International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

Data Backup and Recovery Methods in Cloud Computing

Danish Nazir[1*] and Mir Aman Sheheryar[2]

[1]Department of Information Technology, School of Engineering, Central University of Kashmir, Srinagar, India
[2]Department of Information Technology, School of Engineering, Central University of Kashmir, Srinagar, India

**Corresponding Author: danishnazir18@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 15/May/2018, Published: 31/May/2018

***Abstract*—** Providing different scenarios of service is the prime bedrock of Cloud computing. Cloud computing offer various services to its associated users. The service offered by cloud depends up on the nature of service subscribed by users associated with it. Storage as-a-service is one of the offered services provided by cloud via internet to its clients. Huge data resides in cloud storage management. To ensure security, cloud shall guarantee that our information is protected in all odds and be accessible whenever need arises. In circumstances like Fire, Flood, Tremors, Technical Snags, Equipment Failure and Coincidental deletion, our data access may be lost. To keep up the effectiveness of the outsourced data there need to be timely implementation of Data Backup involving Recovery Services. This paper provides the comprehensive investigation of different backup methods utilized for Cloud Computing with respect to technical snags and provide the gateways to access shortcomings in cloud backup aspect.

***Keywords*—** Technical Snags, Central Repository, Remote Repository, Parity Cloud Service, Seed Block.

I. INTRODUCTION

"Cloud computing is a model for enabling ubiquitous, 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" [NIST]. As we have the definition of Cloud computing, it can provide its end users with those service that fall within cloud domain. Cloud Computing is being able to serve its users through the internet, at anytime from anywhere. This is the reason for the exponential growth of this computing architecture. Cloud Computing is often known for its three basic services like "Infrastructure as a Service (IaaS)", "Platform as a Service (PaaS)", and "Software as a Service (SaaS)". Among them the basic service is IaaS, the other two services PaaS and SaaS are the secondary services.

The infrastructure mainly consists of storage, processors, and network devices. Hence Data Storage is a sub service of Infrastructure as a service. In the current era data is being generated at an unprecedented rate from the social networking sites, web based keeping money, web based shopping destinations, webpages, pharmaceutical companies, bioinformatics and other sources. These companies generate petabytes of data in days and weeks. Ultimately, this data needs to be stored somewhere where it remains safe and easy accessible. In spite of the location from which it is being accessed. The suitable solution to this concern is Cloud data storage, which is safe, and available all the time with proper replicas at different places. Apart from the data storage, data recovery is an important issue, to recover the data once it has been destroyed by some unavoidable reason. Earlier, magnetic tapes were used for the data storage and archival, which were processed and stored at some safe location. All this procedure was a hand operated job, with require full human intervention [1]. This method was tedious and slow, with less security. Recently, with the advancements in the form of cheap, improved and online storage, storage and retrieval have become comparatively better. Cloud Computing presents its user the facility of online storage, along with the backup of data. The data from main sites is being backed up on different geographical location so as to improve the fault tolerance and availability of data. The number backup copies depend upon the replication factor of particular Cloud Service Provider. There are many Cloud Service Providers which serve as capacity sellers, for example, Amazon S3, Glacier, and Rack space. Also there are many other Cloud Service Providers like Zamanda which use Cloud storage like Amazon S3 to provide backup services to its end users. Depending upon the sensitivity level of data we have, organizations may use a hybrid approach to store the data. If the data is confidential and needs more security then it is better to be stored on traditional infrastructure while, if the data is non sensitive it can be stored on the Cloud, which provides better backup and efficient retrieval of data as discussed in **[**1].

Every organization which are generating the data have to do the efficient backup for their critical data. As also the grand associations generate data at a great rate and they are unable to store that data on their local servers and they would like to store their data in the third party control, so that when they can efficiently retrieve the data even if the data gets destroyed by unavoidable natural disasters [2]. This is because of the way that Cloud Paradigm offers us data storage along with the data backup. The Cloud connected users store their data that is critical on the cloud which need to have trust upon the concerned third party associations i.e., cloud Service provider that their data is stored safely without any external manipulation. For any user information Integrity assumes a vital part while recouping the information lost hence the point arises that there is a need of data backup and recovery methods that are highly efficient [2].

Most of the organizations are shifting their infrastructure to the third party control by exercising the Infrastructure as a Service of Cloud platform. This kind of service lets the Cloud user to access the storage, processing units, network devices through an interface i.e. Virtual Machine which are hosted on cloud Infrastructure. The use of IaaS frees the user to buy the costlier storage and to maintain it, as all kind of maintenance is done by the Cloud Service Provider. IaaS providers make use of redundancy in their infrastructure so as to avoid various kinds of disasters [3].

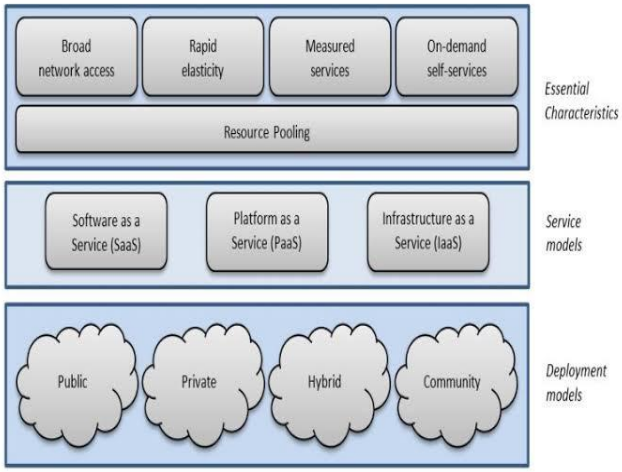

Figure 1: NIST definition of cloud computing

The bloom of Cloud Computing services is getting increased by each passing day and so the is data stored on it. The paramount data being stored on the cloud infrastructure leads to a number of issues such as data loss, data tampering, unauthorized access and other security and privacy concerns. To overcome the issues of data deletion by mistake or by some disaster, there is need of some efficient data storage and retrieval mechanisms so that the data does not get lost at any cost. The data stored on the Cloud may be financial, in case data is lost and there is no efficient data recover mechanism, whole of the business will get lost [6]. Hence, there must be some data storage mechanism in place which will store the data in a cost effective and secure manner. Moreover, it should utilize the storage in an efficient manner with efficacious recovery.

II. RELATED WORK

(Mayuri Tidke et al) [4] in year 2016 presented a paper in which they proposed a framework for data Backup and Recovery in cloud computing. Without network availability the proposed framework helps the clients to gather data from any remote area. In the proposed framework there are two servers: main server and a backup server/remote server, the remote server stores the whole data of a main server and is located at a remote location. Remote data centre comes into place to get the data if the main servers data is lost due to some reasons.

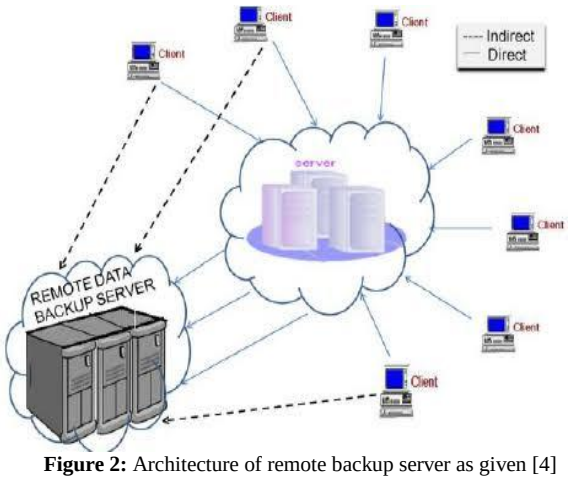

Figure 2: Architecture of remote backup server as given [4]

In the proposed algorithm a random number r is generated and is XORed with client_id of each client(C_i) to get seed block (S_i) of each C_i and then stores S_i in remote repository. If client creates/modify any file z and stores at main repository, z' is created as z'=z' XOR S_i (seed block of that client) and z' is stored at remote repository. Now if server gets crashed or data is lost due to some reasons in main repository then backup repository is used to get the file z as z = z' XOR S_i and z is returned to C_i. Therefore, data backup and recovery is achieved.

(Shilpi U. Vishwakarma, et al) [5] in year 2015 presented a research article for data recovery. By this method user can recover data that is mirrored. There are two parts. First part is uploading and second part is downloading. In uploading module Recent Activity Table(RAT) is maintained which stores the information of most recent file that are uploaded by the user. The steps involved in uploading a file are: Firstly, on cloud, the user uploads his/her file and after the file is uploaded successfully the RAT stores the files

information along with user. Mirroring algorithm is used to mirror the uploaded data to hard drive. In downloading module user requests for data, it will check if the requested file is in the main cloud and if not the respective file is retrieved from mirror hard drive.

(P. S. Challagidad, et al) [6] in year 2017 presented a paper in which a multi-server framework in light of Enriched Genetic Algorithm to recuperate the lost information by utilizing four cloud reinforcement servers is examined. The proposed design consists of three modules Remote repository, main repository and number of Clients. Remote repository keeps up the recreated duplicates of main repository. The main repository stores all the client information. The client transfers the record to main cloud server; the main repository stores every one of the information in backup repository. In the event that client needs to recover the document from cloud at that point record is looked in principle cloud server initially, if the information is absent in primary server then the information is checked in the backup repository to recover lost information. If information is lost due to some reasons. To get the information that is lost due to some reason a recuperation strategy is basic. Recovery of information can be accomplished using proposed calculation effectively. To give the unwavering quality two or four backup information cloud stockpiles could be utilized. The replicated copies of information are kept up in excess of one server to recoup information. At the point when information misfortune happens at one area it can be recovered from other reinforcement server utilizing Enriched Genetic Algorithm (EGA).

(Megha R, et al) [7] in year 2016 proposed a novel double backup framework where the information is moved down in a local drive before it is relived to the cloud. In the event that data lose the framework first tries to discover a duplicate of data put away in a local drive and if inaccessible then just it recovers the data from the cloud. The proposed system takes the benefits of Drobox. The paper discusses the issue that the present framework has likewise got downside at whatever point client manages video. Bandwidth requirements are huge for uncompressed video. So as to defeat the bandwidth issue we should compress the video. Zip files need to be re-encoded as video compression cannot be done with them. Before the video is uploaded automatically into the cloud, it must be automatically compressed at the background.

(Chi-won Song, et al) [8] in year 2011 proposed a new information recuperation benefit structure for cloud foundation, the Parity Cloud Service (PCS) gives a security ensured individual data recuperation benefit. In this method that they have figured client information need not to be transferred on to the third party service provider I.e. server for the need to do recovery. All the critical server-side resources that give the recovery administrations are inside a sensible bound. it uses Ex-or operation to store data for backup. The upsides of PCS are that it gives a solid data recovery requiring little to no effort however the drawback is that its usage many-sided quality is higher.

(Tanay Kulkarni et al) [9] in year 2015 have proposed a smart remote data backup plan using AES and SBA to cater the problem that electronic data generated in cloud computing are large in amount and should be available to user whenever needed. In this method there are three entities the main repository, the backup repository and the clients of main cloud. In this method first a random number 'r' is generated. After the random number is generated, seed block for each client is created. The seed block is created as $S_i = r$ XOR C_i (where S_i is the seed block of C_i). if a client C_i creates or modifies any file say f and stores on main server, then f ' is created after applying AES to f '' and is created as f '' = f ' XOR S_i and then f '' is stored in remote backup server. Now if main server crashes and file f gets deleted then backup server comes into place and we can recover our file back by doing f ' = f '' XOR S_i and return f to C_i after decrypting f '. Hence data backup and recovery achieved.

(Yoichiro Ueno et al) [10] in year 2011 presented a paper in which the HSDRT encompasses three components viz the principal capacities are data centres, supervisor server and customer hubs indicated by administrator. The third component I.e. the customer hubs are connected with a supervisory server by means of protected system. The data centre component encodes the discontinuities again at the stage second and after that circulates to the customer hubs. There are certain limitations in the implementation of HSDRT viz the web application is important to be balanced to utilize the HSDRT unit and another is when the quantity of copied duplicate of information builds the processor execution.

III. COMPARATIVE ANALYSIS

The favourable points of interest and drawbacks of all the above talked about strategies are depicted in the Table-I. Likewise, in light of the high suitability and need of reinforcement process in various associations and endeavours, the piece of a remote data backup server with a successful framework is imperative.

Table 1: Comparative analysis

S. No.	Approach	Advantages	Disadvantages
1.	Seed Block Algorithm [4]	• Data availability • Security without using any present encryption standard	• Memory inefficient • Bandwidth inefficient
2.	Cloud mirroring : A technique of data recovery [5]	• High availability • Integrity • Less cost of recovery	• Memory inefficient • RAT overhead
3.	Efficient and reliable data recovery technique in cloud computing[6]	• Efficient reliable	• Four backup servers
4.	"A cloud based Automatic recovery and backup system with video compression" [7]	• Saves bandwidth • Saves storage space • Higher reliability	• Complex
5.	Parity Cloud Service [8]	• low cost • privacy	• unable to control implementation complexity
6.	Intelligent Cloud Security Backup System [9]	• More security by using AES • Availability	• Memory inefficient • Complex • Bandwidth inefficient
7.	HSDRT [10]	• Increased privacy • Exact Match retrieval	• Increased cost • Increased redundancy

In the method provided by [4] by entering the personal information we have to register first. The user can login using credentials I.e. username and password only if there is successful completion of registration. To identify the user at the time of downloading of the file the user information is stored in the database for that worry. After successful login the next step is uploading a file. The file that is to be uploaded can be browsed from our computer. When we select any file from the computer at that time file type and secret key are automatically generated for that file. The acknowledgement message is generated after the successful completion of uploading. If because of any reason the document gets erased or slammed in the main cloud server then the file can be regenerated back using SBA. To download the file from the remote server a secret key is requested from the data owner and if data owner verifies that and then sends the keyword by which the user can login. After successful login the secret is sent to user email address and finally after entering the keyword and the secret key obtained user will be able to download the file from the remote cloud server. The backup server will send the file to both the user and the main cloud server so that next time the user can get the file from the main server when needed. The final outcome of this method I.e. downloading is shown in the figure 3 below:

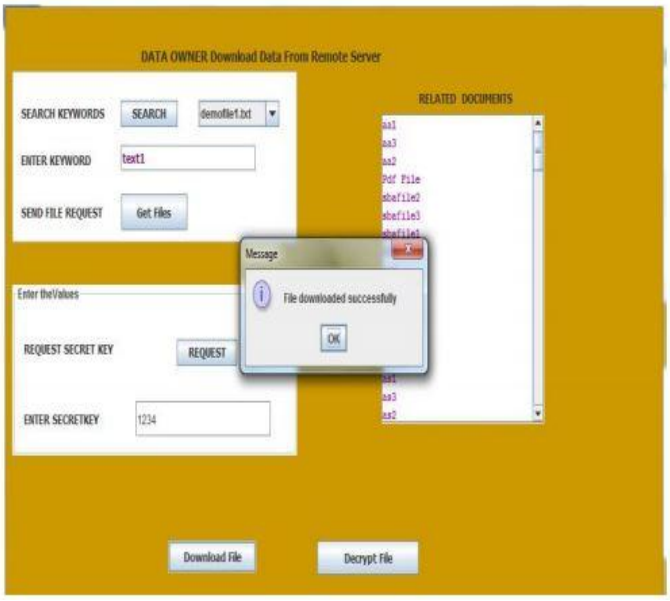

Figure 3: final download [4]

In [6] the proposed system is simulated on windows pc using java language. Different types of files of different sizes are used to conduct the practical implementation of this method. The user uploads the file to the main server and to two or four servers for backup. If file gets deleted from the main repository due to some reasons the fie can be recovered back from the backup servers. The time taken to recover the different files using this method is shown in the figure 4.

In [9] the implementation consists of login page, image upload and image selection view in the user interface. By login page the pages i.e. image upload and image selection view page will be displayed to the authorized user. The procedure of image uploading that takes place in the system will be detailed by image upload and image selection view pages. By using pictorial representation measure of the system is shown Figure 5.

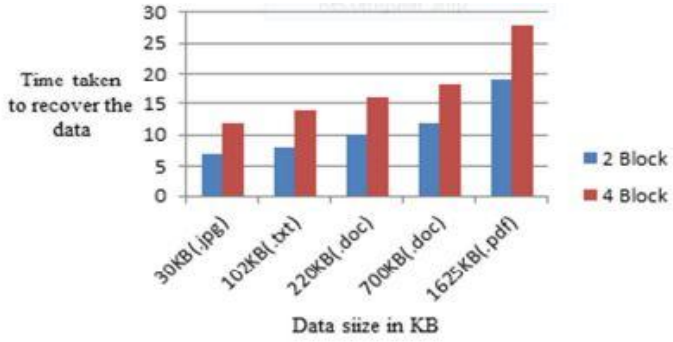

Figure 4: time taken to recover a file in seconds [6]

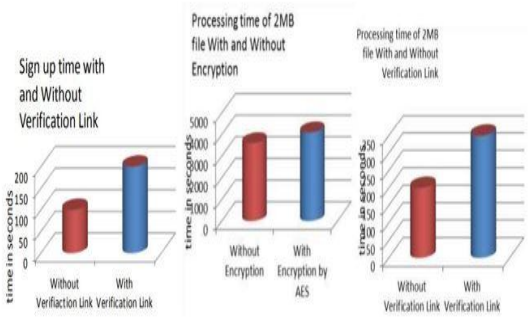

Figure 5: measure of system using graphical representation [9]

IV. CONCLUSION AND FUTURE SCOPE

Because of the regularly expanding utilization of innovation, large amount of data is created and stored. All the above strategies endeavoured to cover different issues of data backup and recovery for Cloud Computing for instance, keeping up the cost of utilization and execution complexities as low as would be reasonable. However, every single one of the backup answer for Cloud Computing can't accomplish every one of the issues of remote data move down server with less storage room. So there is a need to develop a data backup and recovery system in cloud computing that does cover the maximum trade-offs and give the best optimal results.

ACKNOWLEDGMENT

Much significant thanks and profound respect to Mr. Mir Aman Sheheryar for his commendable direction, profitable criticism and steady support all through the term of the review.

REFERENCES

[1] Shubhashis Sengupta and K.M. Annervaz "Multi-site data distribution for disaster recovery—A planning framework," *ELSEVIER*, 2014.

[2] Yashodha Sambrani and Dr. Rajashekarappa "Efficient Data Backup Mechanism for Cloud Computing," *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 5, Issue 7, 2016.

[3] Rodrigo S. Coutoa, Stefano Secci, Miguel Elias M. Campista and Luís Henrique M.K. Costa "Server placement with shared backups for disaster-resilient clouds," *ELSEVIER*, 2015.

[4] Mayuri Tidke, Vijayshree Jadhav, Sonali Parab, Shubhrata Patil and Y. K. Patil "Seed Block Algorithm: A New Approach for Data Back-up and Recovery in Cloud Computing," *International Journal of Engineering Science and Computing*, Vol. 6, no. 4, 2016.doi: 10.4010/2016.941.

[5] Shilpi U. Vishwakarma and Praveen D. Soni "Cloud Mirroring: A Technique of Data Recovery," *International Journal of Current Engineering and Technology*, Vol 5, no. 2, 2015.

[6] Praveen S. Challagidad, Ambika S. Dalawai and Mahantesh N. Birje "Efficient and Reliable Data Recovery Technique in Cloud Computing," *Internet of Things and Cloud Computing*, Vol. 5, No. 5-1, 2017, pp. 13-18. doi: 10.11648/j.iotcc.s.2017050501.13

[7] Megha Rani Raigonda1 and Tahseen Fatima2 "A Cloud Based Automatic Recovery and Backup System with Video Compression," *International Journal of Engineering and Computer Science*, Vol 5, no. 9, 2016.

[8] Chi-won Song, Sungmin Park, Dong-wook Kim, Sooyong Kang, 2011, "Parity Cloud Service: A Privacy-Protected Personal Data Recovery Service," International Joint

[9] Tanay Kulkarni, Sumit Memane, Onkar Nene and Krupali Dhaygude "INTELLIGENT CLOUD SECURITY BACK-UP SYSTEM," *International Journal of Technical Research and Applications*, Vol. 3, no. 2, 2015, PP. 241-245

[10] Yoichiro Ueno, Noriharu Miyaho, Shuichi Suzuki and Kazuo Ichihara "Performance Evaluation of a Disaster Recovery System and Practical Network Applications in Cloud Computing Envionments," *International Journal on Advances in Networks and Services*, vol 4 no 1 & 2, 2

Authors Profile

Mr. Danish Nazir pursed Bachelor of Technology from Baba Ghulam Shah Badshah University Rajouri, Jammu in 2016. He has certificate of training in ASP.net. He is presently research scholar of Master of Technology from Central University of Kashmir, India.

Mr. Mir Aman Sheheryar pursed Bachelor of Technology from Islamic University of science and Technology, J&K India in 2014 and Master of Technology with specilization in Networking from Sharda University in year 2016. He has IIRS certfication, CCNA,CCNP and MCSA certificaions.Served In Telos net Solutions as Network Admin and University of Kashmir. Currently working as Assistant Professor in Department of Information Technology, Central University of kashmir,India . He has published research papers in reputed international journals also available online. His main research work focuses on Network Security, Cloud computing, Networking,, IoT,wireless Communication,Mobile Computing and authentication. He has 1 plus year of teaching experience and 1 year of Technical Experience as Network Admin.

