International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-3, February 2020

Snapshot Based Disaster Recovery on Cloud

Vishnu A., Arokia Paul Rajan R.

Abstract: ***Nowadays, data has been produced in a larger amount. So, it is important to have some recovery issues also related to it. Cloud computing refers to the interconnection of various systems for the purpose of sharing resources. Cloud is one of the largest platforms that is growing rapidly in the IT sector. There are many private data that are related to cloud. Therefor the need for data recovery in the cloud is gaining more importance day by day. For this an efficient and effective data recovery techniques are to be developed. Recovery of data helps the users to collect information on any backup servers whenever the server is down. Many solutions have been developed for the disaster recovery. This research paper mainly discusses about some of the techniques and solutions that are needed to recover the data in cloud architectures. This work proposed snapshot based backup technique for the databases and implemented successfully.***

Keywords: Cloud computing, Disaster recovery, Data recovery, Backup, Disaster recovery planning.

I. INTRODUCTION

Cloud computing is one of the fastest growing technology in the IT sector. Cloud computing refers to the interconnection of various systems for the purpose of sharing resources. It is a technology that is internet based. Cloud computing is implemented mainly for the reason for sharing resources. With the invention of cloud computing it replaces older technologies like grid computing. There are many advantages that the clients receive after the usage of cloud computing. With the use of cloud computing the user was able to get supercomputing and high computing power for a cheaper cost. Cloud computing, mainly involves the sharing of computing resources, therefore there will be a large number of users who shares the same storage and other computing resources. As a result of this there is a strong need for a mechanism to prevent other users to access useful data either intentionally or accidentally or there may be some users who will try to modify your data. The stored data will be also affected by the natural disasters like flood, fire etc. Once it is severely affected by the natural calamities it would be very difficult to recover the data. The term cloud computing may be defined in more details as a system where we will get plenty of computer system resources without the involvement (direct) of the users [1].

Cloud computing may be defined as a data center that is available to an infinite number of users. The cloud storage may be limited to a single or multiple number of organizations depending upon the usage. The experts came into a conclusion that after the introduction of cloud, IT infrastructure cost was able to reduce into a larger amount. The advantages of cloud are not only restricted towards one field. It has got several advantages in different fields. After the introduction of cloud, the users were able to run their application faster by maintaining its manageability, so that IT teams were able to meet up their requirements more than what they expected [2].

There is an infinite amount of data is stored in the cloud. These data may be of high importance for an organization or for a company. So, the data stored in the cloud should be highly protected. When there is a sudden system crash or when power failure occurs, the data stored in the cloud will be lost. It may even cause huge financial loss for some organizations. The data can be lost due to natural disasters also. When disaster occurs, the data stored in the cloud should be protected. The trending IT companies including Google, Microsoft also have faced this type of disaster loss. When something happens to our system or in other words, when something happens in the client side, the user will not have to worry about the data that is stored in the system. Because the data will be automatically saved in the cloud. So that user can easily recover the data from the cloud. But this is not the case when something happens vice-versa. That is, when something happens to the cloud entire data will be lost. Natural disaster occurs due to bad weather, deforestation, sand mining etc. When disaster occurs nothing can be done by human to keep the disaster away. We have to face the consequences of disaster [3].

The background and related works are being discussed in Section II. Section III depicts the problem associated with the previous work and the experimental setup under which the proposed algorithm has been implemented. Section IV presents the results and a discussion of the experiment performed, Section V presents the conclusion and the future scope of the proposed technique.

II. RELATED WORKS

Kriti Sharma proposed an algorithm against the disaster which is known as Seed Block Algorithm [4]. This Algorithm is suggesting a remote backup server. It is located in the remote location. The algorithm deeply explains the solutions whenever there is a data loss. Nowadays, many inventions have made for the problem of data loss in cloud computing.

Revised Manuscript Received on February 05, 2020.
* Correspondence Author
Vishnu A., MSc Student, Department of Computer Science, CHRIST (Deemed to be University), Bangalore, India.
E-mail: vishnu.a@cs.christuniversity.in
Arokia Paul Rajan R.*, Associate Professor, Department of Computer Science, CHRIST (Deemed to be University), Bangalore, India.
E-mail: arokia.rajan@christuniversity.in

Among them most common methods are seed block algorithm, Parity cloud service, High distribution and rake technology, Shared backup router resources, efficient Routing grounded on taxonomy.

The seed block algorithm is defined as the time efficient algorithm that is used to recover the file. These algorithms consist of several advantages like it maintains data integrity, solves problems in implementation, cost and complexity. The algorithm collects information from the user and then recovers the file if the file has been deleted accidentally or intentionally. This Algorithm focuses on simplicity of the back-up and recovery. Seed block algorithm mainly consists of a remote backup server, main cloud, and its clients. It uses the concept of Exclusive– OR (XOR) operation of the computing world. If the user loses his data from the main cloud due to some reasons, the user can get back the original file by EXORing the file with the seed block of that particular file. In seed block algorithm, there will be always a random number and a unique id for the client and whenever client register in the main cloud then the random number and the client id get EXORed each other in order to produce the seed block algorithm for that particular client.

Whenever a client creates a file in the main cloud it is stored in the main cloud. Then these main files are EXORed with the seed block algorithm of that particular client. This EXORed file is then stored in the form of a file on the remote server. If due to any reason, if the file gets destroyed then the user can recover the original file by EXORing the file with seed block algorithm of the corresponding to the client. The seed block algorithm gives more importance to the security issues, so that it protects the data without using any of the existing encrypted techniques. Though it has got several advantages it also got demerits. The files that are stored in the remote server uses the same space in the cloud which leads to wastage of memory space. Thus, this method becomes inefficient. The efficiency of this method can be increased by applying compression techniques to the files.

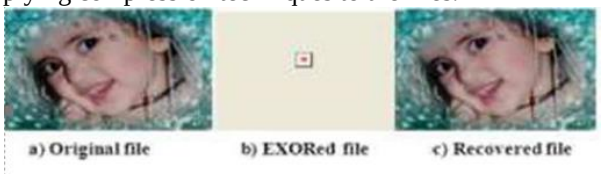

Fig 1. Sample Output Image of SBA Algorithm

Efficient Routing Grounded on Taxonomy method was proposed by Paolo Messier [5]. The method is based on semantic analysis. This method does not depend on time complexity. This model is made of 3 components:

1. (DHT) Distributed Hash Table
2. (SON) Semantic Overlay Network
3. Measure of semantic similarity

This method provides an efficient way of retrieving data that are based on semantic similarity. The semantic similarity is between service description's and service requests. This method proposes a semantic-driven query which will be answered in a DHT based system by implementing a SON over DHT. The main disadvantage of this method is increased time and implementation complexity.

Noriharu Miyaho has proposed High distribution and Rake technology a Backup model called HS-DRT that uses the concept of ultra-widely distributed data transfer mechanism [6]. This mechanism has also got high speed encryption technology. This method is having two sequences. They are Backup sequence and Recovery sequence. In Backup sequence users are storing data that are to be backed-up. Recovery sequence is used when a disaster occurs in the main server. Like any other methods this method has also got certain limitations. Due to this limitation this method cannot be considered as a perfect data recovery mechanism. Although this method can be implemented in devices like mobile phone, laptops etc., The cost of data recovery is very high and there is also chance for increased redundancy.

Cold and Hot Back-Up Strategy method was proposed by Lili Sun, Yang Yang [7]. It is a trigger-based back-up and recovery mechanism which becomes active when failure is detected. In CBSRS (Cold Backup Service Replacement Strategy) recovery mechanism the trigger becomes active when a failure is detected (Service Failure) and when there is no failure trigger becomes inactive. But in HBSRS (Hot Backup Service Replacement Strategy) the process is different. HBSRS method is also known as Transcendental Recovery Strategy, which is used in dynamic network. When the implementation of the process gets started, the backup services will remain in the activated state itself. Then the first obtained results from the service will be used to make sure the successful implementation of service composition. The main disadvantage of this mechanism is that as data increases the cost for implementation also increases.

Backup for Cloud and Disaster Recovery for Consumers and SMBs method was proposed by Vijaykumar Javaraiah for data backup in the cloud and also for disaster recovery [8]. In this method, the cost of backing up the data in cloud will be reduced and also protects the data from disaster. The process of migrating the data from one cloud to another cloud is easy in this process. Since consumers do not depend upon service providers in this method, it reduces the cost of recovering the data. The only thing that is used to recover the data in this method is a hardware box.

Rent out the Rented Resources method was proposed by Sheheryar Malik [9]. Reducing the monetary cost of service is the main aim of this model. This model has three phases as follows:

1. Discovery
2. Matchmaking
3. Authentication

This method uses the concept of cloud vendors that will rent the resources from various ventures and immediately after the virtualization it will rent these resources to clients.

The parity cloud service model proposed by Chi-Won Song [10]. This method provides high service for the personal data. This method is different from other methods, because in this method user need not to upload the data on the server. The recovery services that are provided by the client side are stored within a reasonable bound. The main disadvantage of this method is that its implementation complexity is high.

Shared Backup Router Resources method was proposed by Eleni Palkopoulou [11].

This method mainly aims at reduction of implementation cost and provides a solution when the router fails. This method consists of an IP address which will not get affected even if the router fails.

This model shows how the outer requirements have a direct impact on the SBRR architecture. The main disadvantage of this method as it reduces the implementation cost, it will be unable to optimization concept.

III. PROPOSED ALGORITHM & EXPERIMENT SETUP

Based on the extensive study of the literature, this work proposes Snapshot based Disaster Recovery technique. An application is developed using Python using Django framework. A sample database has been created using MySQL and Node js is used to deploy the application architecture. The proposed Snapshot based Disaster Recovery Technique is presented as algorithm as follows:

Step 1: select database for backup. A sample MySQL database is shown in figure 1.

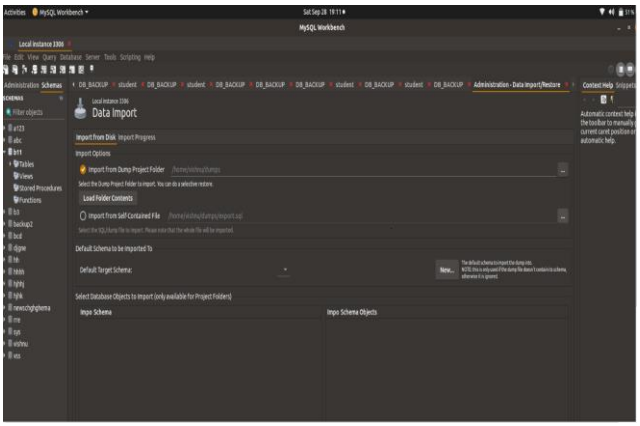

Fig. 1. Sample database

Figure 2 shows the execution of connection from the application environment.

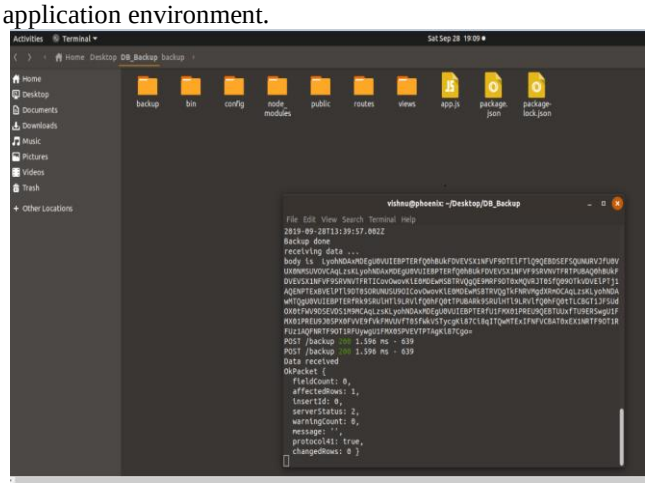

Fig. 2. Connection to the Database

Step 2: Set time interval for backup.

Step 3: Call the backup function in each interval.

Step 4: Create .sql file of database using MySQL dump in node.js

Step 5: Convert the .sql file into base64 string.

Figure 3 shows the creation of repository store for the backup snapshots.

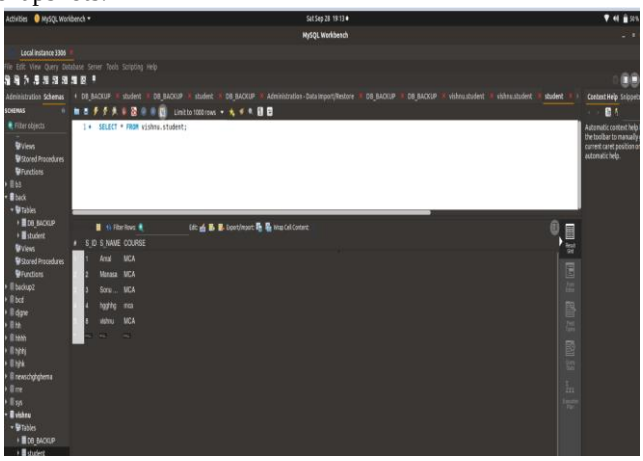

Fig. 3. Repository creation for backup files

Figure 4 shows the execution of the backup process by node js component.

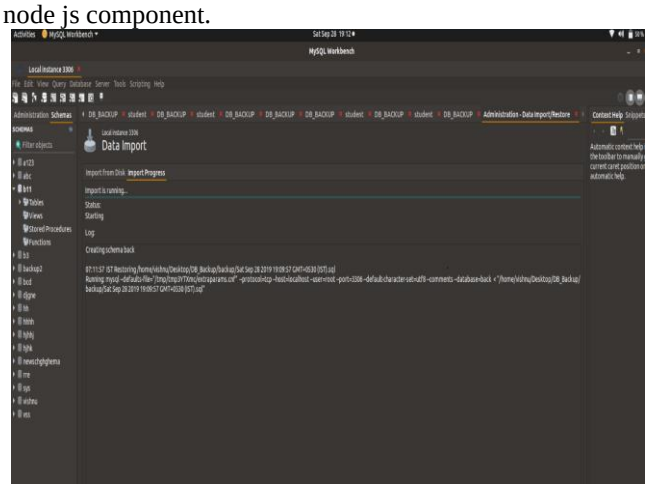

Fig. 4. Execution of backup process

Figure 5 presents the creation of a snapshot of the database in a specified time interval.

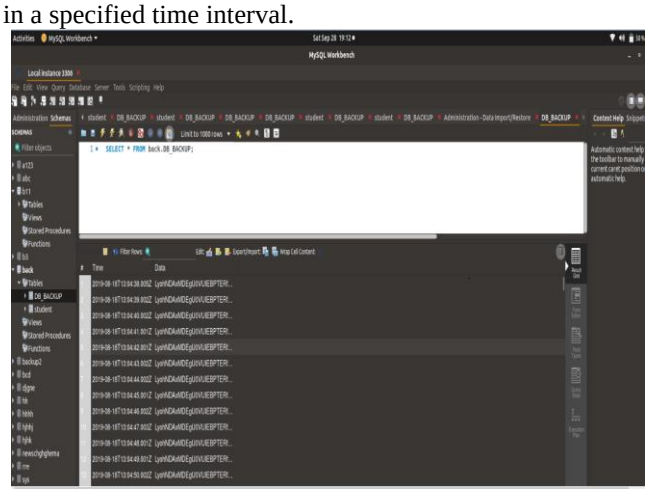

Fig. 5. Creation of snapshot of the database

Step 6: Store the base64 string into another database.

Step 7: Connect the database for the backup which is snapshotted.

Step 8: Fetch data from backup database of particular data and time.

Step 9: Convert the base64 string into .sql file.

Step 10: Snapshot of the database created.

Figure 6 presents the snapshots of the database with time intervals. Each time stamped snapshot can be a recovery backup for the application.

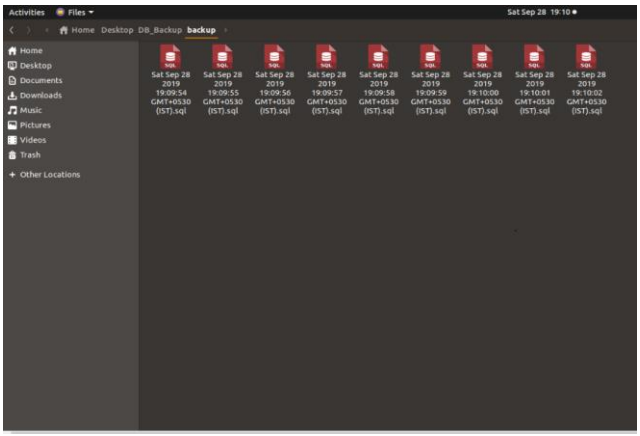

Fig. 6. Automatic Snapshot of the database

Thus the database has been successfully created and stored. The data is stored the database as .sql file at every minute in a folder. Next these .sql file is converted into base64 encrypted data and gets stored in another database in another server. Based on the requirement and application nature, it is possible to change the time interval from one minute to whatever time slice the user wants.

IV. RESULTS DISCUSSION

The implemented technique is efficient because the databases are time stamped as well as gives high importance to the data security. This method stores the database as SQL file at every minute in a folder. Next we convert the SQL file into base64 encrypted data and stored in another database on another server. Later, when we are recovering data, the data is again decrypted as normal data. An application has been created and suitable experiments conducted. The performance of the recovery system is observed with the parameter of Write Time and Read Time.

Table 1. Snapshot creation time and I/O time measurement based on Write & Read Time

Snapshot block size	Snapshot creation time (Seconds)	Write Time (ms)	Read Time (ms)
64kb	2.76	18	8
8kb	0.72	16	9.5
4kb	0.3	15	8
0.5kb	0.06	10	8

From the experimental results, it is obvious that write snapshot performs well for read-intensive workloads, while read snapshot performs well for write-intensive workloads. The proposed technique found to be efficient to handle unexpected damages to the databases, particularly for the cloud application architectures. Although the proposed method got advantages, it also has certain disadvantages [12]. In remote backup server, the data is not highly secure. By encrypting and decrypting of data we are ensuring high security of data. Resolving this issue will be an extension of this research. Fig. 7 presents time taken for creation of snapshots in remote servers in a sample experiment.

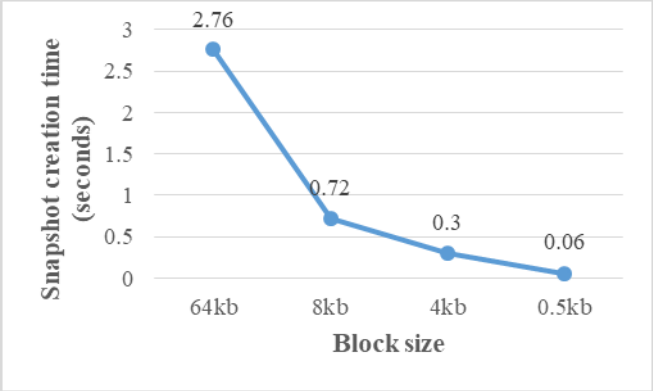

Fig. 7 The snapshot creation time of different blocks

Fig. 8 shows the results in terms of average I/O response time for different snapshots sizes.

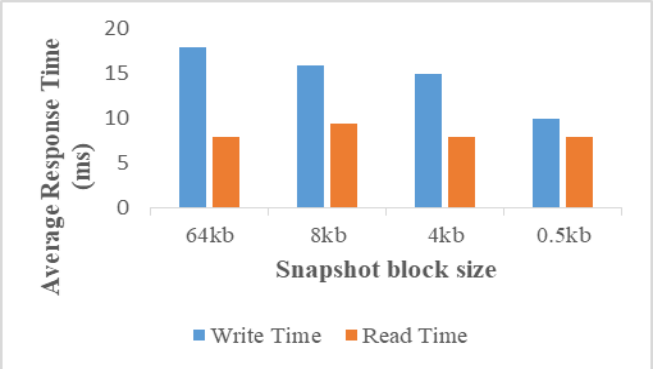

Fig. 8 Average I/O response time of Read and Write of Snapshots

V. CONCLUSION

Cloud computing is becoming very popular in our day today life and every organization are working based on cloud computing. At first, they were not aware about the disasters and also about the recovery mechanisms in cloud. When disaster occurs, the company has to face huge losses, especially financial losses because of the absence of proper recovery mechanisms.

In this paper, we proposed and implemented Snapshot based data restoration technique. Since the proposed method is based on time stamping, it is easy make recovery solutions. The method is able to recover the files that are stored in the databases within seconds. This method also focuses on the security concepts of the data while recovering. For that while recovering the data that is stored in the database, the data is first encrypted for protection. In this method user stores the database as an SQL file at every minute in a folder. After that we convert the SQL file into base64 encrypted data and it is stored in another database on another server. Later, when we recover the data it is decrypted to normal data. So, it ensures the security of the data. The experimental results of this research substantiate the storage solution designer to design economic backup solutions.

The cost of implementing this method compared to other methods is also cheap. Securing the backup servers with more security features will be the extension of research work.

REFERENCES

1. Barrie Sosinky, Cloud Computing Bible, Wiley India Pvt. Ltd, 2010.
2. S. Hamadah and D. Aqel, "A Proposed Virtual Private Cloud-Based Disaster Recovery Strategy," 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), Amman, Jordan, 2019, pp. 469-473.
3. M. M. Alshammari, A. A. Alwan, A. Nordin and I. F. Al-Shaikhli, "Disaster recovery in single-cloud and multi-cloud environments: Issues and challenges," 2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS), Salmabad, 2017, pp. 1-7.
4. K. Sharma and K. R. Singh, "Seed Block Algorithm: A Remote Smart Data Back-up Technique for Cloud Computing," 2013 International Conference on Communication Systems and Network Technologies, Gwalior, 2013, pp. 376-380.
5. G. Pirró, P. Trunfio, D. Talia, P. Missier and C. Goble, "ERGOT: A Semantic-Based System for Service Discovery in Distributed Infrastructures," 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, Melbourne, VIC, 2010, pp. 263-272.
6. Y. Ueno, N. Miyaho, S. Suzuki and K. Ichihara, "Performance Evaluation of a Disaster Recovery System and Practical Network System Applications," 2010 Fifth International Conference on Systems and Networks Communications, Nice, 2010, pp. 195-200.
7. L. Sun, J. An, Y. Yang and M. Zeng, "Recovery strategies for service composition in dynamic network," 2011 International Conference on Cloud and Service Computing, Hong Kong, 2011, pp. 60-64.
8. V. Javaraiah, "Backup for cloud and disaster recovery for consumers and SMBs," 2011 Fifth IEEE International Conference on Advanced Telecommunication Systems and Networks (ANTS), Bangalore, 2011, pp. 1-3.
9. C. Song, S. Park, D. Kim and S. Kang, "Parity Cloud Service: A Privacy-Protected Personal Data Recovery Service," 2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications, Changsha, 2011, pp. 812-817.
10. S. Malik and F. Huet, "Virtual Cloud: Rent Out the Rented Resources," 2011 International Conference for Internet Technology and Secured Transactions, Abu Dhabi, 2011, pp. 536-541.
11. E. Palkopoulou, D. A. Schupke and T. Bauschert, "Recovery Time Analysis for the Shared Backup Router Resources (SBRR) Architecture," 2011 IEEE International Conference on Communications (ICC), Kyoto, 2011, pp. 1-6.
12. O. H. Alhazmi and Y. K. Malaiya, "Evaluating disaster recovery plans using the cloud," 2013 Proceedings Annual Reliability and Maintainability Symposium (RAMS), Orlando, FL, 2013, pp. 1-6.

AUTHORS PROFILE

Vishnu A, pursuing MSc Computer Science from the Department of Computer Science, CHRIST (Deemed to be University), Bengaluru, India. His area of research interest is Cloud computing.

Dr Arokia Paul Rajan R., is currently working as Associate Professor, Department of Computer Science, CHRIST (Deemed to be University), Bangalore, India. He holds Ph.D. in Computer Science & Engineering. His research area is data management in Cloud architectures.