International Journal of Advanced Technology in Engineering and Science
Vol. No. 09, Issue No. 08, August 2021
www.ijates.com
ijates
ISSN 2348 - 7550

Online Data Back-up and Disaster Recovery Techniques:

[1]Vipul Verma, [2]Parth, [3]Nitin Kumar , [4]Nikhil Tyagi

[1]Assistant Professor, CSE Department,

IIMT College of Engineering, Greater Noida(U.P)

[2]U.G Student, B.Tech CSE 4TH Year parthd3040@gmail.com

[3]U.G Student , B.Tech CSE 4TH Year,

IIMT College of Engineering , Greater Noida(U.P)

[4]U.G Student , B.Tech CSE 4TH Year,

IIMT College of Engineering, Greater Noida(U.P)
nitinkumar743@gmail.com1nikhil.iimt@gmail.com

Summary - Today, in the electronic format it is produced in large quantities that require data retrieval services. We know that cloud computing is introducing a new type of computer platform in today's world. This type of computer will generate a large amount of private data in a large cloud. Therefore the demand for data acquisition services is growing day by day and requires the development of an efficient and effective data acquisition method. The purpose of the recovery method is to help the user collect data from any backup server where the server has lost its data and is unable to provide data to the user. To this end, many different strategies have been proposed so far. In this review paper, we explore a few of the latest powerful solutions in the form of "Online Data Backup and Disaster Recovery Techniques". The purpose of this review paper is to summarize the powerful data storage methods used in the cloud computing domain.Reference Terms - Backup; Privacy; Central Storage; Remote Location, Parity Cloud, Parity Cloud Service.

I. INTRODUCTION

The National Institute of Standard and Technology defines it as an example of allowing easy network access, and demands in the pool of configurable computer service sharing (e.g., networks, servers, storage, applications and services) that can be quickly and efficiently managed or managed by a service provider [1]. Cloud computing is no longer the buzzword today. In addition, it changes and improves the way we use a computer platform.

In today's world, there is a huge increase in electronic data. This requires a large volume of data storage devices to store this large data. This requirement leads to the launch of 3 Tera Byte HDD. Therefore, usually the consumer likes to keep a large amount of confidential information in the cloud. Unfortunately, if a cloud is to be damaged or damaged it leads to the loss of all important and confidential information then there must be some way to restore the data, and provide information in case of cloud failure or data loss.

Knowing that open data backup strategies have a lot of reliability and security issues. However, back-to-back strategies are inaccurate and unreliable as well. Therefore, in order to overcome the problem

of data storage and recovery, it requires a secure and efficient system such as RAID (Redundant Array Independent Disk), HSDRT [1], PCS [2], ERGOT [3], Linux Box [5], Cold and Hot backup process [6], SBBR [10], REN [17] etc. These programs provide high level of privacy and security but some increase costs while others cannot maintain use

low weight. Although many backup and restore

proposed strategies over the past few years in the computer field; however, real world conditions remain a challenge. In this review paper, we focus on various ways to backup and recover computer cloud computing. Each onethe process is most affected in a real-time situation either in the point of view of the need or the point of view or the point of view of a complex algorithm.

This paper is structured as follows: Phase II describes the need for cloud computing. In Section III we discuss Remote Data Backup Server. The existing methods that are somewhat effective in the cloud computing environment are reviewed in section IV. Finally, in paragraph V the conclusions are discussed.

II.You NEED TO HAVE A BACK-UP IN CLOUD COMPUTING

Cloud computing provides the required resources for the buyer / user. Requires per client / user resource management management. Such management includes various aspects of the proper use of resources. Resources can be hardware or software. Software like any application interface, program development kit and any type of data file etc. Different options exist between different implementations for backing up data and maintaining its security for different users. Cloud computing should be able to provide reliability such as that users can upload their sensitive and important data. The least expensive method is the main concern over time

to use any cloud.

During the study of cloud computing, we discovered various benefits of using cloud computing. Advantages, we have found that the cloud is capable of storing large amounts of different customer information with complete security such as that the Internet Service Provider (ISP) provides greater cloud storage to the user. Also users are allowed to upload private and important data to a large cloud. And at the same time we have a serious problem with this storage which means that if there is a client data file that is lost or disappears due to some reason or cloud is destroyed due to any natural disaster (such as flood, earthquake, etc.), back-up and retrieval buyer / client must rely on the provider service provider which means data must be stored on a server.

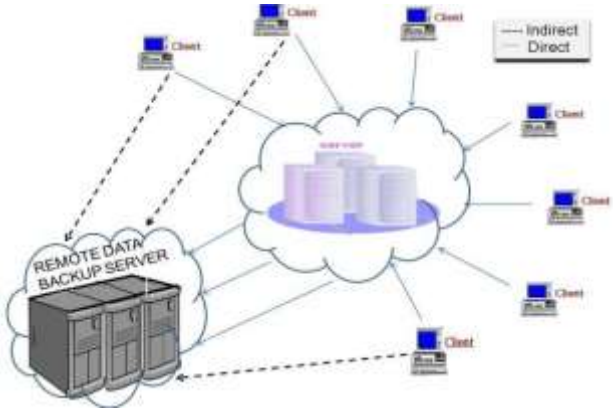
To overcome this problem, you need to

an effective process for backing up and retrieving data so that the client can communicate with the backup server when private data is stored with high reliability and whenever a large cloud fails to provide user information. These methods should have low cost and implementation of a solution to the recovery problem and can easily retrieve data after any disaster. Hence, the need for cloud recovery and renewable cloud computing strategies arises from the massive storage of its customers.

II. FAR DISTRICT SUPPORT

Remote Data Backup Server is a server that keeps all cloud data complete and located remotely (away from the cloud). And if the central location loses its data, then it uses data from the remote location. The purpose is to help clients collect data from a remote location even if the network connection is

not available or the primary cloud cannot provide information to customers. As shown in Figure 1, if clients



found that the data is not available in the central repository, then clients are allowed access to files from the remote location (e.g. indirectly).

Fig. 1. Remote control data and its configuration

Remote backup services should cover the following issues:

1) Privacy and ownership.

2) Server migration to the cloud.

3) Data security.

4) Honesty.

5) Cost efficiency.

6) The Right Time.

1) Privacy and ownership

Various customers access the cloud through a separate login or after any authentication process. They are free to upload their own private and important data to the cloud. Therefore, the privacy and identity of the data should be maintained; The data owner should only be able to access his personal data and perform read, write or other operations. The Remote Server must maintain this Privacy and Identity.

2) Server migrationData recovery should have the server moving to the cloud. Server migration means transferring big data data to another server; however, the new location is unknown to the client. Clients receive information in the same way as before without having access to the primary server delivery, such as providing visibility of the server location transferred to clients and third parties while the data is being transferred to a remote server.

3) Data security

Client data is stored in a central repository for complete protection. Such security should be maintained even in its remote location. In a remote cache, data must be completely protected from access and damage to remote data data either intentionally or unintentionally by a third party or other client.

4) Honesty

The remote cloud must have reliability features. Because in a computer cloud deployment the core cloud stores complete data and each client relies on each master cloud cloud; therefore cloud and remote support cloud must play a reliable role. That means, both servers should be able to provide client information as soon as they need it from a large cloud or remote server.

5) Cost efficiency

The cost of remote server startup and its recovery and backup process also played an important role in the construction of the main cloud structure and its associated remote cloud. The cost of establishing a remote control setup and implementing its own process should be so small that small businesses can afford such a system and a large business can use as low a cost as possible.

6) The Right Time The data acquisition process takes a while to retrieve data from a remote location as this remote location is far away from the big cloud and its customers. Therefore, the time taken for such retrieval should be as short as possible so that the client can access the data as quickly as possible without regard to how far away the client is.

There are many ways to focus on these problems. In the next section, we will discuss some of the latest techniques for backing up and restoring cloud computing.

AVAILABLE WAYS

In our literature review, we have found many techniques that have their own unique ways of backup and recovery. Overall, all of these strategies focus on three different aspects, such as cost control, data duplication and security issues. Each of these approaches is totally focused on their goal of backup and recovery. In addition, we describe a few recent HSDRT techniques [1], PCS [2], ERGOT [3], Linux Box [5], Cold and Hot backup process [6], SBBR [10], REN [17] addressed the issues mentioned above.

Advanced Technology Distribution and Rake Technology (HS-DRT)

HS-DRT [1] is an innovative file support concept, which uses a highly distributed data distribution method and high-speed encryption technology. node various clients specified by the administrator.

Client nodes are made up of PCs, smartphones, Network Storage and Storageservice [1]. They are connected to a management server over the Data Center via a secure network.

The basic process in the proposed network system consists of two consecutive Backup sequences and the second Recovery sequence. In Backup sequence, when the Data Center finds the data to be backed up, writes the text, breaks it into some cracks, and then repeats that data to another satisfaction with the required recovery level according to the pre-determined service level. The Data Center encrypts the division of the second phase and distributes it to client nodes in random order. At the same time, the Data Center sends metadata used to specify fragments of fragments. Metadata is composed of encryption keys (both in the first and second phases), several related details of splitting, duplication and distribution [1]. In the Recovery sequence, which is a recovery process in the event of certain disasters or periodically, the Service Supervisory Server initiates a recovery sequence. It collects confidential segregation from various relevant clients as a rake-receiving process and is removed, merged, and descended in a recurring sequence in the second phase and encrypted. Despite these processes, the Supervisory Server may retrieve the original data that must be backed up.

However there are limitations to this model so, this model somehow cannot be declared as the ideal solution for backup and recovery. Here it is: First, in order to make full use of the HS-DRT processor, web applications are required to be properly configured to use the HS-DRT engine. Second, that when the duplicate copy of the file data increases the performance of the compatible processor will be reduced accordingly by making a web application.

Parity Cloud Service Technique

The Parity Cloud Service (PCS) process [2] is very simple, easy to use and very easy to obtain data based on the integration service. PCS has low acquisition costs and can access data at very high probability. To recover data, PCS implements a new virtual disk drive process in the user's system to back up data, create clusters on a virtual disk, and store cluster unity data in the cloud. PCS algorithms work this way using Exclusive - OR () to create Parity data.

1) First Generation of Parity:

In this case, a block (Si) is created for the visible disk. The PCS server sends a launch message to each Recovery Manager in the group. After sending the start-up message, the PCS server sends a temporary temporary block (r) to the original location. When you find ir block, the first node (node 1) generates an equal medium block with r S1 and sends it to its successor, node 2. Similarly, node 2 creates an intermediate block with XORing the block block detected its block, S2, and send it to its follower, node 3 and so on. The last block transferred to the PCS server from node 4 XORed with a random temporary block, r, and, to produce a seed block of parallel across seed in blocks ((((((r OR

S1) OR S2 OR S3) S4) OR r = S1 OR S2 OR S3 S4) The startup process occurs only once in each group.The seed block is stored separately in the metadata region of each visible disk, for later use.

2) Parity Block Renewal:

Storage Manager for the PCS agent stores a single-generation bitmap (PG-bitmap) that indicates whether the equalized block of each disk data generated has been created or not. The bitmap is started (set to 0) after the startup process of any data block on the visible disk. PG-bitmap is transmitted when the block is updated. When the block (Bold) in nodei will be updated to the new block (Bnew), Storage Manager refers the corresponding value to the PG-bitmap. If it is 0, then Storage Manager creates an internal block (Pt) by XORing a new block with a block (Pt = Bnew Si) and sets the corresponding value in PG-bitmap to 1. Alternatively, the middle block is made with XORing new block and old block (Pt = Bnew _ Bold). For each VDPG, the PCS server also stores the PG-bitmap. Note that parity block updates can be easily performed with PCS data recovery process. In the update process, some nodes are not required to participate.

3) Data Recovery:When a data block is damaged, it can be retrieved using the unity block provided by the PCS server and the embedded data blocks provided by other nodes in the group. Assume that the n-th data block in nodei, Bin, is corrupted. Nodei sends a recovery request message to the PCS server. When you receive the recovery request message, the PCS server identifies which VDPG is the first and reads the equivalent block, Pn. Thereafter, it produces a random temporary block, r, and a temporary block parity, Pr, of the recovery process. When the size of the VDPG is equal to, Pr = Pn

r. Otherwise, Pr = Pn. The PCS server sends Pr and a list of nodes that will send its encoded data block to nodei retrieval and nodei IP address to all other groups in the group. If there are any offline areas, the PCS server sends the message when it is online. In receiving a message, each node creates its own data block, E, by inserting the n-block block with r (Ej = Bin, with node j VD PG, i) and sending it to the nodei. After that, the node returns the damaged data block by

Bin = Pr E1 · No-Ei-1 Ei + 1 · E E | VDPG |. (1)

Note that all visible disk corruption can be detected by installing the above data acquisition process. Apart from its excellent performance provided by the algorithm discussed above PCS has somehow lagged behind in providing complete backup and recovery solutions due to certain liabilitiesaccording to the group size. Generally, it fails to precisely estimate the real data reliability of PCS.

Efficient Routing Grounded on Taxonomy (ERGOT)

Efficient Routing Grounded on Taxonomy [4] is a Semantic-based System for Service Discovery in Distributed Infrastructures in cloud computing. In our survey, we found a unique way of data retrieval. We made a focus on this technique as it is not a back-up technique but it provide an efficient retrieval of data that is completely based on the semantic similarity between service descriptions and service requests. It also exploits both coarse-grain service functionality descriptions and at a finer level.

ERGOT is built upon 3 components. These components include: 1) A DHT (Distributed Hash Table) protocol, which we use to advertise semantic service description annotated using concepts from ontology, 2) A SON (Semantic Overlay Network), enables the clustering of peer that have semantically similar service description. The SON is constructed incrementally, as a product of service advertising via DHT, 3) A measure of semantic similarity among service description [4].

DHTs and SONs both networks architectures have some shortcomings. Hence, ERGOT combines both these network Concept. The ERGOT system proposed semantic-driven query answering in DHT-based systems by building a SON over a DHT. An extensive evaluation of the system in different network scenarios demonstrated its efficiency both in terms of accuracy of search and network traffic. DHT-based systems perform exact-match searches with logarithmic performance bounds, however does not go well with semantic similarity search models.

A. Linux Box

Another technique to reduces the cost of the solution and protect data from disaster. It also makes the process of migration from one cloud service provider to other very easy. It is affordable to all consumers and Small and Medium Business (SMB). This solution eliminates consumer's dependency on the ISP 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 incorporates 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 we found 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.

B. Cold and Hot Backup Service Replacement Strategy (CBSRS)

In Cold Backup Service Replacement Strategy (CBSRS) recovery process, it is triggered upon the detection of the service failures and it will not be triggered when the service is available. In Hot Backup Service Replacement Strategy (HBSRS), a transcendental recovery strategy for service composition in dynamic network is applied [6]. According to the availability and the current state of service composition before the services interrupt, it restores the service composition dynamically. During the implementation of service, the backup services always remain in the activated states, and then the first returned results of services will be adopted to ensure the successful implementation of service composition. On Comparing HBSRS with the CBSRS, it reduced service recovery time. However, because backup services and original services are executed at the same time, the recovery cost increases accordingly.

C. Shared backup router resources (SBBR)

In one of our survey, we found that one technique basically focuses on the significant cost reduction and router failure scenario i.e. (SBBR). It concerns IP logical connectivity that remains unchanged even after a router failure and the most important factor it provides the network management system via multi-layer signaling. However, it concerns with the cost reduction concept there exist some inconsistencies between logical and physical configurations that may lead to some performance problem. Additionally [10], it shows how service imposed maximum outage requirements that have a direct effect on the setting of the SBRR architecture (e.g. imposing a minimum number of network-wide shared router resources locations). However, it is unable to includes optimization concept with cost reduction

D. Rent Out the Rented Resources

Another technique we found in the field of the data backup is a REN (Research Education Network) cloud. As we know the Cloud services are expensive and large number of enterprises and individuals are attracted towards low cost cloud services. The lowest cost point of view we found a model "Rent out the Rented Resources" [17]. It aims to reduce the monetary cost of cloud services. They have proposed a three phase model for cross cloud federation. These three phases are discovery, matchmaking and authentication. Kealey et. al. introduced the concept of Sky Computing [15]. This model is based on concept of cloud vendor that rent the resources from venture(s) and after virtualization, rents it to the clients in form of cloud services. The cooperating venture is paid for its infrastructure utilization [17]. It is based on three core objectives: 1) It minimizes the cloud infrastructure cost. 2) It provides low cost cloud services by reducing infrastructure cost for the cloud vendors to the clients. 3) It gives the monetary benefit with the large under-utilized technology infrastructure to the established enterprises (cooperating ventures).

DISCUSSION AND CONCLUSION

In this paper, we present the details of the latest back-up and redesign strategies that have been developed in the cloud computing platform. A review of the details of this paper shows that these processes have their advantages and disadvantages summarized in Table 1. All of these methods are able to provide excellent play under all uncontrolled conditions such as cost, security, low start-up difficulty, retrenchment and short-term recovery.

Among all the revised strategies PCS are relatively reliable; maintains its privacy on each device and attempts to reduce infrastructure costs. However, it cannot control the startup difficulty. On the contrary, HSDRT has come out with practical solutions for mobile customers such as laptops, smartphones etc. Instead, ERGOT is based entirely on semantic analysis and cannot focus on time and complexity. In addition, the Linux Box model has a very simple data concept

Although each backup solution in cloud computing can meet all of the above-mentioned issues backup and recovery at very low cost. However, in this case of the model protection model is very low. Similarly, in a list of cost-saving strategies, SBBR focuses on cost reduction; but fails to focus on the concept of using and improving the function. With a completely new concept of virtualization REN cloud also focuses on low cost infrastructure with sophisticated implementation and low level of security. All of these strategies have tried to cover a variety of issues that keep startup costs as low as possible. However, there are other strategies where costs increase slightly as data increases. For example, a cold and hot retrieval strategy that makes backup and recovery fail.

TableComparisonbetweenVariousTechniquesof Back-Up andRecovery

S.No	Approach	Advantage	Disadvantage
1	HSDRT[1]	• Used for Movable clients like laptop,Smart Phone	• Costly • Increase redundancy
2	Parity Cloud Service[2]	• Reliable • Privacy • Lowcost	• Implementation • complexity is high
3	ERGOT[4]	• perform exact-match retrieval • Privacy	• Time complexity • Implementation complexity

4	Linux Box[5]	• Simple • Low cost for implementation	• Required higher bandwidth • Privacy • Complete server Backup at a time
5	Cold /Hot Back-up	• Triggered only whenfailure	• Cost increases as dataincreases

REFERENCES

[1] Yoichiro Ueno, NoriharuMiyaho, Shuichi Suzuki,MuzaiGakuendai, Inzai-shi, Chiba,Kazuo Ichihara, "Performance Evaluation of a Disaster Recovery System and Practical Network System Applications," Fifth InternationalConference on Systems and Networks Communications,2010, pp256-259.

[2] Chi-won Song, Sungmin Park, Dong-wook Kim, Sooyong Kang, "Parity Cloud Service: A Privacy-Protected Personal Data Recovery Service," International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11, 2011.

[3] Y.Ueno, N.Miyaho, and S.Suzuki, "Disaster Recovery Mechanism using Widely Distributed Networking and Secure Metadata Handling Technology", Proceedings of the 4th edition of the UPGRADE-CN workshop, 2009, pp.45-48.

[4] Giuseppe Pirr´o, Paolo Trunfio , Domenico Talia, Paolo Missier and Carole Goble, "ERGOT: A Semantic-based System for Service Discovery in Distributed Infrastructures," 10th IEEE/ACM International Conference on Cluster, Cloud and GridComputing,2010.

[5] Vijaykumar Javaraiah Brocade Advanced Networks and Telecommunication Systems (ANTS), "Backup for Cloud and Disaster Recovery for Consumers and SMBs," IEEE 5th International Conference, 2011.

[6] Lili Sun, Jianwei An, Yang Yang, Ming Zeng, "Recovery Strategies for Service Composition in Dynamic Network," International Conference on Cloud and Service Computing, 2011.

[7] XiZhou,JunshuaiShi,YingxiaoXu,YinshengLiandWeiwei Sun, 2008, "A backup restoration algorithm of service composition in MANETs," Communication Technology ICCT 11th IEEE International Conference, pp.588-591.

[8] M. Armbrust et al, "Above the clouds: A Berkeley view of cloud computing,

[9] F.BKashani, C.Chen, C.Shahabi.WSPDS, 2004, "Web Services Peer-to-Peer Discovery Service," ICOMP.

[10] Eleni Palkopoulou¤y, Dominic A. Schupke, Thomas Bauscherty,, 2011, "Recovery Time Analysis for the Shared Backup Router Resources (SBRR) Architecture", IEEE ICC.

[11] BalazsGerofi, Zoltan Vass and Yutaka Ishikawa, "Utilizing Memory Content Similarity for Improving the Performance of Replicated Virtual Machines", Fourth IEEE International Conference on Utility and Cloud Computing2011. [

12] P.Demeester et al., 1999. , "Resilience in Multilayer Networks," IEEE Communications Magazine, Vol. 37, No. 8, p.70-76. S. Zhang, X. Chen, and X. Huo, 2010, "Cloud Computing Research and Development Trend," IEEE Second International Conference on Future Networks, pp. 93-97.

[13] T. M. Coughlin and S. L. Linfoot, 2010, "A Novel Taxonomy for Consumer Metadata," IEEE ICCE Conference.

[14] K. Keahey, M. Tsugawa, A. Matsunaga, J. Fortes, 2009, "Sky Computing", IEEE Journal of Internet Computing, vol. 13, pp. 43-51.

[15] M. D. Assuncao, A.Costanzo and R. Buyya, 2009 , "Evaluating the Cost- Benefit of Using Cloud Computing to Extend the Capacity of Clusters," Proceedings of the 18th International Symposium on High Performance Distributed Computing (HPDC 2009), Munich, Germany.

[16] Sheheryar Malik, Fabrice Huet, December 2011, "Virtual Cloud: Rent Out the Rented Resources," 6th International Conference on Internet Technology and Secure Transactions,11-14 ,Abu Dhabi, United Arab Emirates.