ISSN : 0973 - 8355

www.ijmmsa.com

IJMMSA

INTERNATIONAL JOURNAL OF
MATHEMATICAL, MODELLING, SIMULATIONS AND APPLICATIONS

E-MAIL
editor.ijmmsa@gmail.com
editor@ijmmsa.com

ISSN 0973-8355www.ijmmsa.com
Vol 7,Issuse 2.June 2019

“Enhancing Cloud Productive Storage Performance By Using Data Deduplication”

[1]P.Chandrakalavathi, [2]Fareesa Firdose , [3]Shaik Reshma

ABSTRACT”

To provide users with easy access to cloud storage, cloud computing has emerged as a significant trend in recent years. A widely used method of data compression is deduplication. Use this method to limit the amount of redundant data. Cloud storage uses information deduplication to maintain its capacity and reduce its storage space use. To deal with the following issues, this work tries to codify the concept of safe and productive cloud storage. A key aspect of openness and access to data is achieved in this manner Text files are accepted as input. To begin with, the record is sorted through the text files to identify the most important information in each. These components are used in this way to differentiate between the same kind of information, having a comparison. In advance, the records are sorted based on their spatial relationship. As a result of Tiger's hash creation process, a key is generated that will be encoded using 3DES. The findings reveal that the recommended solutions have a positive impact on both the contrasted and the traditional systems.

1.INTRODUCTION

This Picture Cloud computing has enabled the Internet to spread to new corners and gadgets, not only in the original networks, but also in numerous applications. Data access and cloud network node management and control are important topics to stress since the effectiveness of control techniques greatly influences cloud network performance and quality[1]. However, with the emergence of multiple devices and data.

A. CloudStorage

It may be summed up as "keeping data in the cloud". It is known for its dependability, openness, and speed of deployment. Public, private, and hybrid storage options are all available.

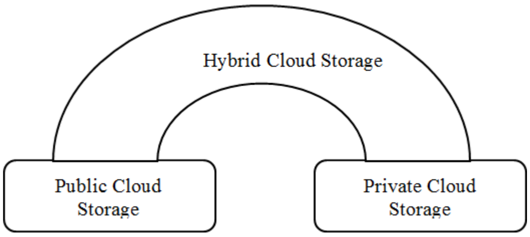

Figure1. CloudDeployment Model

1) PublicCloudStorage:With the storage of unstructured data at global hubs, this is how it is done. Customers must purchase storage space depending on the amount of data each user generates. Cloud storage providers handle the project's cloud storage entirely.
2) Private Cloud Storage: Private cloud is most often used by customers that want greater control over their data. As a result of using private cloud storage, security and performance concerns may be reduced while still reaping the advantages of cloud computing.

[1,2,3]Assistant Professor
[1,2,3] Department of Computer Science & Engineering,
[1,2,3]Dr.K.V.Subba Reddy College Of Engineering For Women

Hybrid Cloud Storage:Cloud storage service is a mix of public and private cloud storage. Typical data on the endeavor private cloud location. Meanwhile, a public cloud storage service is used to store and make accessible additional data. This is followed by [3][4].

Symmetric encryption's intrinsic safety vulnerabilities are discussed in this study, along with cloud deduplication's benefits and advantages. As a result, the benefits of both deduplication and symmetric encryption are retained. Tosummarizeourcontributions:

- Through the use of symmetric encryption, it provides block-level deduplication while still maintaining the secrecy of data.
- Reducing hardware backup and storage costs via data deduplication was achieved.
- It improves storage capacity.
- Confidentiality and privacy are protected even from potentially hostile cloud storage because of this technology.
- This means that cloud deduplication is completely compatible with existing cloud storage providers.

In today's world, cloud computing providers are becoming more and more popular since they provide a more environmentally friendly method of storing data. Cloud computing's fundamental drawback is its inability to manage massive amounts of data. To help us do this, a deduplication approach is used. There are several advantages of deduplication. [5] There are a few safety concerns. Deduplication's security issues have prompted us to provide a strategy that allows for permitted deduplication on the cloud server while also managing those issues. When it comes to cloud data storage security, we have designed a security framework to ensure that the data file can only be read by those who need to access it [6].

B. DataDeduplication

When a user saves data, numerous copies of the same file are made in various locations, resulting in increased storage complexity. The technology of data deduplication plays a significant role in cloud computing architectures. This is the method that eliminates the cloud's redundant data. ' If the user uploads the same file that is already saved, the cloud provider will add the user to its owner list, so he/she may access the file at any time. Storage complexity may be reduced by using dedupes. Deduplication and encryption are two separate methods. In the event that customers fail to encrypt their data, a cloud storage provider will still be able to safeguard it because of the aftereffect of encryption: Ithastwolevels-

i. FileLevelDeduplication:In file-level deduplication, the data in a duplicate file is deleted. It is also often referred to as "single-instance" storage (sis).

ii. BlockLevelDeduplication:When two files are not identical, they share the same blocks of data. If you are looking for a way to increase storage capacity, block-level deduplication is the way to go.

2.LITERATURESURVEY

It was previous to outsourcing that Jin Li et al. [1] invented a data encryption technology. File tokens created by a private cloud server are used to execute a duplication check in a hybrid cloud. There is a unique file token for each file that is linked to this. You will not get the secret key via email or any other means from us. In other words, the privilege key will not be controlled because of a privately hosted cloud server. Each private cloud server issues a file token to the user who requests it after verifying his or her identity with a private cloud service provider. Before file encoding, the public cloud does a check for duplicates. Compared with regular operations, it will have a negligible impact on expenses.

The safe and effective cloud storage solution is provided by S. A. Maindakar and Dr. M. Z. Shaikh. MaindakarDeduplication for hybrid cloud storage and sharing was suggested by them. Proof of ownership of files and two-factor authentication are also used. Using a one-time password to keep your account safe. As a consequence, in a real-world setting, the additional overhead is minimal. There is a proposal and implementation on OpenStack by NesrineKaaniche and Maryline Laurent. To store and share outsourced data in the public cloud, another customer-side deduplication approach is put out here. Using it, you may do two kinds of access control checks at once. The data is encrypted using symmetric encryption, while the metadata is encrypted using asymmetric encryption. Data is encoded using a unique key for every piece of information on each customer's side [2] Adding permissions to a metadata document ensures that the client may decode the information contained inside. It can only be decrypted with his private key. In addition to this setup, it seems to be impenetrable against unwanted access to data. Pasquale Puzio and his colleagues suggested To provide data security while also ensuring duplication at the block level, Cloudedup uses reagent encryption records and Cloudedup. Deduplication at the block level was employed instead of at the file stage. On a block-by-block basis, there is considerable difficulty controlling the keyboard Additionally, the server and optional HSM provide a new level of security while still using little resources. Current and full techniques may readily benefit from this approach.

3.RELATED WORK

A. "DifferenceBetweenSHA1/SHAandMD5"

"Text FeatureExtractionTechniques"

i. Bag-of-words(BOW):A word is commonly counted in a text or file using this approach.
ii. Term frequency/inversedocumentfrequency (TF-IDF):When calculating TF-IDF, the approach de-emphasizes repeated terms in numerous documents while highlighting words that appear just once in a single file or document.

PROBLEMANDSOLUTION

A data deduplication technique for cloud storage difficulties is suggested in this paper. Here, a model is described that uses data to find duplicates and then removes them from storage. In addition, the data in storage that is discovered to be a duplicate is now included. The usage of an inverted index may aid in locating certain pieces of data on a cloud server. Additionally, it is made simple to access data searches (availability). Along with the hash tree, we were able to locate comparable files based on their content and eliminate server duplication.

A vast quantity of data is stored on a cloud server. Images, music, video, text, and other types of data may all be included. The user may upload a text file from their computer by selecting it and then uploading it to the cloud storage. An example of a TF/IDF approach would be the removal of stop keywords and special characters such as separators (i.e. After this, the indexing method is employed for domain identification. It is expected that the server has a certain amount of data and its choices. A list of features is then compared to this file-based choice. Using the most matched extracted feature, this is taken into consideration. After defining the domain of information, the file is regarded as a result of the domain. After this step, the server will be able to store encrypted data. For the purpose of encrypting data, two separate algorithms are combined to create a merging algorithm. During this step, the original file is encrypted and sent to the system as input. The tiger hash generation technique is used to produce the hash key initially during the encryption procedure. This generates a 512-bit key block, but 3DES can not accept keys that long, thus the algorithm discards the last bits of the tiger hash.

- SHA/SHA1:160bits (20-byte digest) is the key length. While the MD5 algorithm is faster, it is more vulnerable to brute force attacks.
- MD5: It has a digest key length of 128 bits/ 16 bytes. However, despite its superior speed, this algorithm is less safe than SHA1. Compiling costs less money.

B.CryptographicTechniqueComparison

Data from many fields may be easily sorted using indexing. Information structures that store the field cost and pointer to the document it pertains to may be created by creating an index on any given topic. A basic set of processes may greatly minimize the amount of time it takes to compute.

The remaining 168 bits are utilized to encrypt the file using the 3DES technique. A decrypted file is utilized in the next step of the procedure Using the encrypted file, the data is then divided into blocks of a certain size. The 512-bit data block is produced in this step. Designing their own size is up to the individual, since there is no predetermined standard. File blocks are not used in this step of the data processing procedure. The SHA1 algorithm is used to compute a hash of the block contents for each individual block. Each block of data is mounted in binary fashion using the tree. The binary method is used to mount each block of data on the tree. Additionally, the block hashes are compared to the leaf nodes of the current binary tree in order to discover duplicate data before the tree is constructed. There is no further construction of the tree if the data is discovered in any of the leaf nodes. Once the file's availability has been verified, it is placed in the server's storage space, where a tree search may quickly locate other files with the same name.

A."AlgorithmUsed:

Inthissectionweusedtwotypesofalgorithm:

1) ForFileUploading
2) ForFileDownloading

```
ForFileUploading:
Begin
   Step-1Readfile
   Step-2 Duplication checks by cloud
   serverStep-3 Sends the responses i.e. file
   exist or notStep-4
   If File does notexist
            4.1 File Uploaded
   successfullyStep-5If
   Filealreadyexist
            5.1DisplayFilealreadyexist
End
ForFileDownloading:
Begin
   Step-1ReadFile
   Step-2 Duplication checks by cloud
```

```
  serverStep-3 Sends the responses i.e. file
  exist or notStep-4
  4If Filedoesnotexist
          4.1 Display File does not
  existStep-5If Fileexist
          5.1FileDownloadedSuccessfully
End”
```

CONCLUSION

Suggested cryptography uses a symmetric key encryption for secure data retrieval by end-users With a dispersed network, it is possible to start transmission while maintaining security with a single storage location. There are other precautions in place, such as the use of well-known and trusted methods for authenticating the user and the maintenance of duplicate copies of all provided data on a cloud server that is updated on a regular basis. The availability of public cloud data is ensured through the use of multi-replica data blocks, which ensure data integrity between two parties. Using 3 DES for encryption and document indexing for hash trees, the data stored and sent across the trustworthy network was further protected. It is hoped that this method would reduce the complexity of data storage and boost the accessibility of data.

REFERENCES

1. Xin Yao, et al., “A Secure Hierarchical DeduplicationSystem in Cloud Storage” IEEE/ACM 24th InternationalSymposiumonQualityofService(IWQoS).Beijing,China, 1-10, 2016.
2. JunbeaomHur, et al., “Secure data deduplication withdynamic ownership management in cloud storage” IEEEInternationalConferenceonDataEngineering.SanDiego,CA, USA, 3113–3125,2016.
3. JingweiLi,etal.,(2015)“SecureAuditingandDeduplicationDatainCloud”IEEETransactionsonComputers,65(8),2386–2396.
4. ZuhairS.Al-Sagar1,etal.,“Optimizingthecloudstorage by data deduplication : A Study” InternationalResearchJournalofEngineeringandtechnology(IRJET),2(9), 2524-2527, 2015.
5. JadapalliNandiniandRamireddyNavatejaReddy“Implementation of hybridcloudapproach forsecureauthorizeddeduplication”InternationalResearchJournal of Engineering and technology (IRJET), 2(3),1297-1306,2015.
6. NimgireReshma,etal.,“Deduplication&secureauthorized data using hybrid cloud”. Imperial Journal ofInterdisciplinaryResearch, 2(6),415-419,2016.
7. Sumedha A Telkar and Dr. MZ Shaikh “Secured andefficientcloudstoragedatadeduplicationsystem”InternationalJournalofAdvancedResearchinComputer and communication Engineering, 5(1), 301-304,2016.
8. DifferencebetweenHashingandindexingfromhttps://www.quora.com/what-are-the-major-differences–between-hashing-and-indexing.
9. DifferencebetweenMD-5andSHA-1fromhttp://lnxsysadm.blogspot.in/2010/12/what-is-
10. difference-between-md5-and-sha.html.
11. WhatisTripleDESfromhttps://www.tutorialspoint.com/cryptography/triple_des.htm.