International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 5, Issue 03, March-2019

Survey on Data Recovery Techniques such as HS-DRT, Parity Cloud Service and Mirroring Technique for Lost Data in Cloud Computing

Miss. Ashwini T. Chauhan[1] , Prof. J.M. Patil[2]

[1] *Computer Science And Engineering , Shree Sant Gajanan Maharaj Collage Of Engineering,Shegaon,India,*
[2] *Computer Science And Engineering , Shree Sant Gajanan Maharaj Collage Of Engineering,Shegaon,India,*

Abstract- Storing the data on the cloud is one of the methods for the computation and data saving. Data has been saved by clients to the one or multiple servers via web. While saving those data over the servers or clouds there are many chances that data could get deleted or get lost. If the data is corrupted. There are many more disastrous or manmade chances for data corruption. Some of these corruptions are able to get recovered but some can't get recovered. So for theses reason there should be have some backup of lost data for clients whenever the client requires. The recovery of the data is main aim to get backup of lost data.

Keywords-Cloudcomputing ,Storage , data , loss , encryption , backup , key , distributed.

I. INTRODUCTION

A distributed cloud computing for storing the data has been serves to the client on the basis of the demands[6] .It Is serves via internet. It has been prioritize and visualized on demand .It consists of a various types of system that holds a large amount of application programs and data[1,6]. It is considered as Internet-based computing where sharing and virtualization of hardware, software and information resources are served as per on demand. Cloud computing use the internet and central remote servers to maintain data and applications and also have the ability to create, update and store files via any computer that has access to the web[6].

II. RELATED WORK

Chi-won Song proposed the innovative file back-up concept HS-DRT, that makes use of an effective ultra-widely distributed data transfer mechanism and a high-speed encryption technology[3,1]. This system has been consists of two sequences such one is Backup sequence and other is Recovery sequence. The data to be backed-up will get received In Backup sequence. The recovery sequence is used when there is a disaster or any data loss occurs the Supervisory starts the recovery sequence[3].

S.S.Ganorkar have proposed a novel data recovery service framework for cloud infrastructure, the Parity Cloud Service i.e. PCS provides a privacy protected clients personal data for recovery service. In this proposed framework user data is not required to be uploaded on to the server for data recovery. All the necessary server-side resources for that data has been provide the recovery services are within a reasonable bound[2].

TABLE 1
ADVANTAGES AND DISADVANTAGES OF DIFFERENT APPROACHES

Sr. No.	Approaches	Advantages	Disadvantages
1	HSDRT (Chi-won Song)	Used for movable clients	Costly, Increased redundancy
2	Parity Cloud Service (S.S.Ganorkar)	Reliable privacy low cost	High Complexity
3	SBA (S Sankareswari)	Simple to implement	Inefficient

III. STORAGE OF DATA IN CLOUD

Cloud storage is amorphous today, with neither clearly defined set of capabilities nor any single architecture[6]. Storing the data is primary use in clouds**.** Cloud has been serve with third-party server rather than the dedicated servers**.** while data get stored**,** user will store the data on virtual server which exist in imaginary but not in reality that's where data get stored only one data server is needed to be connected to internet. A client sends copies of to the Internet to the data server, after that the records are save the information[6]. When the client wishes to retrieve the information, he Web-based interface is useful to share the data. The server will send that data back to the client.

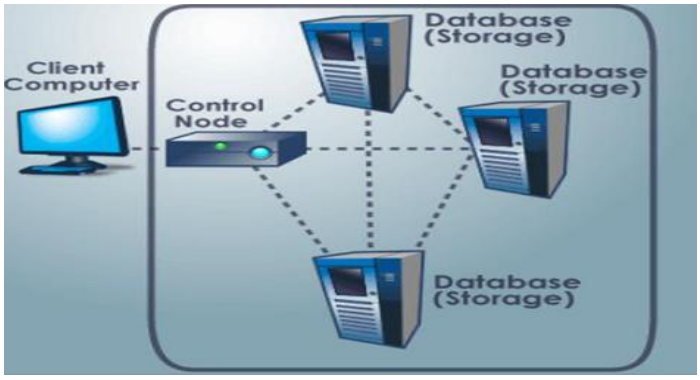

Fig 1:Cloud Storage Architecture

IV. DATA CURROPUTION

Cloud computing is a technology where in IT infrastructure has been given as various 'services' to various end-users under based on usage payment models[1]. It permits the user to store huge amount of data in cloud storage and can use whenever required, anywhere from any part of the world, via any terminal equipment .But it is important to understand some of the factors like causes of data corruption, how much responsibility a cloud service provider holds, some best practices for utilizing cloud storage safely, and some best methods and standards for providing the integrity of data regardless of whether that data resides locally or in the cloud[1,6]. Integrity checking is essential in cloud storage as providing integrity is critical for any data centre in the servers. Data corruption can happen at any level of storage and with any type of media rather than data like images or videos[1].

.Example of different media types causing corruption.

- Bit rot controller failures,
- Reduplication metadata corruption
- Tape failures

Metadata corruption can be the result of any of the vulnerabilities or viruses listed above such as bit rot means slow corruption, but are also susceptible to software glitches outside of hardware error rates Unfortunately, a side effect of reduplication is that a corrupted file, block, or byte affects every associated piece of data tied to that metadata[3]. The truth is that data corruption can happen anywhere within a storage environment[1]. Cloud storage systems are still data centres, with hardware and software, and are still vulnerable to data corruption. Recently, Amazon failure is one of the famous examples of data corruption. It has caused prolonged downtime and also 0.07 percent of their customers actually lost their data. It responses as actual data loss due to ESB.

V. DATA LOSS

Data loss is so much common it could be viral attack, natural or manmade disaster which more often get solved by senior technicians[5]**.** Only in the most severe cases of platter damage, magnetic degradation or a file over-write will the data be labelled as unrecoverable and main issue for data loss .The data loss scenarios caused the damage to servers, the Possibilities are as followed[5]:

- Hardware failure
- Human error
- Software corruption
- Theft
- Hardware Destruction

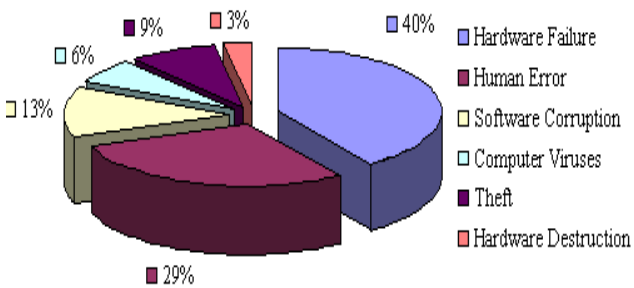

Fig 2: Data loss scenarios

VI. DATA RECOVERY TECHNIQUES

A. High Security Distribution And Rake Technology[HS-DRT]

The HS-DRT is an improved file recovery or backed up service, in which it makes use of an effective ultra widely distributed transfer of data mechanism and a technology of encryption[3].

Three Components are given:

a) main functions are Data Centre

b) Supervisory server

c) various client nodes specified by admin.

Two sequences are driven backup and Recovery sequence.

In Backup sequence, when the Data Centre will get the data for recovery, it will encrypts scramble, after encryption scrambles will get divides into some fragmentations, after that there after duplicates that data up to satisfy with the required recovery rate according to the pre-determined service level[3]. The Data Centre will again encrypt the fragmentations and also at the second stage then distributes them to the client nodes in a random order. While this process is running, the Data Centre sends the metadata used for deciphering the series of fragments. The metadata which are composed of encryption keys (in both stages), several related information of fragmentation, duplication and distribution.

In Recovery Sequence, it is the recovery process for data when some disasters occur periodically or unpatriotically, the Supervisory Server will get starts the recovery sequence[3]. It is going to collect all the encrypted fragmentations from various users or clients like rake-reception procedure and then they are decrypted, merged, and descrambled in the reverse order at the second stage of recovery and the decryption will be completed for recovery. Though these processes, the Supervisory Server can recover all the original lost data that should be backed-up[3].

B. Parity Cloud Service Technique[PCS]

Parity Cloud Service technique (PCS) is a very simple and easy technique for the recovery process which is based on parity recovery[2]. A PCS can recover data having very high probability. For data recovery, PCS uses a new technique by creating virtual disk in user's systems for backup the data, It will make parity groups across the virtual disks, and store parity data of the specific parity group in clouds. The algorithms for PCS by using the Exclusive–OR ()[XOR] for creating Parity information[2].

1) The First Step Is Parity Block Generation

In this, the seed blocks (Sb) are get generated for specific virtual disks. PCS server sends the initialize message to each Recovery Manager in the group of blocks. After sending the it will initialize message, firstly the server sends temporary random block (rb) to the first node. On receiving the random block, the first node

(node 1) generates an intermediate parity block via rb S1 and sent it to its successor, node 2. Accordingly, node 2 generates an intermediate parity block via XORing the received parity block with its seed block, S2, and sends it to its successor, node 3 and so on. The final block transferred respectively to the PCS's server from node 4 is going to XORed with the temporary random block rb again to generate the seed parity block across all seed blocks (((((r S1)S2) S3) S4) r = S1S2 S3S4). The initialization process occurs only once for each parity group. The seed parity block stored separately in the metadata region of each virtual disk, for further use[3,2].

2) The Second Step Is Parity Block Update

The Storage Manager in PCS agent maintains parity generation bitmap (PG-bitmap) .It indicates whether the parity block for each data block in the virtual disk has been generated or not. The bitmap is get initialized and set it to 0 after the initialization process for any data block in the virtual disk. The PG-bitmap is referred when a block is considered as updated. When a block (Bold) in nodei is to be updated to a new block (Bnew), the Storage Manager refers to the corresponding value in the PG- bitmap. If it is set to 0, then the Storage Manager generates an intermediate parity block (Pt) by XORing the new block with the seed block(sb) (Pt = Bnew sb), and now setting the corresponding value in the PG-bitmap to 1. Otherwise, the intermediate parity block is generated by XORing the new block and the old block (Pt = Bnew _ Bold). For each Virtual disk parity generation, the PCS server also maintains the PG-bitmap. Note that the parity block update can be easily done by the data updating node and the PCS server for all the blocks updating[2].

3) Data Block Recovery

When a data block is get corrupted due to disastrous scenarios or manmade scenarios, it can be recovered using the parity block which has been provided by the PCS server and encoded data blocks provided by other nodes in the parity group in servers[2]. Assume that the n-th data block in node i Bin, has been corrupted. Node i sends a recovery request message to the PCS server. On receiving the recovery request message, the PCS server identifies to which Virtual Disk Parity Generation's the node belongs to and reads the corresponding parity block, Pbn. Then, it is going to generates a temporary

random block rb and a temporary parity block, Pbr, for recovery process. When the size of the VDPG is even, Pbr = Pbn rb. Otherwise, Pbr = Pbn. The PCS server sends Pbr along with the list of nodes that will send their encoded data block to node i for recovery along with the IP address of node i to all other nodes in the group. If there are any off-line nodes, the PCS server sends the message when they become on-line. On receiving the message, each nodes will generates their own encoded data blocks E after the XORing the n-the data block with rb (Ej = Bin rb, for each node j VDPG, j i) and sends to node i. Then, the nodei recovers the corrupted data block[2].

VII. PROPOSED WORK

A. Mirroring Technique

The mirroring is given for maintaining real time copies known as mirror and replication. . Both mirroring and replication use the same terminology for the roles of databases: the original, updateable database is called the master. From one master database, one or more slave copies can be created and dynamically maintained. The terminology comes from the idea that the

1. Master database controls the generation of data, and
2. The slaves respond only when changes have been made on the master.

Server instances are proposed as two copies of single database resides on different computers.

The primary server instance provides the database to clients. The mirror server instance acts as a standby that can take over in case of a problem with the principal server instance[5].

The time mirroring happen the original data is get updated. When at the same time mirrored data and updated data is get updated then the term is known as synchronous operation or hot standby mirror. And the term where only original data is get updated than mirrored data then it is known as asynchronous operation or warm standby mirror [4].

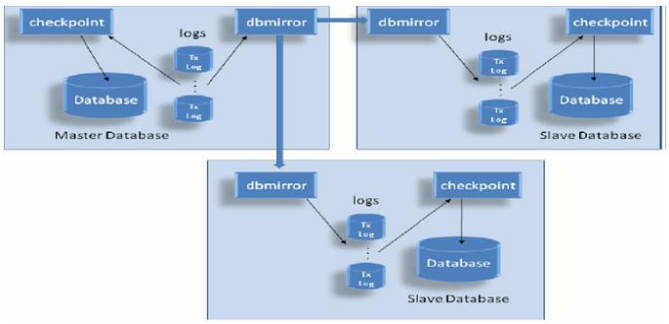

Fig 3: Mirroring Architecture

The procedure of database mirroring process as shown above:

1. Every transaction is committed to a master database by creating a log file
(Containing modified page images) in the database's directory. The log files are named after the transaction number they represent. Transactions are numbered serially with no gaps.

2. The checkpoint process scans the directory for log files. When one or more new log files are found, they are "checkpointed" into the database. The entire process is safe and repeatable so that there will be no loss of data.

3. To facilitate mirroring, the checkpoint process will not delete used log files, but will rename them so that they can be found by the dbmirror process.

4. The slave database, dbmirror process requests the "next" log from the master dbmirror process if a TFS is running, for normal transaction processing.

4.1.When it receives it, it sends it to the local TFS, If a TFS is not running together with the slave dbmirror, there may or may not be a check pointing process running.

4.2.If there is, the presence of the log causes the check pointer to copy these changes into the slave database. This is repeated forever, or until the dbmirror slave process is terminated.

5. The master data base mirror i.e.dbmirror process searches the database directory for log files and responds to slave dbmirrors when certain log files are requested.

6. The master dbmirror can respond to any number of slave dbmirrors

B. Mirroring algorithm

The proposed technique is broadly categories into two parts i.e. uploading and downloading. In the first part, user or clients data has been consist of files, documents etc. can be uploaded by the user on cloud whether it is in plain text or in encrypted format or in any other form[5].

1) Upload Module

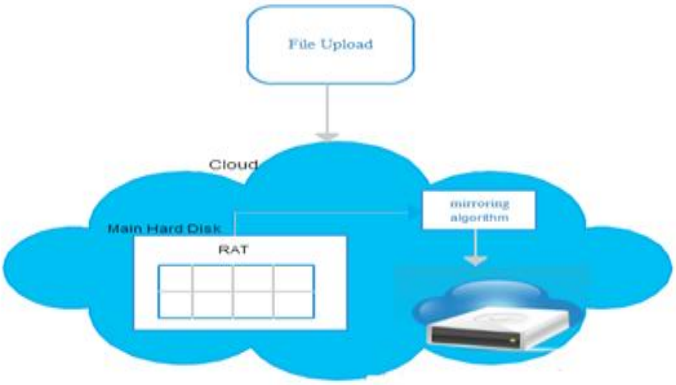

Fig 4: Uploading Module

Mirroring scheduling algorithm will check the mirror copies of the user's or client's data. Mirroring is going to starts when the CPU utilization goes below the threshold value (we assume the CPU threshold value is 50% or it could be any variable), and daily we will do the mirroring according to time (we assume time threshold is midnight (2 a.m.))[6].
By using the concept of CSP (Cloud Service Provider) we will maintain the log through which we will continuously(say after 5 minutes) check the row mirror counter for the mirroring process so that reach row will get mirrored simultaneously, after analyzing the log, CSP can dynamically change the threshold values[6].
Mirroring algorithm is as follows:-

- Notations

CPU_Threshold --- CPU Threshold
Time_Threshold --- Time Threshold
Event_Threshold --- Event Threshold
Current_CPU --- Current CPU
Current_Time --- Current Time
MHD --- Main Hard Disk

Pseudo code for mirroring :

```
No_of_rows_mirrored= 0;
If (Current_Cpu< Cpu_Threshold)
While(RAT.length!empty||Current_Cpu<cpu_Threshold)
{
Mirror the current row of RAT and delete it.
No_of_rows_mirrored++
}
If(Current_Time= Time_Threshold)
while(RAT != Empty)
{
Mirror the current row of RAT and delete it.
No_of_rows_mirrored++;
}
Return No_of_rows_mirrored;
End of Pseudo code;
```

2) Downloading module

The main aim of this proposed technique is to provide the recovery of user data (files) though it has been corrupted or loss etc. so the main role of mirroring technique' comes in downloading part, when user wants to download his requested data from the base cloud and if unfortunately the original data of user are gets corrupted in the base cloud where data is stored, then with the help of mirroring technique we will provide the same data stored by the user from mirror cloud[5].
Steps for downloading are as follows:-
1. Firstly the user will request for the user's respective data (file).
2. The user's request will moves to the retrieval algorithm where we check the presence of desired data (files), as well as its integrity.

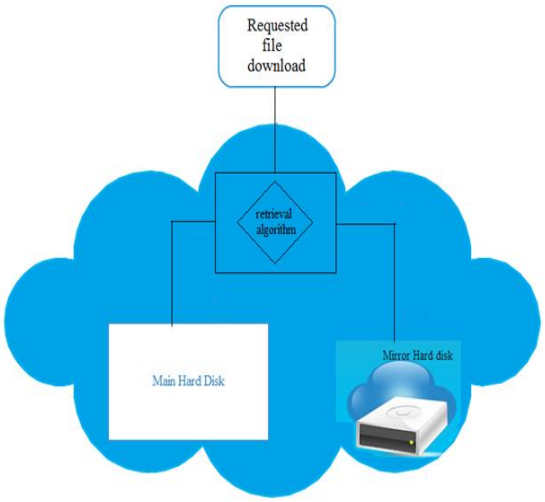

Fig 5: Download Module

File retrieval algorithm are as follows: -

1. When the request will arrive firstly we will check whether the requested data (files) is present in the MHD or not.
2. But if the file is corrupted and its integrity is improper then we will retrieve the respective file from mirror hard dis

VIII. CONCLUSION

The data stored by the user is always valuable for him in the case of needed or not, but no one can assure whether his data cannot be corrupted or lost so recovery plays a vital role in such scenarios so after this client need a backup of data. Various techniques have been proposed for data recovery but these techniques have certain limitations which need to be overcome. With the help of HS-DRT and parity cloud service and proposed mirroring technique we provide the high availability, integrity as well as recovery of user data (files). So for this issue we need file recovery mechanism for recovering the corrupted file. We have proposed file recovery technique by the concept of cloud mirroring.. This technique will focus on entire data recovery in cloud.

ACKNOWLEDGEMENT

I would love to acknowledge and thank to respected guide Prof.J.M.Patil, Professor in Computer Science And Engineering department, who has been a constant source of support and recommendation developing the background information and providing review comments and suggestions for this document.

REFFERENCES

[1] 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.

[2] S.S.Ganorkar, S.U.Vishwakarma, S.D.Pande, *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, issue1.

[3] Chi-won Song, Sooyong Kang,(march 2013) *Parity Cloud Service:- A Privacy-Protected Personal Data Recovery Service*, International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11.

[4] Kruti Sharma, 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.

[5] Vijaykumar Javaraiah,(jan2013*) Brocade Advanced Networks and Telecommunication systems (ANTS), Backup for cloud and Disaster Recovery for Consumers and SMBs*, IEEE 5th International Conference.

[6] Cloud Computing Journal
http://cloudcomputing.sys-con.com/node/640237.com