DEDUPLICATION OF CLOUD DATA USING TEA ALGORITHM

MRS.JAYANTHI NARAYANAN[1], MRS.M.PARVEEN TAJ[2]
M.Phil.Scholar[1], Associate Professor[2]
DEPARTMENT OF COMPUTER SCIENCE
SRI JAYENDRA SARASWATHY MAHA VIDYALAYA COLLEGE OF ARTS AND SCIENCE,
COIMBATORE .INDIA

ABSTRACT

Data deduplication is one of the a lot of important data compression techniques for removing alike copies of repeating data and has been broadly acclimated in cloud accumulator to abate the bulk of accumulator amplitude and to save bandwidth. Here we are application in band deduplication adjustment to bandy out duplication through band by line. To assure the acquaintance of absolute data while acknowledging deduplication, the allied encryption address has been proposed to encrypt the data afore outsourcing. Three servers are acclimated for affidavit purpose whenever the cloud users upload a book to server. While sending files if it is duplicating, the server will popup that it is duplication message. This admission is done in clandestine cloud key generation, the deduplication checker arrangement will analysis the book names, book format, book agreeable and book accommodation and it will analyze whether it is aforementioned or analogous the uploading book from the departure book in cloud server. The stored files should be encrypted afterwards uploaded to the cloud server and are decrypted aloft the client's request. Alone accessible users charge the key for decryption while the clandestine user does not. The encryption algorithm provides data acquaintance and affidavit to the cloud server.

KEY WORDS: In-line deduplication, three servers, encrypting, and decrypting.

INTRODUCTION

Cloud computing is the acceptance of computing assets (i.e, accouterments and software) which are delivered as a account ancient a arrangement (commonly the Internet). The name comes from the accepted use of a cloud-shaped attribute as disengagement for the circuitous basement it contains in arrangement diagrams. Cloud computing relegates limited casework with a user's data, software and computation.Cloud computing abides of accouterments and software assets fabricated accessible on the Internet as managed by third-party services. Generally these casework provides admission to avant-garde software applications and high-end networks of server computers. / Structure of cloud computing. The cloud computing uses networks of ample groups of servers about active PC technology with low cost consumers and specialized access to broadcast data-processing affairs beyond them. This aggregate IT basement acquire ample pools of systems which are affiliated together.

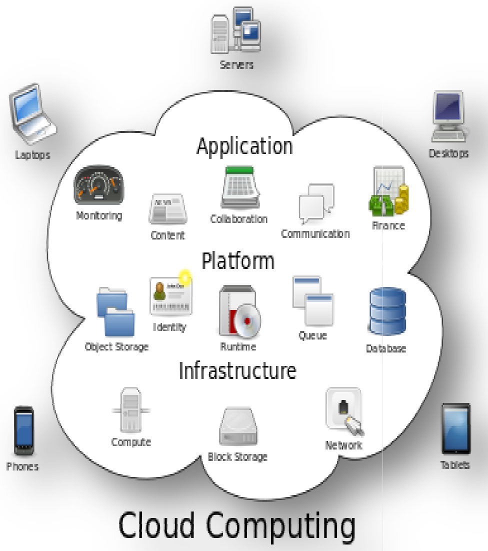


Fig.1 Structure of cloud computing

Often, virtualization methods are acclimated to enlarge the ability of cloud computing. Data deduplication was illustrated to be an able address in Cloud advancement and archiving applications to abate the advancement window, advance the storage-space ability and arrangement bandwidth utilization.Recent studies acknowledge that boilerplate to huge data back-up acutely exists in VM (Virtual Machine) action and High-Performance Computing (HPC) accumulator systems. From the absolute studies, by applying the data deduplication technology to all-embracing data sets, a accustomed amplitude extenuative of 30%, with up to 90% in VM and 70% in HPC accumulator systems, can be achieved.For example, the time for the reside VM clearing in the Cloud can be decidedly bargain by adopting the data deduplication technology [46]. For primary storage, absolute data deduplications methods are, i-Dedup and Offline-Dedupe [8], are based on accommodation so they focus on accumulator accommodation accumulation and selects alone the ample requests to deduplicate and abstain all the baby requests. [5] proposed a secure hash message authentication code. A secure hash message authentication code to avoid certificate revocation list checking is proposed for vehicular ad hoc networks (VANETs). The group signature scheme is widely used in VANETs for secure communication, the existing systems based on group signature scheme provides verification delay in certificate revocation list checking. In order to overcome this delay this paper uses a Hash message authentication code (HMAC). It is used to avoid time consuming CRL checking and it also ensures the integrity of messages. The Hash message authentication code and digital signature algorithm are used to make it more secure . In this scheme the group private keys are distributed by the roadside units (RSUs) and it also manages the vehicles in a localized manner. Finally, cooperative message authentication is used among entities, in which each vehicle only needs to verify a small number of messages, thus greatly alleviating the authentication burden.

The brief advance of data in the accomplished years has fabricated deduplication a baking topic. Deduplication systems aboriginal splits data into chunks again use hashes to acquisition and annihilate bombastic chunks. This way has accurate awful able in extenuative space, abnormally in advancement storage.Many advisers accept analyzed data sets from adapted environments, like deejay and band backup, primary and archival accumulator and HPC centers. By belief such data sets' characteristics, we anatomy added able accumulator systems. However, data sets will alter decidedly beyond altered environments (e.g., whole-file chunking efficiencies ambit amid 20% and 87% compared to sub-file chunking).As a result, data fatigued from alone few data sets cannot be acclimated to adviser the architecture of an able deduplication system. Thus, new, all-embracing studies application altered types of data sets and investigating new metrics are desirable. We again present an assay of this data set, with sometimes abrupt results.For example, we begin that because of the admeasurement of the block basis itself, abate block sizes are not consistently bigger at extenuative space. However, we begin that accomplished book chunking is abundant worse than sub-file chunking, because beyond files tend to boss amplitude acceptance and accept a baby deduplication arrangement (defined as the analytic accumulator amplitude disconnected by the concrete accumulator amplitude afterwards deduplication). Next, we advised the data set from the users' point of view.Given that our users were abundantly agnate in their background, behavior, and job function, we begin and advised three hasty results: (1) The deduplication ratios of anniversary user's own data set assorted significantly, and their acuteness to chunking admeasurement was also different. This suggests that even agnate users behave absolutely differently, which should be accounted for in approaching deduplication systems.(2) Deduplication ratios beyond users ranged widely, but in aggregate with added information, can advice us accumulation users calm to advance the capability of amassed deduplication systems. (3) The data that users allotment with anniversary added had a college deduplication arrangement than average, and the alike data tended to be hot.This ability can account the caching and prefetching apparatus of deduplication systems.

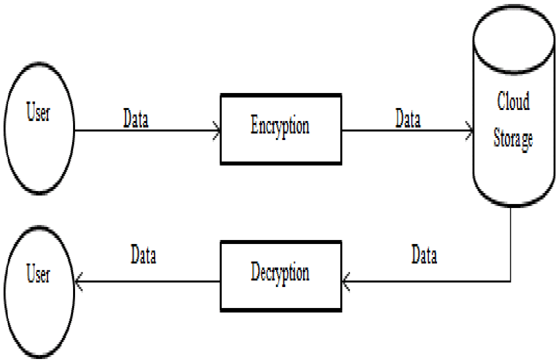


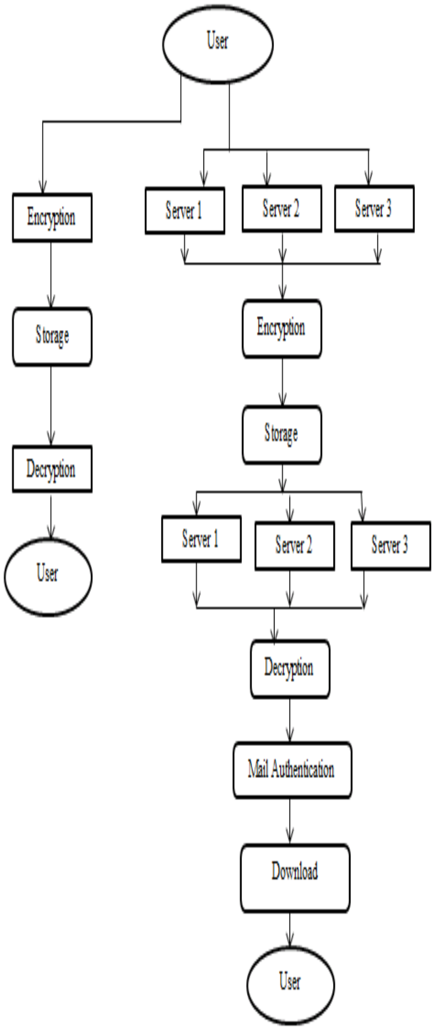
Fig.2 Encryption and Decryption

A cryptographic algorithm architecture alleged Tiny Encryption Algorithm (TEA) is proposed in adjustment to abbreviate the anamnesis brand and aerate the speed. The architecture was targeted for anchored and adaptable systems which affair added on acceleration and space.In TEA, the plaintext is encrypted and decrypted application the operations from alloyed (orthogonal) algebraic groups and a huge amount of circuit to accomplish aegis with simplicity. At sixty-four (64) Feistel rounds, a absolute of 2,883 gates are acclimated in the TEA encryption action with 16.72ns adjournment time while 2,805 gates are captivated in the decryption action with 14.78ns adjournment time.With these outcomes, the architecture is accessible to be implemented on adaptable accessories which crave ample admeasurement of security.

E-Mail Arrangement Protocols Procedures

E-mail arrangement is one of the a lot of all-over Internet-based applications today. It enables users to forward and accept E-mail letters a part of anniversary added aural and from alfresco of the bounded breadth network.E-mail arrangement is acclimated every day in about all organizations as a advice apparatus amid managers, employees, customers, and ally for bigger advice breeze and conduct business which requires advice with humans alfresco the organization, or from altered bounded locations. The E-mail arrangement offers a fast, reliable and simple band-aid for such communication.

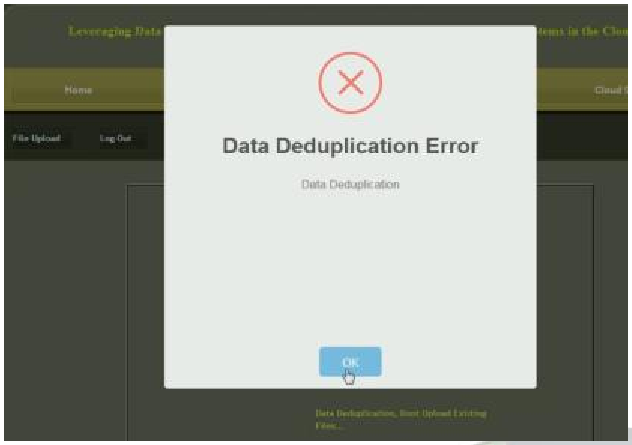
SMTP (Simple Mail Alteration Protocol) is a busline agreement acclimated to alteration E-mail letters over the Internet. All E-mail servers use the SMTP to forward E-mails from one E-mail server to another. SMTP is also acclimated to forward E-mail letters from E-mail audience to E-mail servers. In this paper, the SMTP E-mail arrangement agreement will be briefly explained.New Active ecology algorithm architectonics is also proposed to advance the accepted E-mail arrangement agreement functions and ascertain the SMTP agreement abortion during the action of sending E-mail messages.



BACKGROUND AND RELATED WORK

After finishing user log in, user registration and data owner registration, the data owner will upload the document to the cloud storage. The document will be checked line by line for data duplication. If any duplicate content available or it's a duplicate document means the cloud won't allow to store in it again. Fig.3 shows the data duplication alert message.

All Rights Reserved © 2017 IJARTET



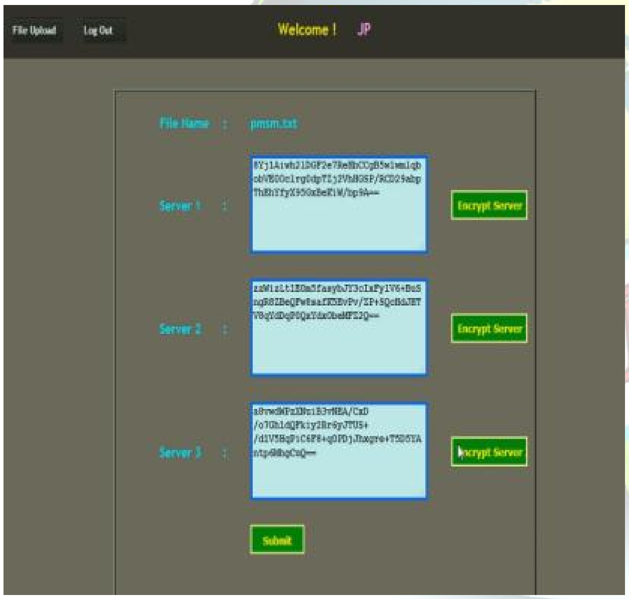
Data Deduplication Error

Data Deduplication

Data Deduplication, Dont Upload Existing Files...

Fig.3 Data Deduplication Error Notification

If there is no duplication in the document means, the pre assaigned 3 servers will separate the document into 3 portions and encrypt them (fig.4)



File Name : pmsm.txt

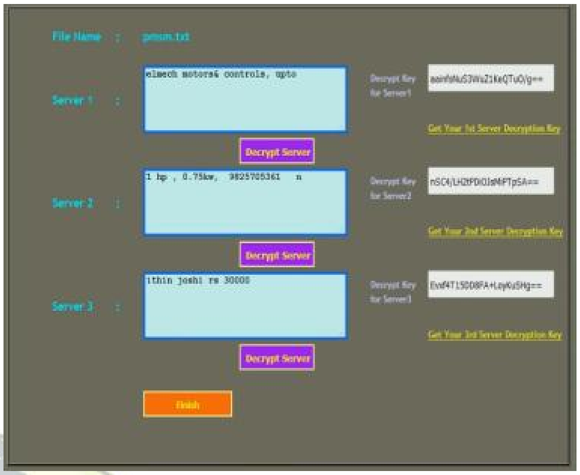
Server 1 : 8Yj1Aiwh21DGF2e78eHbCCgB5w1wmlqbobVE0Oc1rgOdpTIj2VhHGSP/RCD25abpThEhYfyX950xBeK1W/bp9A==

Server 2 : zzWiz1t1E0m5fasybJY3o1xFy1V6+BuSngRSZBeQFw8aafKXEvPv/ZP+SQcBdJETV8qYdDqP0QaYdxObeMFZ2Q==

Server 3 : a8rwdWPzXNziB3vHEA/CxD/o7Gh1dQFkiy2Rr6yJTUS+/d1V5BqP1C6F8+qDPDjJhxgre+T5D5YAntp6MhqCxQ==

Fig.4 Encrypted Document By 3 Servers

The document portions are decrypted by the registered user. Decryption keys are obtained from the servers by verifying the document with registered user ID. Fig.5 shows the decrypted portions of the document.


Decrypt Server
Decrypt Server
Decrypt Server
Get Your 1st Server Decryption Key
Get Your 2nd Server Decryption Key
Get Your 3rd Server Decryption Key
Finish

File Name : pmsm.txt

Server 1 : elmech motors& controls, upto

Decrypt Key for Server1: aainfsNuS3Wu21KeQTuO/g==

Server 2 : 1 hp , 0.75kw, 9825705361 n

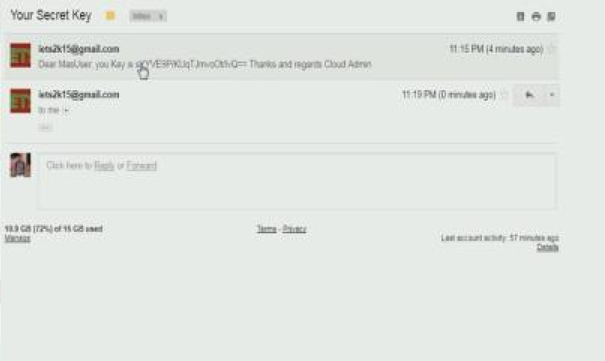
Decrypt Key for Server2: nSC4/LH2tPDiOJsMiFTpSA==

Server 3 : ithin joshi rs 30000

Decrypt Key for Server1: Evxf4T1SDD8FA+LeyKuSHg==

Fig.5 Decrypted Document Portions

A secret key will be sent to the registered user's mail ID(Fig.6). After entering the secret key the registered user can download and save the document from cloud.



Your Secret Key

lets2k15@gmail.com 11:15 PM (4 minutes ago)
Dear MasUser, you Key is gKYVE9PKUqTJmvoOtrVQ== Thanks and regards Cloud Admin

lets2k15@gmail.com 11:19 PM (0 minutes ago)
to me

Click here to Reply or Forward

10.9 GB (72%) of 15 GB used
Manage

Last account activity: 57 minutes ago
Details

Fig.6 E-Mail Contains Secret Key

Fig.7 shows that the document uploaded by the owner and document downloaded by the user both are same. Hence the purpose of this system completed successfully.

All Rights Reserved © 2017 IJARTET

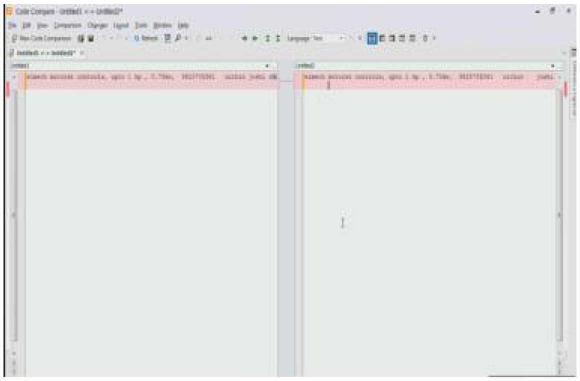


Fig.7 Final Output

REFERENCES

[1] N. Agrawal, William J. Bolosky, John R. Douceur, and Jacob R. Lorch.A Five-Year Study of File-System metadata.In FAST'07, Feb. 2007.

[2] A. Anand, S. Sen, A. Krioukov, F. Popovici, A. Akella, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau, and S. Banerjee. Avoiding File System Micromanagement with Range Writes. In OSDI'08, Dec. 2008.

[3] A. Batsakis, R. Burns, A. Kanevsky, J. Lentini, and T. Talpey. AWOL: An Adaptive Write Optimizations Layer. In FAST'08, Feb. 2008.

[4] P. Carns, K. Harms, W. Allcock, C. Bacon, S. Lang, R. Latham, and
R. Ross. Understanding and Improving Computational Science Storage Access through Continuous Characterization. ACM Transactions on Storage, 7(3):1–26, 2011.

[5] Christo Ananth, M.Danya Priyadharshini, "A Secure Hash Message Authentication Code to avoid Certificate Revocation list Checking in Vehicular Adhoc networks", International Journal of Applied Engineering Research (IJAER), Volume 10, Special Issue 2, 2015,(1250-1254)

[6] A. T. Clements, I. Ahmad, M. Vilayannur, and J. Li. Decentralized Deduplication in SAN Cluster File Systems. In USENIX ATC'09, Jun. 2009.

[7] L. Costa, S. Al-Kiswany, R. Lopes, and M. Ripeanu. Assessing Data Deduplication trade-offs from an Energy Perspective. In ERSS'11, Jul. 2011.

[8] A. El-Shimi, R. Kalach, A. Kumar, A. Oltean, J. Li, and S. Sengupta. Primary Data Deduplication - Large Scale Study and System Design.In USENIX ATC'12, Jun. 2012.

[9] FIU traces. http://iotta.snia.org/traces/390.

[10] D. Frey, A. Kermarrec, and K. Kloudas. Probabilistic Deduplication for Cluster-Based Storage Systems.In SOCC'12, Nov. 2012.

[11] M. Fu, D. Feng, Y. Hua, X. He, Z. Chen, W. Xia, F. Huang, and Q. Liu.Accelerating Restore and Garbage Collection in Deduplication-based Backup Systems via Exploiting Historical Information.In USENIX'14, Jun. 2014.

[12] Garth Gibson. Storage at Exascale: Some Thoughts from Panasas CTO. Interview. May 2011.

[13] B. S. Gill, M. Ko, B. Debnath, and W. Belluomini. STOW: A Spatially and Temporally Optimized Write Cache Algorithm. In USENIX ATC'09, Jun. 2009.

[14] A. Gupta, R. Pisolkar, B. Urgaonkar, and A. Sivasubramaniam.Leverag- ing Value Locality in Optimizing NAND Flash-based SSDs.In FAST'11, Feb. 2011.

[15] Y. Hu and Q. Yang. DCD - Disk Caching Disk: A New Approach for Boosting I/O Performance. In ISCA'96, May 1996.

[16] Y. Hua and X. Liu.Scheduling Heterogeneous Flows with Delay-aware Deduplication for Avionics Applications. IEEE Transactions on Parallel and Distributed Systems, 23(9):1790–1802, 2012.

[17] K. Jinand and E. L. Miller.The Effectiveness of Deduplication on Virtual Machine Disk Images. In SYSTOR'09, pages 1–12, May 2009.

[18] Stephanie Jones. Online De-duplication in a Log-Structured File System for Primary Storage.Technical Report UCSC-SSRC-11-03, University of California Santa Cruz. May 2011.

[19] S. Kiswany, M. Ripeanu, S. S. Vazhkudai, and A. Gharaibeh. STD- CHK: A Checkpoint Storage System for Desktop Grid Computing. In ICDCS'08, Jun. 2008.

[20] R. Koller and R. Rangaswami. I/O Deduplication: Utilizing Content Similarity to Improve I/O Performance. In FAST'10, pages 1–14, Feb. 2010.

BIBLIOGRAPHY

Mrs.Jayanthi Narayanan studying M.phil in the Department of Computer Science, Sri JayendraSaraswathyMahaVidyalaya College of Arts and Science, Singanallur, Coimbatore. She completed MCA in Madurai Kamaraj University. She is interested in data mining. She3 has more than 17yrs experience in school teaching.

Mrs.M.Parveen Taj working as Associate Professor in the Department of Computer Science, Sri JayendraSaraswathyMahaVidyalaya College of Arts and Science, Singanallur, Coimbatore. She has 11.5 years of teaching experience. Her area of interest Networking.

All Rights Reserved © 2017 IJARTET