The Seybold Report

ISSN 1533-9211

DOI 10.17605/OSF.IO/CDAJ9

PERFORMANCE ANALYSIS OF PLATFORM AS A SERVICE (PAAS) FOR STORAGE IN MULTI CLOUD ENVIRONMENT

KALP CHOBISA[1], VITTHAL KAUL[2], DAKSHAY KUMAR[3], KRISHNAKANT PHATE[4] and VINODINI GUPTA[5]

[1, 2, 3, 4, 5] School of Computer Engineering and Technology, MIT Academy of Engineering, Pune, India.
Email: [1] krchobisa@mitaoe.ac.in, [2] vskaul@mitaoe.ac.in, [3] djkumar@mitaoe.ac.in,
[4] kpphate@mitaoe.ac.in, [5] vinodini.gupta@mitaoe.ac.in

Abstract

Cloud computing is the delivery of computing, storage, networking, and other resources over the internet with a pay as go pricing. You pay for the duration you have used the service without any upfront costs. Most companies use public cloud providers like Amazon Web Services (AWS) Microsoft Azure, Google Cloud Platform (GCP) for their Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) solutions for their workflows. Every cloud service provider has some unique features in their services and they have different pricings points for the services based on your configurations. Organizations want to take advantage of these unique features and pricing points by using more than one cloud platform. This necessitates the development of a multi cloud platform that provides the services of more than one cloud provider. Other than that multi cloud provides better disaster recovery, higher availability, failure tolerance and prevents vendor lock in. A highly secure, available storage service is required to store sensitive data on the cloud. In the view of this requirement we have proposed a multi cloud platform to store data. To provide these functionalities along with low redundancy of data and low cost storage, we have implemented the Erasure code algorithm in python. We concluded that using a multi cloud platform for storage we got high availability of the data stored on the cloud at low cost. It also reduces redundancy as only the partitions of the files are stored over the cloud. This will also increase the data availability of the file as the files will be recovered in case of any cloud provider being down.

Keywords: Cloud Computing, Multi Cloud Platform, Erasure Code, Data Availability, Data Recovery, Low Redundancy, Low Cost

I. INTRODUCTION

The on-demand availability of computer computing resources, in particular data storage (cloud storage) and computing power, without the user actually maintaining the resources, is known as "cloud computing." Functions in large clouds are frequently dispersed among a number of data centers, each of which serves a single purpose. Cloud computing relies on resource sharing and generally uses a "pay-as-you-go" approach, which can aid in decreasing capital expenses.

The domain of our project revolves around cloud computing and to be precise it is regarding the emerging technology known as Multi-Cloud or Cloud Federation. The cloud computing systems of two or more service providers are linked together through the process of "cloud federation" in order to load balance traffic and manage demand spikes.

The main research area of ours is the object storage service provided by different cloud providers and to apply concepts of multi-cloud to avail the best Quality of Service (QoS), cost efficiency, reliability, effective utilization of cloud resources, and excellent performance by merging the different services available to get the best out of it.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/CDAJ9

Customers can optimize the delivery of enterprise IT services thanks to the federation of cloud resources. In terms of flexibility, price, and service availability, the federation of cloud resources enables a client to select the finest cloud service provider for their organization in order to satisfy a certain business or technological demand. Applications can run in the most suitable infrastructure conditions thanks to numerous cloud resource pools. Additionally, a company can leverage the services of multiple cloud providers to disperse workloads throughout the world and transfer data between various networks. Some of the benefits of multi-cloud

1. Freedom of choice.
2. Reliable architecture.
3. Better disaster preparedness.
4. ROI optimization.

We intend to develop our multi-cloud platform with the help of the Software Development Kit (SDK). Cloud SDK is a set of tools that you can use to manage resources and applications hosted by the various cloud platforms with the help of a piece of code. The SDK builds and deploys apps that integrate with cloud services. It allows developers to access the cloud service providers with a particular code to create applications for that specific platform, system, or programming language. A basic SDK typically includes a compiler, a debugger, and application programming interfaces (APIs). We will use the SDK to deploy the storage services on the cloud platform by taking the necessary information from the user.

One more part of the project is the utilization of erasure code. The erasure code we have used is the Reed-Solomon erasure code. The use of the erasure code distinguishes the project from other standard capacity multi-cloud staging solutions. This also helps us provide the end user with greater security and data availability functionality. Erasure code will assist us in creating partitions of the file that the user wishes to upload to the cloud. Now the partitions of these files will be uploaded accordingly on the various cloud platforms. Whenever the client needs to download the files they have uploaded, the document will be recovered from the partitions transferred, and the client can see the document they have transferred without the partitions being downloaded in the client's framework. Regardless of whether a few partitions are not accessible, the file can be recovered.

Erasure code works on the principle of parity bits and uses the inversion of the recovery matrix to recover the file from the partition even if the partitions are not available. There are two parameters when calling erasure code: partitions and shards. Partitions indicate the number of partitions into which the document will be divided. Shards represent the number of least-used partitions required to recover the document. Let's look at an example to see how the erasure code works. For a text file, text.txt, the accompanying boundaries are given: 15 partitions and 7 shards this implies that the document will be separated into 15 partitions. We needed any 7 of the 15 partitions we created to recover the document. We considered that we have erased 8 partitions, and we are currently left with just 7. We can recover the document because the

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/CDAJ9

number of partitions we have is equal to the number of partitions we need. This shows that regardless of whether a few partitions are erased, we can recover the document. This feature of erasure code will greatly improve the data availability of the file uploaded to the cloud. Regardless of whether any of the cloud providers are unavailable or down, the end user can still download their document. Likewise, since partitions are transferred, this will fundamentally lessen the expense of transferring the document on all the cloud service providers.

The erasure code will allow the end user to choose between security and the availability of the file they have uploaded. If the user selects "ensure availability," then the number of shards required to recover the file will be less, as we want to retrieve the file quickly with minimal disruption for the user. If the user wants to ensure security, then the number of shards required to download the file will be higher. This is to ensure that any unknown person will not be able to download or change the file with the few partitions available. The number of partitions that are uploaded on each platform will depend on what option the end user chooses with respect to file availability or file security.

Further, Section 2 focused on the brief literature of prevailing research issues and techniques in this domain. The methodology for the development of the project, along with the system design, was briefly explained in Section 3. Section 4 discussed the various functional parameters that show the enhancement of the workings of the multi-cloud platform while using erasure code. The results were analyzed and discussed in Section 5. Finally, conclusions from the study were made in Section 6. Section 7 tells us about the future scope and further improvements that can be made to the project.

II. LITERATURE SURVEY

Cloud Federation [1] is an idea of administration collection that resolves the monetary issues of merchant secure and supplier combination. Federation can be either horizontal or vertical. Horizontal federation can be achieved across provider domains and regions, and vertical federation can be achieved at multiple levels.

Horizontal federation can be used for redundancy and migration. [2] Describes the latest technologies in multicloud design and identifies the gaps multi-cloud developers will need to fill in the near future... There are several research projects to support the development of innovative solutions for multi-cloud environments. These solutions aim to remove vendor lock-in barriers. Following current developments in support for multiple cloud use cases, it is possible to shape a concrete understanding of what MultiCloud is and is not. However, more research is needed to present a modern solution that supports multi-cloud.

Cloud Computing [3] is a pay-as-you-go service business model that redefines the security challenges associated with external customer data. It provides a secure and cost-effective multi-cloud storage model that provides customers with data availability and secure storage. Distributes data across multiple service providers in the market to ensure data availability and secure storage for customers. To ensure the best possible availability of your data, you should

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/CDAJ9

store your data across multiple cloud service providers. In this article, they have proposed a secure and cost-effective multi-cloud storage model for cloud computing that respects the user's budget and provides the best quality of service.

[7] Distributed computing is an unclear procedure wording that can be utilized in numerous application situations. Distributed computing is a framework that comprises three layers: the resource layer, the platform layer and the application layer. There are diverse dimensions to classify cloud computing, including service boundary and service type. Public cloud, private cloud, and hybrid cloud are examples of cloud computing. Distributed computing administrations are delegated Infrastructure as a Service, Platform as a Service, and Software as a Service.

[10] In a distributed computing framework, the client's PC no longer needs to do all the truly difficult work with regards to running applications. Distributed computing management improves maintenance, provides support for different phases, and accelerates progress. It also allows for high volume prototyping and load testing, much easier than stressed or enterprise servers. Distributed storage will catch the whole market as we see investment properties rented to the inhabitants.

David Bermbach et.al [4] Trust, security and compliance with data privacy laws are some of the concerns potential customers have when using cloud computing. This can lead to a lack of investment. In this work, we attempt to reveal some insight into these compromises and check whether normal suspicions are valid. Redundant Deployment (RD) in this strategy, different providers are used for the same application logic, and requests are routed to one of the providers. This improves availability and response time. In Redundant Computation (RC), the same application logic is distributed to different providers, but each request is redirected to every cloud. Depending on your goal, you can either wait for all results or use the first result returned. Parallel computing (PC) is a data processing strategy in the cloud where requests are split at the bit level so that each cloud processes only a subset of the data. This white paper compares five different compute redundancy strategies and shows their impact on availability, processing time, work distribution across clouds, correct results, security, and cost.

Erasure Code [5] Reed-Solomon code can be used to ensure data integrity in cloud storage, but has two drawbacks: high access latency and low access throughput due to CPU limitations, and high recovery traffic that congests network links. Assume that a data object can be stored on n disks, and k disks are sufficient to access every bit of the original data. The framework needs to get to k different coded blocks from k various plates to get to the first information article. In this paper, we give an outline of the development of coding strategies for distributed storage frameworks. We consider careful recovering codes, fix-by-move recovering codes, and locally repairable codes, which can save computational assets and circle I/O above.

This paper [6] sums up the exploration status of eradication coding in appropriated capacity frameworks. We, right off the bat, present the essential thought and principal ideas of eradication coding, and point out the super specialized difficulties of incorporating deletion coding into huge scope conveyed capacity frameworks. Plan of deletion codes, information fix

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/CDAJ9

advancements, information update innovations, etc. give an extensive correlation of normal deletion codes from the parts of information encoding and information fix. Further improvement of deletion coding in dispersed capacity frameworks, including simultaneous information encoding, recovering codes with low overt repetitiveness and information disappointment anticipating.

[8] Data robustness is a key requirement for storage systems. Distributed erasure codes are a convenient way to store data across storage servers but storing data in third-party cloud systems raises serious concerns about data confidentiality. To provide strong confidentiality for messages in the storage server, users can encrypt messages using encryption schemes before encoding and storing them using erasure key schemes. In this client A encodes his message M and sends it to capacity servers. The capacity servers directly consolidate the ciphertexts with randomly picked coefficients to make a codeword image and store it. Client A requests a message from capacity servers, and key servers recover codeword images for the message. The key servers then to some degree unscramble the codeword images, and client A joins the images to get the first message.

Hussam Abu-Libdeh [9] ongoing patterns show that more organizations and associations are moving their information to distributed storage suppliers or cloud service providers. Erasure coding allows you to tolerate the loss of one or more storage providers without data loss. It's also acceptable without the overhead of strict replication. RACS mimics the Amazon S3 interface and uses the same data model. It stores data in named buckets and allows partial writes to objects, but not partial reads. RACS is a proxy that splits an object into m data shares, creates m redundant shares, and sends the m shares to n repositories. Metadata is replicated to all servers and RACS stores a small amount of metadata per release. RACS empowers distributed storage clients to investigate compromises between above cost and seller versatility.

Guanfeng Liang [11] their paper presents a strategy for optimizing the delay performance of uploading and downloading data in and out of cloud storage by using erasure coding, parallel connections to storage cloud and limited chunking. They present TOFEC, a technique that can further develop throughput-defer execution for record getting to on distributed storage using eradication coding. Files are divided into k chunks, each of which is 1/k the size of the original file, and encoded into n > k of "coded chunks" using a (n, k) Forward Error Correction (FEC) code. This improves delay performance under light workload, but reduces system capacity. User requests have two components: queuing delay (Dq) and service delay (Ds). Queuing delay is the time a request waits in the request queue, and service delay is the time it takes a thread to process a task. They had checked by using both Shared key as well as unique key and found the performance to be very similar.

To conclude, cloud computing is a technology we can use to make things easier as we don't have to do the hectic task of setting up the computing, networking, security etc for our requirement. Cloud providers will look into this on our behalf. Apart from this there are many issues that we need to face as it works on the concept of distributed computing so security is one of the major issues and there are many more. Cloud providers provide us data availability but to increase the data availability and to reduce the redundancy we can use the concept of

ISSN 1533-9211

DOI 10.17605/OSF.IO/CDAJ9

Erasure code, so that we can break the file and then store them to different servers to increase the data availability.

III. METHODOLOGY

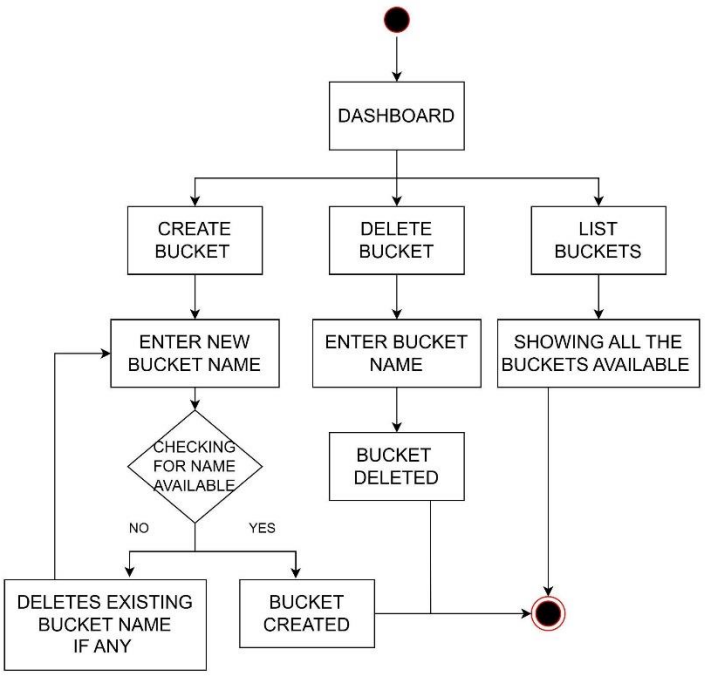

Figure 3.1. Bucket operations on multi cloud platform

This Figure 3.1 explains the processes related to creation of bucket, deletion of bucket, and listing of bucket. User will be asked to enter the name of bucket the process will check whether the name is valid or not ie. If it's not unique across the region and cross platform they will be asked to reenter the name of the bucket and when it's valid a storage space will be created on the different providers.

Similarly while deleting, the bucket name will be asked to user and whether the bucket is empty or not based on this the deletion operation will be performed. To view the available bucket list bucket option can be used.

ISSN 1533-9211

DOI 10.17605/OSF.IO/CDAJ9

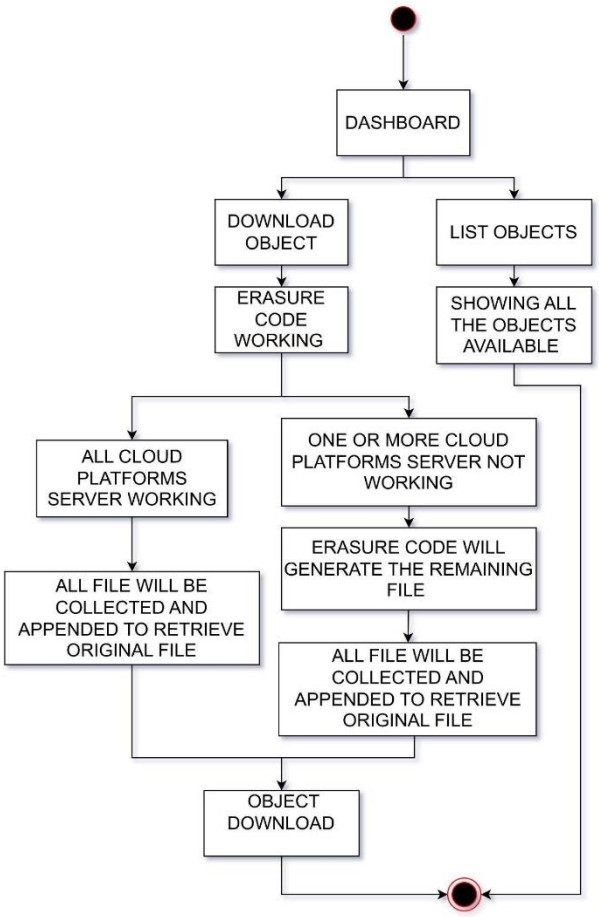

Figure 3.2 (a): Object operations on multi cloud platform

This Figure 3.2(a) explains the functionality of downloading files from the platform and listing the available objects.

For Downloading the the file name sis asked then it is checked that file parts are available or not on a particular provider (since we have partitioned the file and we require a minimum parts of file to retrieve the file) if so the the file will be fetched and given to user, but if the system is not able to fetch parts from a particular provider (ie. vendor lock in situation) then we can get parts uploaded on another provider and erasure code will generate the whole file and will be given back to the user.

In this way the files that are present on the cloud could be downloaded and users can access those files.

ISSN 1533-9211

DOI 10.17605/OSF.IO/CDAJ9

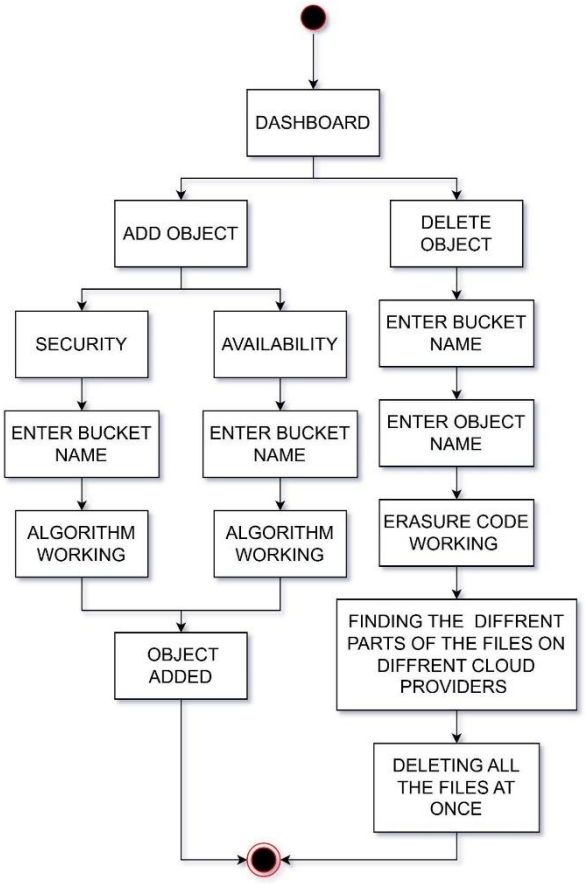

Figure 3.2 (b): Object operations on multi cloud platform

In Figure 3.2(b) here the main options are Add object and Delete object, here the user has to upload the object they will be given two options whether they have to go with security or availability. So security and availability simply means the number of shards and partitions. If its security then then the shards value will be high means more number of file parts are required to fetch the file and if its availability then the shards value is bit less as compared to security.

Similarly while deleting an object, the user will be asked the object name and the file name will be appended and then deleted from the bucket.

In this way users will be able to interact with the Multi cloud storage system and similarly other services also can be configured.

ISSN 1533-9211

DOI 10.17605/OSF.IO/CDAJ9

IV. RESULT AND ANALYSIS

The following performance matrix was analyzed on a CentOS8 64-bit virtual machine, which had the following configuration:

- Memory allocated: 6.2 GB
- Total processor core: 8
- Hard disk size: 50 GB.

1. Time taken to upload file based on number of partitions

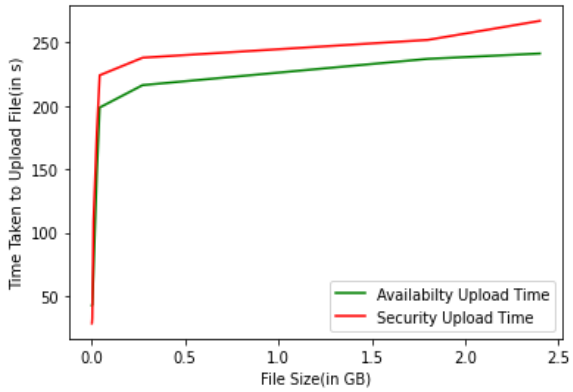

Fig 4.1

From Fig 4.1 we observed that the time taken to upload the file partitions increases as the file size increases. Further we also observe that the time taken to upload the partitions for providing high security is more than as compared to time taken to upload the file partitions for providing higher data availability.

2. Time taken to retrieve file based on no of partitions

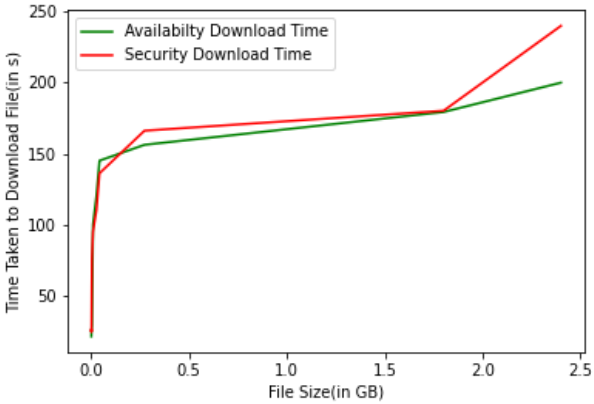

Fig 4.2

From Fig 4.2 we observed that the time taken to download the file partitions increases as the

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/CDAJ9

file size increases. Further we also observe that the time taken to download the partitions for providing high security is more than as compared to time taken to download the file partitions for providing higher data availability.

3. Time taken to partition file based on their size

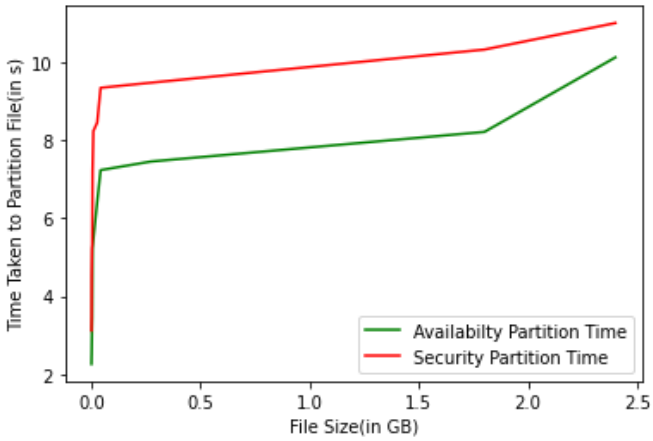

Fig 4.3

From Fig 4.3 we observe that the time taken to partition the file increases as the file size increases. Further we also observe that the time taken to partition files for providing high security is more as compared to time taken to partition the file to provider high availability.

4. Size of partitions uploaded compared to the actual file size.

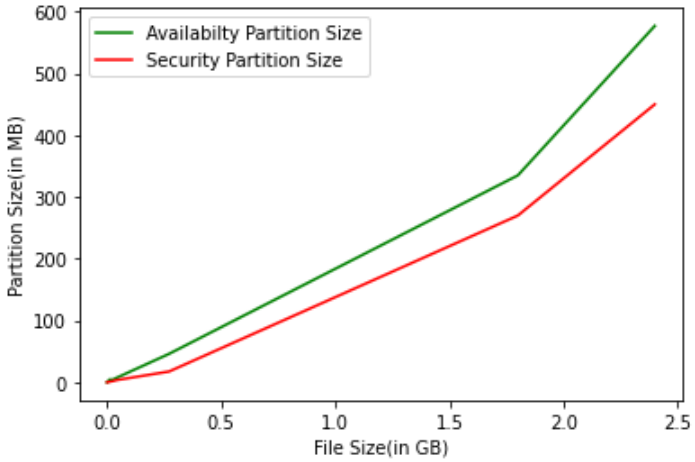

Fig 4.4

From Fig 4.4 we observe that the file size of the partition increases as we increase the file size. Furthermore we observe that the partition size for providing more availability is more as compared to the partition size for providing higher security.

ISSN 1533-9211

DOI 10.17605/OSF.IO/CDAJ9

5. Time taken to complete operations on AWS, GCP as compared to time taken to complete operations on multi cloud platform

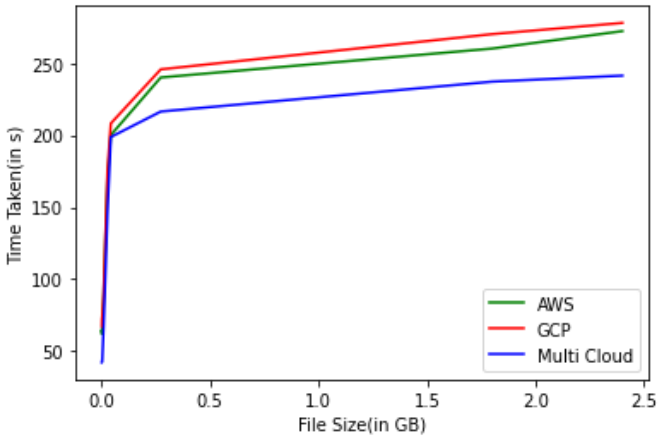

Fig 4.5

From Fig 4.5 we observe that the time taken to complete object storage operations like create bucket, upload object, etc. on AWS,GCP is more as compared to completing the same operations on a multi cloud platform .

V. CONCLUSION

In this paper we proposed a mechanism to implement storage services on a multi cloud platform. To our knowledge we have yet to come across any multi cloud platform that uses erasure code to reduce the redundancy of files stored over a multi cloud platform. We observed that the time taken to complete object storage operations like create bucket, upload object, etc. on AWS, GCP is more as compared to completing the same operations on a multi cloud platform. Time taken to upload and download the partitions for providing high security is more than as compared to time taken to upload and download the file partitions for providing higher data availability. We observe that the file size of the partition increases as we increase the file size. The time taken to download the file partitions increases as the file size increases. Furthermore we observe that the partition size for providing more availability is more as compared to the partition size for providing higher security.

VI. FUTURE SCOPE

In this paper we analyzed the performance of PaaS for storage services over a multi cloud platform. However, future work can still be performed for enabling the concept of versioning of our proposed model. Also the concept of implicit sharing the data over multi cloud platforms for a certain period of time can also be implemented in future.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/CDAJ9

References

1) Kurze, Tobias & Klems, M. & Bermbach, David & Lenk, Alexander & Tai, Stefan & Kunze, Marcel. (2011). Cloud federation. The Second International Conference on Cloud Computing, GRIDs, and Virtualization. 32-38.

2) Petcu, Dana. (2013). Multi-Cloud: expectations and current approaches. MultiCloud 2013 - Proceedings of the International Workshop on Multi-Cloud Applications and Federated Clouds. 1-6. 10.1145/2462326.2462328.

3) Singh, Yashaswi & Kandah, Farah & Zhang, Weiyi. (2011). A secured cost-effective multi-cloud storage in cloud computing. 619 - 624. 10.1109/INFCOMW.2011.5928887.

4) Bermbach, David & Kurze, Tobias & Tai, Stefan. (2013). Cloud Federation: Effects of Federated Compute Resources on Quality of Service and Cost. 10.1109/IC2E.2013.24.

5) Erasure Coding for Cloud Storage Systems: A Survey Jun Li and Baochun Li Tsinghua Science and Technology ISSNll1007-0214ll06/11llpp259-272 Volume 18, Number 3, June 2013

6) Wang, Y.-J & Xu, F.-L & Pei, X.-Q. (2017). Research on erasure code-based fault-tolerant technology for distributed storage. 40. 236-255. 10.11897/SP.J.1016.2017.00236.

7) Qian, Ling & Luo, Zhiguo & Du, Yujian & Guo, Leitao. (2009). Cloud Computing: An Overview. 5931. 626-631. 10.1007/978-3-642-10665-1_63.

8) Abu-Libdeh, Hussam & Princehouse, Lonnie & Weatherspoon, Hakim. (2010). RACS: a case for cloud storage diversity. 229-240. 10.1145/1807128.1807165.

9) Liang, Guanfeng & Kozat, Ulas. (2013). TOFEC: Achieving Optimal Throughput-Delay Trade-off of Cloud Storage Using Erasure Codes. Proceedings - IEEE INFOCOM. 10.1109/INFOCOM.2014.6848010.