IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Data Storage and Retrieval using Multiple Cloud Interfaces

Ms. Prathyakshini[1], Mrs. Ankitha K[2]

4th Semester M. Tech, Computer Science and Engineering, Sahyadri college of engineering and management, Adyar, Mangaluru, India[1]

Assistant Professor, Computer Science and Engineering, Sahyadri college of engineering and management, Adyar, Mangaluru, India[2]

Abstract: In this growing world there is enormous demand for cloud computing which is popular technology with huge benefits. As the data is increasing in business as well as in personal perspective there is demand for storing large amount of data. Cloud computing is a platform for storing the data and hence there should be some care taken towards the disaster that happens in the cloud. A simple recovery service from disaster is proposed using multiple cloud service providers. These multiple cloud service providers are set aside at different locations so that if any of the cloud fails the data could be retrieved from the remaining cloud interfaces. The proposed technique has greater availability as the data can be recovered from disaster.

Keywords: Recovery, cloud computing, Disaster, Cloud service provider

I. INTRODUCTION

For many of the services in the sectors such as financial services, medical services and other business services it is very important to protect the data. It does not matter whether the data is small or large even if there little amount of data loss it might cause lot of problems. Therefore many of the services design data recovering services in order to retrieve the data in case of any disaster that happens in cloud. Data could be periodically saved for modified data stored on the cloud.

For the sake of disaster recovery constructing a separate cloud service provider will be costlier. Whether the service provider is idle or in use the infrastructure and other costs should be paid. And hence enormous amount of investments should be done for doing this.

Pay-as–you-go model of cloud computing is beneficial because only on demand the resources are allocated and user should pay for it. There are pretty lot of experiments done on the basis of using cloud service for disaster recovery with existing resources.

The challenge is now how to do disaster recovery when the cloud fails. There might be several reason for disaster such as hardware failure, natural calamities etc. Some of the service providers use data centres of their own for storing backup data. But with complete data centre failure might cause full data loss.

This is because the infrastructure used by the various data centres will be same even if their stored at different locations. The additional risk is here more than one cloud interface failure. As more protective measures are taken the risk of data loss cannot be pushed a side.

Using various data centres from different cloud service providers are used for recovery service. The disaster recovery provider uses resources of other cloud interfaces and also data centres can be added as when it is required. This could produce higher data availability and lower the price as the resources are shared.

II. RELATED WORK

A least common denominator cloud interface is used in order to back up a file system which is termed as cumulus [2]. This type of backup system uses only single cloud to keep the backup and hence if the cloud fails the entire data is lost. The system does not give support for de-duplication. There are set of operations such as get, put, list and delete. Aggregation is used for small files for the purpose of storage. For effective storage LFS segment cleaning is used. In this only limited interface between client and server is used.

A disaster recovery model is created for web site applications which showed the low backup cost [3]. Public cloud based disaster recovery service is used and the price is lower when compared with private cloud. In this pricing estimate is also performed for the sake of analysing the price for running public cloud.

Causality based de-duplication method checks for the redundant data and removes those from transmitting to backup server [14]. As the duplicate data are removed, there is decrease in transmitting time which improves the performance. It internments the relation among the data blocks and it also removes the unchanged data.A data backup algorithm with two objectives namely seed block encryption where information can be retrieved without

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

network connection also if the data is corrupted it can be retrieved back [15]. The replica of the data is kept at separate location.

R-ADMAD utilizes ECC code in order to encode the data blocks and the place it different nodes [5]. It makes use of distributed dynamic restore for handling data recovery. This method reduces the storage space and also gives increased data availability. The storage application used in RAID and R-ADMAD are alike.

Differentiated replication method where the user is permitted to pick the replication method [4].

In order to provide greater availability the replication method is applied. This strategy offers different services based on the data availability aspect. The semantic based data backup technique which performs proper match searches and the system enables semantically driven query answering method [7].

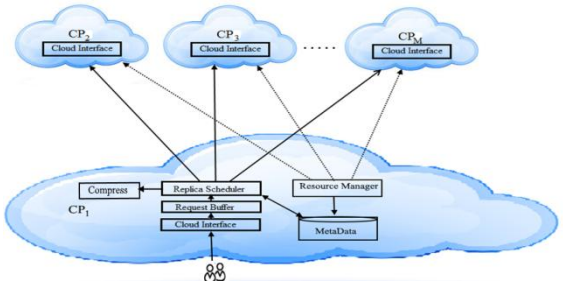

Fig.1. Backup model for recovery service

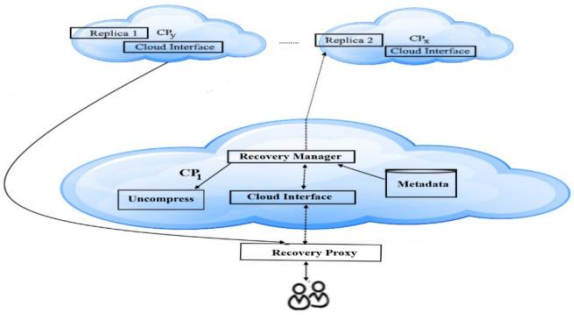

Fig. 2. Recovery model for data restore

III. SYSTEM MODEL

The Description regarding the system model is being mentioned as follows:

A. Design Methodology

The proposed model should make use of different cloud interfaces which are kept at different locations and these put forward various kinds of infrastructure, storage resources and network resources. Proposed system architecture for recovery service in cloud is depicted in the Fig. 1. and Fig. 2.which constitute data backup and recovery model. There are multiple cloud interfaces which comprise of data disaster recovery service provider. In Fig.1. Data disaster recovery service provider is given by CP_1 and the customers are the clients who are supposed to register before getting the service. In addition there are other clouds which are from CP_2 to CP_m. The request buffer‘s job is to hold the user request and the replica scheduler does replicas and place it in different locations. Especially for doing the replication and placing the data blocks in various locations erasure coding algorithm is used. This algorithm moves on by dividing the data segments and encoding it so that nobody can see the actual data. The supervision of resources is controlled by the resource manager, who is also responsible for charging the clients later based on the amount of storage as well as access cost. The information concerning where the files are stored is kept in metadata which is a database. There are cloud interfaces in Fig. 2.namely $CP\,x$ and CPy where the replicated data is stored. These locations are designated by the replica scheduler. The recovery manager is answerable for verifying the user request which includes checking their cloud accounts. At the client side the recovery proxy is answerable for restoring data from the service providers.

B. System Workflow

The whole working procedure is mentioned as follows:

1) Cloud user register in the cloud for authentication purpose with necessary requirements and by using the registered username and password user will be able to login.
2) Once the cloud user login to cloud, user can save and download required files of their choice by browsing and selecting the file from the directory.
3) The file contents are encoded and replicated using the erasure coding algorithm. Initially it is broken down into fragments and in the later stage it is encoded by adding extra bytes to the text which increases the file size.
4) The compression algorithm is additionally added for decreasing the file length. Gzip algorithm is used for compressing which is based on removing the repeated characters from the text and replacing it with the pointer. The text is then stored on to the multiple cloud interfaces which offers comparably less access and storage cost.
5) When the file should be downloaded it is fetched from the cloud interface and here the text is decoded and uncompressed before handing over to the user.

IV. EXPERIMENTS AND RESULT ANALYSIS

Using cloud sim simulator simulation is done by considering the following factors. There are three modules in the proposed model such as user portal, Storage manager and cloud interface. As user of cloud interface should be registered before using the service it should be done by giving the necessary details. Later by providing the username and the password the user can login and use the cloud services. There are set of user functions which are listed as follows: (1) File save choice where the file location is browsed and selected for save choice. Once the

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

file is saved there will be response given by the storage manager. (2) File downloads choice which is done by manually entering the file which user wants to download. Even for file download choice upon successful download the relevant messages are given by the storage manager. The cloud interfaces details such as id, access cost, storage cost and IP address is specified by the storage manager. Accordingly the clients are charged based on the file size.

A. Snapshots of Result

The user Login page is represented by the Fig. 3. The necessary requirements are supposed to be given by the user for login. After the cloud user login to the system the user operations can be performed.

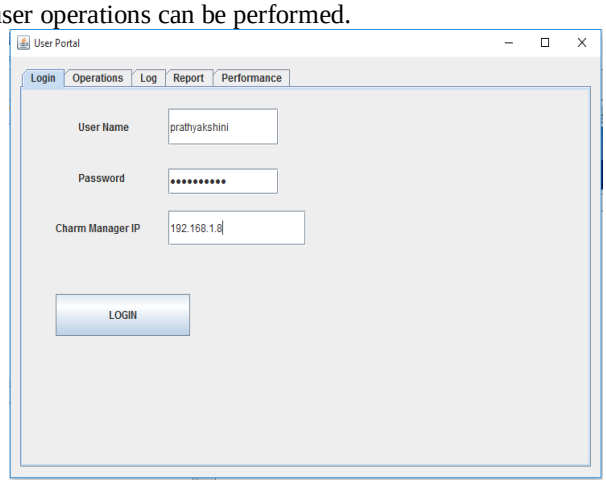

Fig.3. Cloud User Login

The cloud user operations are represented by the Fig. 4. The file save and download operations can be performed on the files. The necessary access costs and storage costs are calculated based on the file size and charged to the user later.

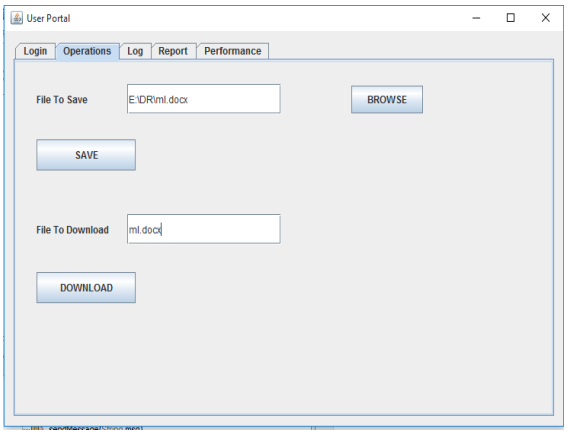

Fig. 4. User operations module

The simulation results are given by the Fig. 5. The Stepwise result is represented where the multiple cloud service providers are created with data centres which are located at different locations. From the multiple cloud interfaces the one with lowest cost is chosen and data is stored. The text in the file is divided as two shares and the copy of the same is created once it is being encoded. These are stored at multiple locations so that if any of the servers fails it can be retrieved easily from available servers. Fig. 6.shows the data blocks which are encoded and after estimation of cost the user will be charged for the same.

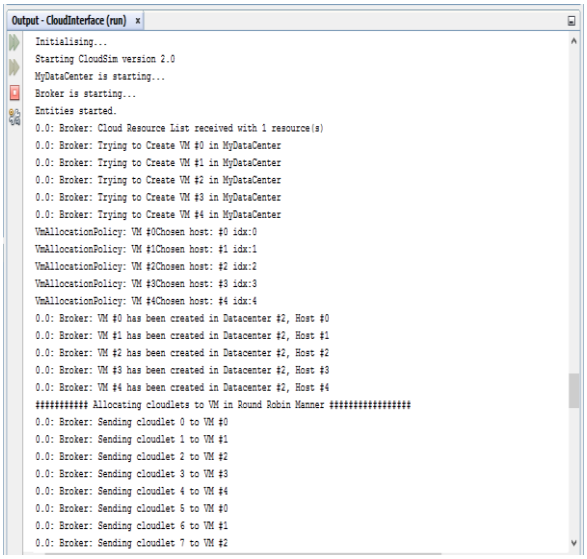

Fig. 5. Simulation Result

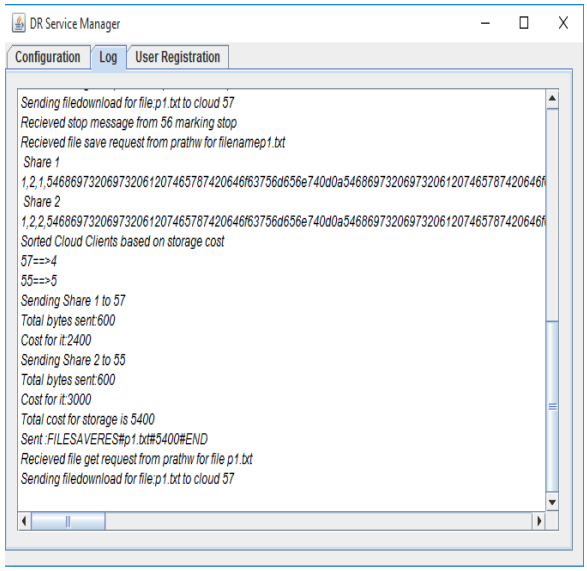

Fig.6. Allocating data blocks to the cloud interface

B. Result Analysis

The performance of the system differs as the file size varies. The access cost for the file download and the storage cost for the file save on the cloud varies according to each cloud service provider.

The system chooses the cloud interface which is having low access cost and storage cost as location to store the user data.

The access price plays an important role when the user wants to download the file. The Performance graph with access cost is shown by the Fig. 8. It is represented for different files with various file length.

The access cost is considered for estimating the overall cost charged for the cloud user.

DOI 10.17148/IJARCCE.2016.54230

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

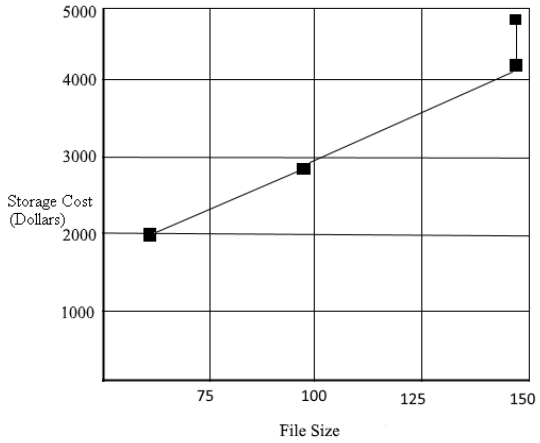

Fig. 7.Performance graph with storage cost

The performance of the system could be estimated by checking the variations with performance when the file size varies. The Performance graph with storage cost is represented by the Fig. 7. It is shown for different file sizes which represented by bytes. There are four different files are considered with different file length. The estimate price is calculated by considering the total number of bytes. With more file size the cost is estimated to be greater. This graph is shown for smaller file sizes with varied length.

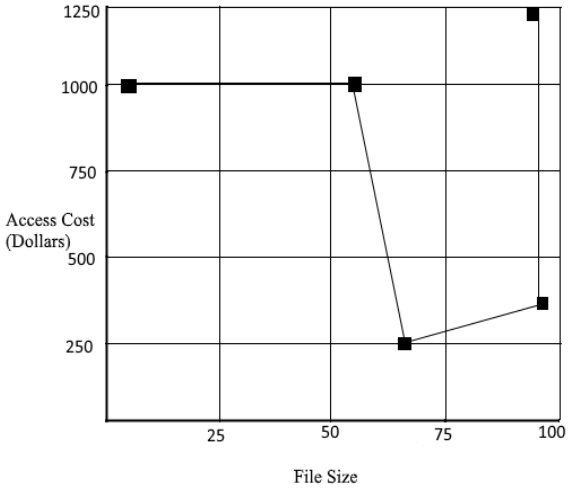

Fig. 8. Performance graph with access cost

When the file length is seemed to be smaller the cost charged will also be lower. But File size is higher the cost charged will also increases accordingly.

V. CONCLUSIONS AND FUTURE WORK

Using multiple cloud interface the recovery service from the disaster is developed where the resources are used supportively by the cloud service providers.

The user data is replicated and stored on to different locations by using replication method. The data is also compressed and stores which minimizes the storage space required. This gives maximum data availability and also greater reliability.

In future for the sake of improving network resource consumption, resource usage de-duplication method can be supported to enhance the performance of the system.

REFERENCES

[1]. Yu Guet.al. "DR-Cloud: Multi-Cloud Based Disaster Recovery Service", IEEE, Tsinghua Science and Technology, vol.19, pp.13-23, February 2014.
[2]. Michael et.al. "Cumulus: File system Backup to the Cloud", ACM Transactions on Storage, vol.5, December 2009.
[3]. Timothy Wood et.al. "Disaster Recovery as a Cloud Service: Economic Benefits and Deployment Challenges", USENIX Workshop on Hot Topics in Cloud Computing, 2010.
[4]. zung Nguyen et.al., "Differentiated Replication Strategy in Data Centers", IFIP International Federation for Information Processing, pp.277–288, 2010.
[5]. Chuanyi Liu et.al., "R-ADMAD: high reliability provision for large-scale de-duplication archival storage systems", in Proceedings of the 23rd International Conference on Supercomputing, vol.1, pp.370-379, 2009.
[6]. Timothy Wood et.al., "Pipe Cloud: Using Causality to Overcome Speed-of-Light Delays in Cloud-Based Disaster Recovery", in Proceedings of the 2nd ACM Symposium on Cloud Computing, October 2011.
[7]. Giuseppe et.al., "ERGOT: A Semantic-based System for Service Discovery in Distributed Infrastructures", 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing , 2010.
[8]. Chi-won Song et.al., "Parity Cloud Service: A Privacy-Protected Personal Data Recovery Service", International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11, 2011.
[9]. C. Liu, Y. Gu, L. Sun, B. Yan, and D. Wang, RADMAD: High reliability provision for large-scale de-duplication archival storage systems, in Proceedings of the 23rd International Conference on Supercomputing, 2009, pp. 370-379.
[10]. James S P lank, "Erasure Codes for Storage Systems", usenix Vol. 38 No. 6, December 2013.
[11]. Lili Sun, et.al. "Recovery Strategies for Service Composition in Dynamic Network," International Conference on Cloud and Service Computing, 2011.
[12]. Mohammad Ali et.al., "Disaster Recovery in Cloud Computing: A Survey", Canadian Center of Science and Education Computer and Information Science, Vol. 7, 2014.
[13]. Kruti Sharma, Kavita R Singh, "Online Data Back-up and Disaster Recovery Techniques in Cloud Computing: A Review", International Journal of Engineering and Innovative Technology (IJEIT), Volume 2, November 2012.
[14]. Y. Tan, H. Jiang, D. Feng, L. Tian, and Z. Yan, "CAB dedupe: A causality-based de duplication performance booster for cloud backup services", in Parallel & Distributed Processing Symposium (IPDPS), 2011, pp. 1266-1277, 2011.
[15]. Ms. Kruti Sharma and Prof. Kavita R Singh, "Seed Block Algorithm: A Remote Smart Data Back-up Technique for Cloud Computing", International Conference on Communication Systems and Network Technologies, IEEE, 2013.