

REFERENCES

- [1] P. Mell et T. Grance, « The NIST definition of cloud computing », 2011.
- [2] D. U. R. Pol, « Developers Harness Open Source Cloud Management Platforms for Novel Applications », *Urmila R PolIJCSIT Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no 5, p. 6483-6486, 2014.
- [3] A. Shawish et M. Salama, « Cloud Computing: Paradigms and Technologies », in *Inter-cooperative Collective Intelligence: Techniques and Applications*, vol. 495, F. Xhafa et N. Bessis, Éd. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, p. 39-67.
- [4] T. HATEM, M. Ali, K. E. El KADRI, et A. NOURA, « Expansion of Architecture Cloud Storage with Snapshot and Disk-to-Disk in Cloud Computing », *Glob. J. Eng. Sci. Res.*, avr. 2016.
- [5] W. Zeng, Y. Zhao, K. Ou, et W. Song, « Research on cloud storage architecture and key technologies », in *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, 2009, p. 1044-1048.
- [6] G. DeCandia et al., « Dynamo: Amazon's Highly Available Key-value Store », in *Proceedings of Twenty-first ACM SIGOPS Symposium on Operating Systems Principles*, New York, NY, USA, 2007, p. 205-220.
- [7] F. Chang et al., « Bigtable: A Distributed Storage System for Structured Data », *ACM Trans Comput Syst*, vol. 26, no 2, p. 4:1-4:26, juin 2008.
- [8] « PNUITS: Yahoo!'s Hosted Data Serving Platform », 2008. [En ligne]. Disponible sur: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKewjcrP8KbKAhUBXh4KHdl1BKEQFggdM AA&url=http%3A%2F%2Fwww.vldb.org%2Fpvlldb%2F1%2F1454167.pdf&usg=AFQjCNG6XxVmfbS7tTcGUyRlZMoOcRclTw&sig2=uz3bxqCo4g75530p104qww>. [Consulté le: 13-janv-2016].
- [9] « Cisco Global Cloud Index: Forecast and Methodology, 2014-2019 White Paper », Cisco. [En ligne]. Disponible sur: http://cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.html. [Consulté le: 15-janv-2016].
- [10] « Performance advantages of the new Google Cloud Storage Connector for Hadoop », Google Cloud Platform Blog. .
- [11] F.Gens, « New IDC Cloud Services Survey: Top Benefits and Challenges ». IDC Exchange », 2009.
- [12] « Cloud object storage makes scalability a reality: An expert podcast », SearchCloudStorage. [En ligne]. Disponible sur: <http://searchcloudstorage.techtarget.com/podcast/Cloud-object-storage-makes-scalability-a-reality-An-expert-podcast>. [Consulté le: 16-oct-2016].
- [13] « From infrastructure delivery to service management in clouds ». [En ligne]. Disponible sur: <http://www.sciencedirect.com/science/article/pii/S0167739X10000294>. [Consulté le: 15-oct-2016].

-
- HATEM Taha, Abdelmalek Essaâdi University, Morocco, hatemtaha99@gmail.com
 - ALI Manea, Sana'a University, Yemen
 - El KADIRI. Kamal Eddine, Abdelmalek Essaâdi University, Morocco