Journal of Innovative Engineering and Research (JIER)
Vol. - 6, Issue -2, October, 2023, pp. 6-10 (5 pages)

Data Protection in the Cloud: A Backup and Recovery Approach

H. NAGA CHANDRIKA[1] and DR. PRAMOD PANDURANG JADHAV[2]

[1,2]*Department of Computer Science and Engineering, Dr. A. P. J. Abdul Kalam University, Indore 452010, India*
Corresponding Author Email: ppjadhav21@gmail.com

Abstract—

The business occasionally makes use of unregistered functions in its database, which can greatly boost speed. The only recovery possibilities available with this method are backups, thus any essential steps must be made right away after the session. In today's commercial context, this could result in data loss and up to a week of downtime, which could cause bankruptcy. Despite having backup solutions in place, the organization has not been able to restore data within the SLA time range. In order to improve backup and recovery processes in cloud computing, a secure database monitoring mechanism has been proposed to address this problem. With a minimum of 30% yearly data increase, this strategy makes sure that the backup speed is proportional to the amount of data. three to four years. The backup speed can remain constant while the data can at least double and, in some cases, increase. To fulfil the needs of the business, SLAs and business requirements for data recovery, such as recovery point target and recovery time objective, must be doubled.

Index terms: ***Data Protection, Cloud Computing, Backup, Recovery Approach***

I. INTRODUCTION

The company is undergoing a transformation where all business processes are being moved to IT systems, causing a disappearance of paper documents, such as copies, originals, and scans. Data loss in this context could result in a complete loss of everything. In the cases examined, data unavailability caused companies to be unable to function, leading to direct and indirect losses, including reputational damage that is difficult to measure in monetary terms. We have considered the recovery time objective, where the actual recovery time should grow with the data, and SLA requirements become more stringent. Most clients have received information about when the actual graph time matches the required time. In practice, many errors can occur, making it crucial to adopt appropriate measures to protect data in the cloud. Data loss can result in the loss of a portion of data, and traditional backup tools often require restoring the entire system. For instance, restoring a 15 TB database can take days. In today's world, where everything is stored in IT systems, every second creates data that needs immediate protection. Traditional backup systems are not adequate for this purpose, as they cannot provide immediate protection for data from the moment it is created. Each data has a specific lifetime and exists in one copy all over the world. Customers want their data protected from the moment it is created. Restoring from a backup often requires restoring data from a day earlier and retrieving the data from somewhere else a day later, which is a time-consuming task for administrators that can take several days. In the worst-case scenario, it can result in the loss of vital information. The issue is not limited to backup; it concerns the overall design of the system. The title of the IT system is crucial, and it cannot be overlooked. However, there are no easy and inexpensive options to verify the quality of the backup. Periodic test recoveries can be performed but they are costly in terms of human effort and IT resources, requiring a separate team and hardware. Many customers do not carry out this practice. Although backups are created, they often cannot be restored, despite the proper functioning of the RMS, due to various reasons. An example best illustrates this issue. The backup system is a service subsystem of a data center and has specific features. The loss of data is not critical to solving information system (IS) problems, and a backup failure does not affect the availability of critical information services. However, the backup process creates a computational burden that is not beneficial to the provision of IS information services. Unfortunately, this critical aspect was overlooked by the maintenance industry for many years, leading to the complete loss of some data. In some cases, it took days to recreate the infrastructure services due to the absence of backups of operating systems, binaries, configurations, and other critical information. One of our customers used an SAP system with an Oracle database and relied on inbuilt SAP tools and external vendors for backups. They had two backup policies: one was file-based, which copied the operating and software systems data, and the other was database-based. As they were sent to the same system, an exclusion list was created and entered into the database. The file policy considered this list and did not allocate directories in the database.. However, due to the uniqueness of the architecture, the database backup policy ignored the exclusion list and correctly copied the required data.

II. RELATED WORK

Numerous backup and recovery strategies, including HSDRT, PCS, ERGOT, Linux, Cold and Hot backup schemes, have been discussed in the literature. However, a comprehensive study indicates that these methods do not always yield optimal results in terms of data protection, costs, duplicates, and recovery following lost files. Among these techniques, PCS is a reliable, fast, user-friendly, and efficient method for data recovery. It focuses mainly on parity recovery and creates a virtual disk in the user system, where parity groups are formed using parity data for storage. The parity knowledge is generated by EXOR. Although it sounds promising, the implementation complexities can be challenging to manage. On the other hand, the HSDRT technique is suitable for mobile users such as laptop and smartphone users, but it involves high implementation costs and does not address duplication issues. It also uses high- frequency symmetric cryptography for widely distributed data transfers and offers a data backup process. Several backup and recovery strategies have been reviewed in the literature, such as HSDRT, PCS, ERGOT, Linux, Cold and Hot backup schemes, among others. However, these technologies do not always provide the best possible results and outcomes in terms of protection, costs, duplicates, and recovery following lost files. PCS has been identified as a reliable, quick, and easy-to-use technique for data recovery. It focuses mainly on parity recovery, where a virtual disk is created in the user system, and parity groups are formed to store data using parity data. Although this technique sounds efficient, its implementation can be complex and costly. HSDRT is useful for mobile consumers, such as laptop and smartphone users, but it entails high implementation costs and does not address duplication. This technique uses high-frequency symmetric cryptography for broadly distributed data transfers and provides a data backup process. The Efficient Routing Grounded on Taxonomy (ERGOT) method, on the other hand, focuses on semantic design to aid cloud computing in service discovery. This approach considers the semantic relationship between service descriptions and service requests in the data recovery component. The Linux Box model offers a low-cost implementation and fast data migration from cloud to cloud. However, this model lacks protection. This model is suitable for all kinds of users, especially small and medium-sized businesses. Although each of these techniques has its strengths and limitations, there is no perfect solution due to the costs and replication problems. The proposed approach is cost- effective and accessible to all types of users, including small and medium businesses. Unlike the Linux box model, this approach eliminates high deployment costs and synchronizes data from the cloud service provider to the user at the block or file level. According to K. Deerthana (2019), cloud computing is an internet-based computing system that provides shared resources, data, and information to computers and various devices. Cloud computing enables the creation, configuration, and customization of business applications over the web. To mitigate the risk of unauthorized access and data breach, it is essential to ensure data protection. This paper proposes a combination of cryptography (using the Blowfish algorithm) and steganography for secure data protection.

III. SECURITY & PRIVACY

Security has, as was already said, been a major issue in the adoption of cloud computing. Trusting someone else's hardware and computer to store your data and applications can be unsettling. The safety of data and software is seriously threatened by security vulnerabilities like data loss, phishing, and botnets. Additionally, the new features of the Cloud, such multi-tenancy and shared computing resources, have created new security concerns that call for new defenses. For instance, hackers can simply launch assaults since cloud infrastructure services are more dependable and less expensive. When migrating their data to the Cloud, organizations are nonetheless concerned about data privacy. Compared to IaaS, IT management and personal applications are thought to be simpler to migrate to the cloud. This is due to the common practice of outsourcing simple tasks. The backup speed can remain constant while the data can at least double and, in some cases, increase. To fulfil the needs of the business, SLAs and business requirements for data recovery, such as recovery point target and recovery time objective, must be doubled.

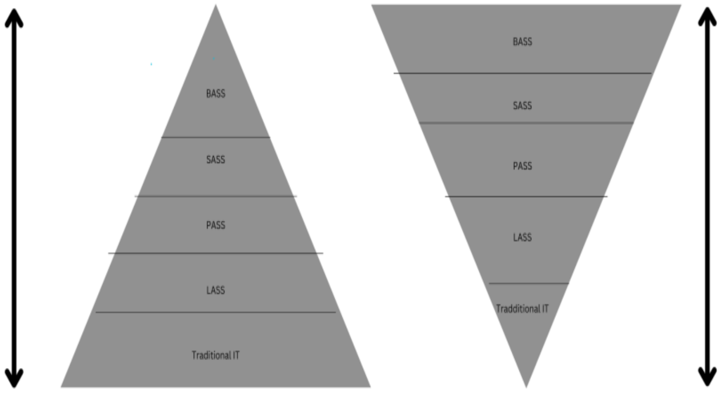

FIGURE 1.: CLOUDS BUILDING BLOCKS IN HIERARCHY

IV. LOSS OF CONTROL AND DATA

Due to the difficulties involved, businesses have traditionally been cautious to move their data and services to the cloud. The fact that the cloud service provider can host data anywhere in the cloud presents one such difficulty. Organizations are concerned about the security of their important data because of this lack of control and visibility. One of the advantages of the cloud, multi-tenancy, can potentially be a security concern if the right security measures are not taken. Sensitive information may be accessed by unauthorized people, which could result in data loss or leakage and cause harm to an organization's finances or reputation. Unauthorized access, poor authentication and accounting controls, conflicting encryption keys, operational failure, service interruption, and unreliable data are some of the factors that can cause data loss. centers. Strong network authentication and authorization procedures and transparent key exchange techniques, which are well-known to both cloud users and providers, could be implemented as a solution to these security issues.

V. CLOUD COMPUTING ARCHITECTURE

Spectrum Enterprise states that most organizations have a hybrid IT structure, which combines physical infrastructure with cloud-based infrastructure. By 2017, Gartner predicted that half of mainstream enterprises would adopt hybrid infrastructure. The hybrid cloud model provides easier access to and management of private cloud solutions located on-premise. This approach allows organizations to assess resources for each job and determine which application is best suited to perform the task. Applications with varying levels of network resource requirements are better suited for public or private clouds. The complexity of hybrid cloud architectures can vary, and some organizations use hybrid cloud approaches by connecting SaaS expenditure tracking applications to their billing systems. This provides them with seamless access to cloud-based solutions.

VI . PROPOSED MODEL

The goal of the suggested remedy is to make the Seed Block Algorithm more user-friendly and secure. The method entails putting a file on a Cloud Server and a backup file on a Remote Server at the same time. The encoded version of the file is kept on the Remote Server, while the ordinary version is saved on the Cloud Server. The Seed Block Algorithm and a cyclic redundancy check (CRC) that is applied to the file are both included in the encoding. The file is first transformed into an EXOR file, followed by a CRC encoded file, after the Seed Block Algorithm has been applied. The file is thus twice protected in this scenario. However, since a rogue user can still upload, encoding by itself is insufficient to provide total protection. To mitigate this risk, an authenticated login is required. Users must register themselves and have their parameters checked for genuineness before gaining access to the data storage. To improve backup and recovery speed, it is important to separate it from the computer volume. Various tools recommended by data storage systems, application software, and RMS manufacturers can be used to achieve this. One such tool is snapshots, which allows for quick data backup and recovery with minimal impact on performance. By using sequence and controlling SRK, snapshots can be made a part of an organization's backup policy. Another solution involves using utility tools like Oracle Standby, DB2 HADR, and MS SQL Always On to create a working copy of the production system that can be quickly deployed in case of failure. Secondly, restoring only the necessary data is crucial. Rather than copying the entire system, the ability to deploy or use existing systems that contain the required data is more efficient. Snapshots can also be used here to open a snapshot on a neighboring server and extract the necessary data. Continuous data protection technologies like Oracle Standby with Flashback can also be employed for quick deployment of a working copy of the data. To retrieve a specific logical volume, such as a row or database table, various tools are available to make the task easier and enable selective data restoration without the need for a complete copy recovery. The third challenge is to minimize the time gap between the data source and its protection, which can be accomplished in multiple ways, depending on the importance of the data and the specific requirements. For instance, for less critical systems, the backup interval can be shortened to a few hours using snapshots as restore points, which can be taken every hour. Some modern storage arrays handle this process efficiently and can store a large number of system snapshots, offering a useful solution when it is necessary to revert to a previous state of the system. To simplify the task of restoring a logical volume, such as a row or a database table, various tools are available that allow for selective restoration of data, without restoring the entire copy. The frequency of backups for less critical systems can be increased to a few hours, using snapshots as restore points. Some modern arrays are capable of handling such processes and storing large numbers of system snapshots. However, for critical systems, continuous data protection is essential, and there are many solutions available such as Oracle Standby with Flashback, which records all changes and allows the database to be rolled back to any point in time. Other options include general-purpose software and hardware systems that record all changes and enable restoration to any point in time. While regular backups are still an important tool for disaster recovery or data retrieval from the distant past, they are now viewed as a last resort option.

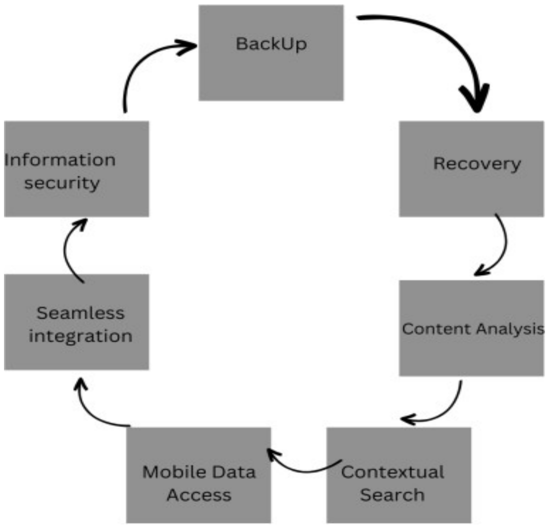

FIGURE 2. BLOCKS FOR PROPOSED METHOD

The fourth strategy is aimed at reducing hidden errors. The only surefire way to ensure that a backup is functioning correctly is to test it by attempting to restore it, but unfortunately, this approach is rarely taken by clients. To address this issue, we suggest having easily recoverable instances of computers, such as snapshot and standby systems that can be swiftly deployed and tested. This process requires far less time and effort than restoring an entire backup. While this may not always be a panacea, it at least provides some additional confidence that data can be restored in the event of an emergency. Additionally, some SRCs offer automated testing, which enables virtual machines to be run on a predetermined schedule in an isolated environment. These machines use predefined algorithms to verify that data has indeed been retrieved, that the application is operational, consistent, and responsive to necessary requests. By utilizing this method, administrators can reduce their workload on lengthy and repetitive tasks. The centralized backup system's multi-layered architecture includes the following components, as illustrated in Figure 3.

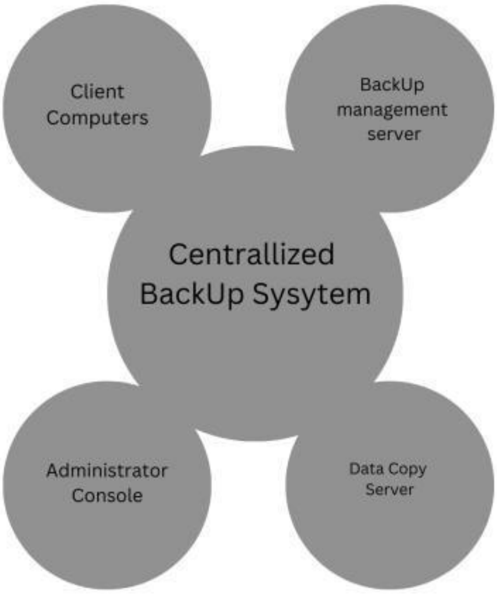

FIGURE 3: COMPONENTS ON BACKUP SYSTEM

The fourth objective is to minimize hidden errors in the backup process. The best way to ensure that the backup system is working properly is to attempt to restore it. However, this approach is rarely used. Instead, we suggest having easily recoverable instances of computers, such as snapshot and standby systems that can be rapidly deployed and tested. This is a much quicker and easier process than restoring an entire backup. Automated testing is another option, which is available through some SRCs. At predetermined times, virtual machines can be launched in a secure environment, and predefined algorithms can be used to check that the data has been retrieved, the application is available, it is consistent, and it is responding to necessary requests. This relieves administrators of routine tasks. A centralized backup system has a multi-layered architecture that includes a backup management server, one or more data copy servers, client computers with backup agent programs, and a Backup System Administrator Console. The fifth objective is to make the backup system transparent. Creating an integrated system using different technologies from various manufacturers can be challenging. There are two ways to approach this issue: customers can implement the system independently, with our assistance in creating necessary processes, regulatory frameworks, instructions, and plans to enable their IT department to operate and expand the system more independently, or they can partially or entirely outsource the system to us if they are unsure whether they can maintain the SRK system continuously. Some of our clients are already using this service, and as our IT outsourcing engagement levels and SLA requirements continue to grow, we expect to see more.

VII. CLOUD SYSTEM

Various criteria can be used to categorize cloud computing. The two most frequently utilized subcategories among these are service border and service type. Public cloud, private cloud, and hybrid cloud are the three different types of cloud computing from the perspective of the service boundary. While private cloud is created and run by businesses for their own use, public cloud services are offered to third parties. On the other hand, a hybrid cloud uses a secure network to transfer resources between private and public clouds. Examples of hybrid cloud services are Google's and Amazon's Virtual Private Clouds (VPC). Cloud computing can be divided into three categories from the perspective of the service type: infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). SaaS delivers services straight to customers, whereas IaaS and while IaaS and PaaS provide services to Independent Software Vendors (ISVs) and developers, allowing room for third-party application developers.

VIII. CHARACTERISTICS OF CLOUD COMPUTING

Cloud computing is a technology that has gained a lot of attention in recent years, but it can still be confusing to those who are not familiar with it. Understanding the essential characteristics of cloud computing is crucial for any business that is considering implementing this software offering. There are five main characteristics that define cloud computing. i. On-demand self-service: This allows users to provision computing resources automatically without the need for human interaction with the service provider. This includes server time and network storage. ii. Broad network access: Capabilities are accessible over the network and can be accessed through standard mechanisms, making it possible for users to access them through a variety of devices, including mobile phones, tablets, laptops, and workstations. iii. Resource pooling: Providers of cloud computing services use a multi-tenant model to pool their computing resources, which are dynamically assigned and reassigned to meet the demands of multiple consumers. The exact location of the provided resources is generally unknown to the customer.

IX. CONCEPTUAL MODEL OF CLOUD STORAGE SYSTEM

Cloud storage is a collection of storage devices linked together by a network, distributed file system, and other middleware to offer customers a cloud-based storage service. Software-defined storage, standalone arrays, converged infrastructure, hyper converged infrastructure, and public cloud storage are just a few of the options that are possible for the storage. Various network infrastructures, including fibre, iSCSI, NFS, and SMB, are utilized to connect these storage systems. Block, file, or object storage may be used. Some of these storage options include NVM e-based arrays, hybrid arrays, HCI, public cloud for primary and backup storage, and container storage. Typical organizational structures for cloud storage include storage resource pools, distributed file systems, service level agreements, and service interfaces. A five- layer Cloud storage idea is shown in F

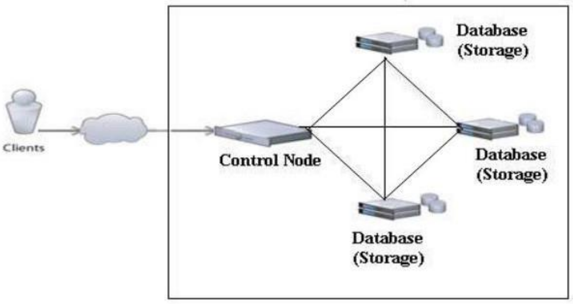

FIGURE 4. PROPOSED MODEL CLOUD STORAGE

X. CONCLUSION

Daily, enormous amounts of data are generated and processed as computing and technology use grow. Data must be securely preserved in order to ensure effective backup. A suggested plan combines two strategies to make content backup easier and provide secure transaction techniques. The suggested approach enables fine-grained access control, but it makes the cloud out to be a reliable third party and ignores network problems. Traditional cloud service providers offer a variety of security measures, but they might not always inspire their clients with enough trust. This work offers a holistic approach that supports and collaborates with clients and partners as it handles security frameworks like NIST, CSA, and HIPPA. To satisfy requirements for consumer protection, the device can be updated. Data is tested during the transmission and power phases before being combined and delivered to the cloud to serve as the suggested data and service protection framework. Customers can utilize the cloud without worrying about their security or quality thanks to the network protocol's transformation of the cloud into a safe, effective, and reliable environment. In addition to the dangers associated with virtualization and multi- tenancy, data in the cloud can also be at risk if it is not sufficiently secured. The study addressed data in multiple stages and effective encryption approaches employing block cypher, stream cypher, and hash functions. Its main focus was data security and hazards in cloud solutions.

REFERENCES

[1] Jyoti, A., Shrimali, M., Tiwari, S., & Singh, H. P. (2020). Cloud computing using load balancing and service broker policy for IT service: a taxonomy and survey.*Journal of Ambient Intelligence and Humanized Computing*, *11*, 4785-4814.

[2] Cao, S., Zhang, G., Liu, P., Zhang, X., & Neri, F. (2019). Cloud-assisted secure eHealth systems for tamper-proofing EHR via blockchain.*Information Sciences*,*485*, 427-440.

[3] Li, G., Yan, J., Chen, L., Wu, J., Lin, Q., & Zhang, Y. (2019). Energy consumption optimization with a delay threshold in cloud-fog cooperation computing.*IEEE access*,*7*, 159688-159697.

[4] Anawar, M. R., Wang, S., Azam Zia, M., Jadoon, A. K., Akram, U., & Raza, S. (2018). Fog computing: An overview of big IoT data analytics.*Wireless Communications and Mobile Computing*, *2018*. [5]Humayun, M. (2020). Role of emerging IoT big data and cloud computing for real time application. *International Journal of Advanced Computer Science and Applications*, *11*(4). [6]Deerthana, K., & Jayamala, B. D. S. R. Enhancing Security using Cryptography and Steganography and Providing Data Backup and Recovery in Cloud.

[7] Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. *Journal of network and computer applications*, *34*(1), 1-11.

[8] Yang, C., Liu, Y., & Tao, X. (2020). Assure deletion supporting dynamic insertion for outsourced data in cloud computing. *International Journal of Distributed Sensor Networks*, *16*(9), 1550147720958294. [9]Challagidad, P. S., Dalawai, A. S., & Birje, M. N. (2017). Efficient and reliable data recovery technique in cloud computing. *Internet of Things and Cloud Computing*, *5*(1), 13-18.

[10]Kollu, P. K. (2021). Blockchain techniques for secure storage of data in cloud environment. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(11), 1515-1522.