

A survey on various backup and recovery technique in cloud

Shilpi U. Vishwakarma

Computer Science and Engineering
P R Patil College of Engineering. & Technology
Amravati, Maharashtra, India
aditi12vish@gmail.com

Prof. Pravin D. Soni

Computer Science and Engineering
P R Patil College of Engineering. & Technology
Amravati, Maharashtra, India
pravindsoni@gmail.com

Abstract— Cloud computing is an emerging technology that allows their user to store large amount of data in cloud storage and it can be use whenever required, from anywhere. It provides numerous kinds of services to its users such as Software as a service, Infrastructure as a service, and Platform as a service etc. Storage-as-a-service is one of the services provided by cloud infrastructure in which large amount of data is stored in cloud. As valuable and vital data of enterprises are stored at a remote location on cloud so we must be assured that user's data must be safe and be available at any time. In situations like Flood, Fire, earthquakes or any hardware malfunction or any accidental deletion our data may no longer remain available. To maintain the data security and availability there must be some data backup technique for cloud platform to recover valuable and vital data efficiently. This paper provides a review on various backup techniques used for Cloud Computing platform regarding this concern.

Keywords: Cloud Computing, cloud mirroring, Recent Activity Table (RAT), message digest.

I. INTRODUCTION

Cloud Computing is a large-scale distributed computing model that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet as pay per use [1]. Therefore we call a cloud an umbrella term that it refers to Internet based development and services which can be access over the network on-demand basis that are provided by the various Service Provider's (SP's). Also it provides a variety of computing resources from servers and storage to enterprise applications for example email, security, backup/DR, voice, all delivered over the Internet.

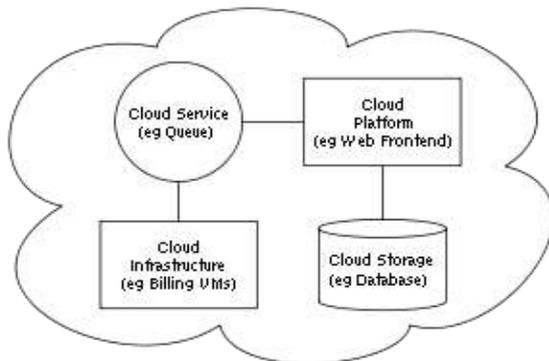


Diagram 1. Cloud Architecture

Cloud Services: -_Cloud Computing provides various types of services like Software as a service, Platform as a service, Infrastructure as a service, Storage as a service etc.

1. *Software as a service:* This service is highly scalable Internet based applications because it is hosted on the cloud and offered as services to the end user. For eg. Google Docs, acrobat.com, Salesforce.com etc.
2. *Platform as a service:* This service provides the best platform to design, develop, build & test applications are provided by the cloud infrastructure. For example

Azure service platform, force.com, Google App Engine etc.

3. *Infrastructure as a service:* This service is pay per use model, because it provides services like storage, database management & compute capabilities that are offered on demand. For example Amazon Web Services, GoGrid, 3 Tera.
4. *Storage as a service:* This service provides the data storage where the user can store the data at remote location (disk) and accessible from anywhere anytime. It maintains user's data and information.

As we know that Cloud Computing provides sharing of computing resources that is there are large number of users that share the same storage and other computing resources. Therefore there is a strong need for a mechanism to prevent other users to access your important and useful data either intentionally or accidentally also if it happens that some other user on the cloud access your data and makes some modification or any deletion then it must be recoverable to its original state in an efficient way. Also the stored data is at danger due to any natural calamities such as any Flood, Fire etc. Natural disasters for example flood can make recovering data an impossible task. Flood water contains contaminated water which may contain dust, sand dregs and other materials that may affect the platters and sectors of the hard drive. The parts of the hard drive may seize up just as an engine that will start slugging in it. Also many businesses that relied on electronic data suffer total or temporary data loss due to hardware damage and failure. Data integrity is another issue while recovering lost data [2].

Backup and Recovery

Now a day's computers are essential part of life, as computers are used for research, in business applications, or in particular organization or in any such field. As the use of computers in our day-to-day life increases tremendously, so the computing resources that we need also grow up.

For big companies like IBM, Google, Microsoft, they all can control and make use of resources properly when they found problems like crashes, server failure etc. But for smaller

enterprises, it becomes a huge factor. With the huge infrastructure come problems like machines failure, hard drive crashes, software bugs, etc. This might be a big headache for such enterprises. Cloud Computing offers a solution to this situation. Cloud Computing is the idea of storing files, accessing files, software and computing services through the Internet. But the big issue is data recovery.

II. LITERATURE REVIEW

In our study, we found many techniques that has their own unique ways to create backup and recovery. These all techniques focused on three different aspects, such as cost controlling, data duplication and security issues but in all of these techniques performance of the system is not satisfactory, low implementation complexity, redundancy and recovery in short span of time. Each of the technique has the complete focus on their aim of backup and recovery. Further, we detail few recent techniques PCS [3], HSDRT [4], Linux Box [5], ERGOT [6], Cold and Hot back-up technique [7], SBBR [8], REN [9], and SBA [10]. The detailed survey of the following technique is as follows:

1. PCS(Parity Cloud Service):

This technique provides a novel data recovery service framework for cloud infrastructure, The Parity Cloud Service (PCS) provides a privacy-protected personal data recovery service. In this framework user data is not required to be uploaded on to the server for data recovery. [3] All the necessary server-side resources that provide the recovery services are within a reasonable bound. The advantages of Parity Cloud Service are that it provides a reliable data recovery at a low cost but the disadvantage is that its implementation complexity is higher. This technique work on the following three steps those are:

- 1.1. Initialization Process
- 1.2. Parity Block Update
- 1.3. Data Block Recovery

2. HSDRT(High Security Distribution and Rake technology):

HSDRT makes use of an effective ultra-widely distributed data transfer mechanism and a high-speed encryption technology. It consists of two sequences one is *Backup sequence* that is it stores the data which is backed up and other is *Recovery sequence* which is used whenever there is disaster or any data loss occurs the Supervisory Server (the components of the HSDRT) starts the recovery sequence. [4]

Because of some limitations in this approach this model cannot be declared as a perfect technique for Cloud back-up and recovery. Even though this model can be used for movable clients such as laptops Smart phones etc. the data recovery cost is comparatively increased and also it increases the redundancy. HSDRT technique consist of following three component those are:

- 2.1. Data Center
- 2.2. Supervisory server

2.3. Client node specified by admin (PCs, Smart phones etc.).

3. Linux Box:

Linux box technique reduces the cost of the solution and protects data from disaster. This makes the process of migration very easy that is transfer data from one cloud service provider to other. This technique is affordable to all consumers and Small and Medium Business (SMB). It eliminates consumer's dependency on the ISP (Internet Service Provider) and its associated backup cost. *A simple hardware box can do all these at little cost named as simple Linux box which will sync up the data at block/file level from the cloud service provider to the consumer.* It integrates an application on Linux box that will perform backup of the cloud onto local drives.

The application will interface with cloud on a secured channel, check for updates and sync them with local storage. The data transmission will be secure and encrypted. After a valid login, the application secures the channel using IP Security and in-flight encryption techniques. The application then interacts with the application stack at the cloud service provider and does a onetime full backup. During subsequent check, it backs up only the incremental data to the local site. The limitation is that a consumer can backup not only the Data but Sync the entire Virtual Machine[5] which somehow waste the bandwidth because every time when backup takes place it will do back-up of entire virtual machine.

4. ERGOT(Efficient Routing Grounded on Taxonomy):

ERGOT technique is fully based on the semantic analysis which does not focus on time and implementation complexity. It provides the support for Service Discovery in cloud computing. [6] Building components of ERGOT is:-

1. A DHT (Distributed Hash Table) protocol
2. A SON (Semantic Overlay Network)
3. A measure of semantic similarity among service description.

This technique provides the retrieval of data in an efficient way that is totally based on the semantic similarity between service descriptions and service requests. This technique proposes a semantic-driven query answering approach in DHT-based systems by building a SON over a DHT but it does not go well with semantic similarity search models. The downside of this model is an increased time complexity.

5. CBSRS(Cold/Hot Backup Service Replacement Strategy):

Suggested technique is a gradual increase in cost with the increase in data i.e. The Cold and Hot back-up strategy that performs backup and recovery on trigger

basis of failure detection. In **CBSRS** (i.e. Cold Backup Service Replacement Strategy) recovery process, it is triggered when a service failure is detected and it will not be triggered when there is no failure i.e. when the service is available. The **HBSRS** (i.e. Hot Backup Service Replacement Strategy), is a transcendental recovery strategy for service composition that is used for dynamic network.[7] During the implementation of process, the backup services remains in the activated state and the first returned results of services will be used to ensure the successful implementation of service composition.

6. SBBR(Shared backup router resources):

This technique mainly focuses on the significant reduction of cost and router failure scenario i.e. (SBBR). It involves logical connectivity of IP that will be remain unchanged even after a router failure. The most important factor of this model is that it provides the network management system via multi-layer signaling. [8] Additionally this model shows how service imposed maximum outage requirements that have a direct effect on the setting of the SBBR architecture (e.g. imposing a minimum number of network-wide shared router resources locations).The problem with model is that it is unable to include optimization concept with cost reduction.

7. REN(Rent Out the Rented Resources):

This technique provides the lowest cost point of view a model “Rent out the Rented Resources”. This technique focuses on reducing the cloud service’s monetary cost. It proposed a model for cross cloud federation which consists of three phases that are 1) Discovery, 2) Matchmaking and 3) Authentication. This model is simply based on the concept of cloud vendors that rent the resources from different venture(s) and after virtualization, rents it to the clients as cloud services.[9]

8. SBA(Seed Block Algorithm):

The remote Backup server is a replica of original cloud server which is physically situated at a remote location. This method is based on the concept of Exclusive-OR (XOR) operation of digital computing. The whole mechanism consists of three main parts

1. The Main Cloud Server
2. Clients of the Cloud and
3. The Remote Server.

The SBA uses a random number and a unique client id associated with each client. Whenever a new Client is get registered with the cloud its unique client id is get XOR with a random number. The result of this XOR operation is called as a Seed Block which will be used only for that particular client. Whenever a client stores any Data on to the Cloud it is saved in Cloud and at the same time it is XORed with its Seed Block and the resultant Data’ is stored in the remote server. If any accidental data loss occurs in the main Cloud then in such cases the original data is

recovered by XORing the Data’ with the Seed Block of that particular client to obtain Data’’ i.e. the original Data file. [10]

This technique is fully capable of recovering the data files accurately in any data loss situation also at the same time it maintains data integrity. The downside of this technique is that it is inefficient because the data files on the remote server uses the same space as in the main Cloud so in this way there is wastage of storage space. The storage space in the remote Server can be reduced by applying the compression techniques to achieve high efficiency.

Overall Comparison of the below following techniques: -

Sr. No.	Techniques	Pros	Cons
1	Parity Cloud Service[3]	Privacy, Economical	High Complexity
2	HSDRT[4]	Used for movable clients like Laptop, Smart Phones	Increase Redundancy, Costly
3	Linux Box[5]	Economical, simple Implementation	High bandwidth, Complete server backup at a time
4	ERGOT[6]	Exact match retrieval, privacy	Increased complexity Time and Implementation
5	CBSRS/HBSRS[7]	Triggered only when failure detected	Cost increases as data increases
6	SBBR[8]	Works even if router fails, It concerns with cost reduction	Unable to include optimization concept with cost reduction, Inconsistencies between logical and physical configurations
7	REN[9]	Cost depends on the infrastructure utilization	Resources must kept under special attention due to rented concepts
8	SBA[10]	Simple to implement	Inefficient

III. PROPOSED METHODOLOGY

Data is always important assets of any user but no one can assure that his data cannot be corrupted or lost, so recovery plays a crucial role in such scenarios. We have seen that

various techniques have been proposed till date for data backup and recovery but all of these techniques have their own limitations which need to be overcome. With the help of cloud mirroring technique we provide the high availability, integrity as well as recovery of user data (files). We have proposed file recovery technique by the concept of cloud mirroring. [13] The advantages of using this technique is: though we are using hard disk for file recovery, eventually the cost for recovery will be reduced along with this the proposed technique is applicable to any kind of cloud. This technique will focus on entire mirroring of cloud as we are using the asynchronous mirroring the overhead of the RAT.

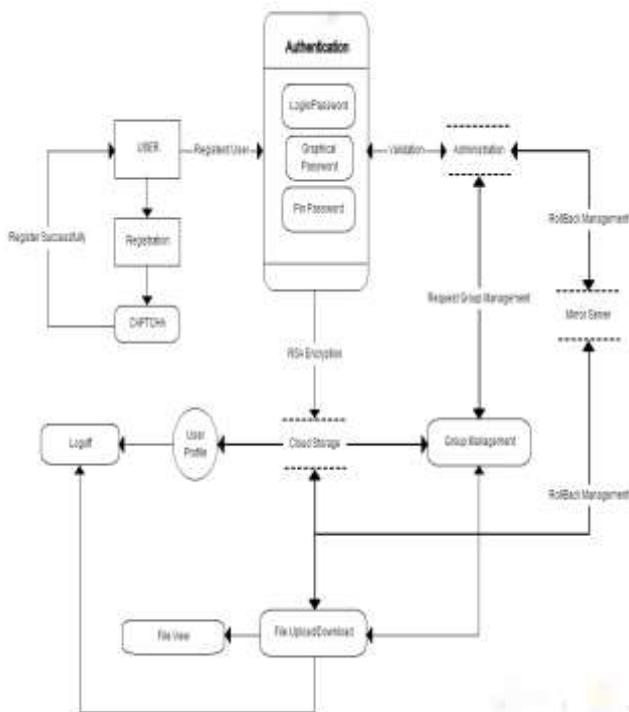


Diagram 2: DFD of Basic Block Diagram

IV. CONCLUSION

All the above techniques tried to cover different issues of data backup and recovery for Cloud Computing such as maintaining the cost of implementation and implementation complexities as low as possible. However each one of the

backup solution for Cloud Computing is unable to achieve all the issues of remote data back-up server with less storage space.

REFERENCES

- [1] Ian Foster, Yong Zhao, Ioan Raicu, Shiyong Lu, *Cloud Computing and Grid Computing 360-Degree Compared*, Grid Computing Environments Workshop, 2008..
- [2] Somesh P. Badhel *et al*, International Journal of Computer Science and Mobile Computing, Vol.3 Issue.12, December-2014, pg. 538-542 © 2014, IJCSMC All Rights Reserved 539
- [3] Chi-won Song, Sungmin Park, Dong-wook Kim, Sooyong Kang, 2011, "Parity Cloud Service: A Privacy-Protected Personal Data Recovery Service," International Joint Conference of IEEE TrustCom-11/IEEE ICSS-11/FCST-11.
- [4] Yoichiro Ueno, Noriharu Miyaho, Shuichi Suzuki, Muzai Gakuendai, Inzai-shi, Chiba, Kazuo Ichihara, 2010, "Performance Evaluation of a Disaster Recovery System and Practical Network System Applications," Fifth International Conference on Systems and Networks Communications, pp 256-259.
- [5] Vijaykumar Javaraiah Brocade Advanced Networks and Telecommunication Systems (ANTS), 2011, "Backup for Cloud and Disaster Recovery for Consumers and SMBs," IEEE 5th International Conference, 2011. M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
- [6] Giuseppe Pirr'ò, Paolo Trunfio, Domenico Talia, Paolo Missier and Carole Goble, 2010, "ERGOT: A Semantic-based System for Service Discovery in Distributed Infrastructures," 10th IEEE/ACM International Conference on Cluster, Cloud and Grid] Lili Sun, Jianwei An, Yang Yang, Ming Zeng, 2011, "Recovery Strategies for Service Composition in Dynamic Network," International Conference on Cloud and Service Computing.
- [8] Eleni Palkopoulou, Dominic A. Schupke, Thomas Bauscherty, 2011, "Recovery Time Analysis for the Shared Backup Router Resources (SBRR) Architecture", IEEE ICC.
- [9] B. Cully, G. Lefebvre, D. Meyer, M. Feeley, N. Hutchinson, and A. Warfield, emus: high availability via asynchronous virtual machine replication, in Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation, 2008, pp. 161-174.
- [10] Ms. Kruti Sharma, Prof. Kavita R Singh, 2013 "Seed Block Algorithm: A Remote Smart Data Back-up Technique for Cloud Computing" International Conference on Communication Systems and Network Technologies IEEE.
- [11] S Sankareswari, S Hemanth "Attribute Based Encryption with Privacy Preserving using Asymmetric Key in Cloud Computing" , (IJCSIT) International Journal of Computer Science and Information Technologies, may 2014.
- [12] S.U.Vishwakarma, S.D.Pande, S.S.Ganorkar" An Information Security Scheme for Cloud based Environment using 3DES Encryption Algorithm" International journal of recent and development in engineering and technology(IJRDET), volume 2, issue 1, Januray 2014.
- [13] Shilpi Vishwakarma, Praveen D. Soni "Cloud Mirroring: A Technique of Data Recovery" International Journal of Current Engineering and Technology, Vol.5, No.2 (April 2015).