International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

Data Deduplication with Encrypted Parameters Using Big Data and Cloud

Mohd Akbar, K. E. Balachandrudu, Prasadu Peddi

Abstract: ***Distributed computing empowers organizations to devour a figure asset, for example, a virtual machine (VM), stockpiling or an application, as a utility simply like power as opposed to building and keep up registering frameworks in house. In distributed computing, the most significant part is server farm, where client's/client's information is put away. In server farms, the information may be transferred various time or information can be hacked along these lines, while utilizing the cloud benefits the information should be encoded and put away. With the consistent and exponential increment of the quantity of clients and the size of their information, information deduplication turns out to be increasingly more a need for distributed storage suppliers. By putting away a one of a kind duplicate of copy information, cloud suppliers significantly decrease their stockpiling and information move costs. As a result of the approved information holders who get the symmetric the encoded information can likewise be safely gotten to. Keys utilized for unscrambling of information. The outcomes demonstrate the predominant productivity and viability of the plan for huge information deduplication in distributed storage. Assess its exhibition dependent on broad examination and PC re-enactments with the assistance of logs caught at the hour of deduplication.***

Keywords: big information, distributed computing, information deduplication, intermediary re-encryption.

I. INTRODUCTION:

Distributed computing offers another method for Information Technology benefits by revamping different assets (e.g., capacity, processing) and giving them to client's dependent on their requests. Cloud clients transfer individual or classified substance information to a cloud specialist organization (CSP) server farm and permit keeping this information. With the possibly vast extra room offered by cloud suppliers, clients will in general use as a lot of room as they can and sellers continually search for methods expected to limit repetitive information and expand space reserve funds. A strategy which has been broadly embraced is cross-client deduplication. Deduplication has demonstrated to accomplish high space and cost investment funds and many distributed storage suppliers are at present receiving it.

Revised Manuscript Received on November 30, 2019.
***** Correspondence Author
Mohd Akbar*, Research Scholar, Shri JJT University, Rajasthan, akb.mtech@gmail.com
Dr. K. E. Balachandrudu, Principal, MALLA REDDY ENG CLG, HYDERABAD
Dr. Prasadu Peddi, Assistant Professor, Shri JJT University, Rajasthan

Deduplication can lessen capacity needs by up to 90-95 percent for reinforcement applications and up to 68 percent in standard document frameworks.

While the point of deduplication is to identify indistinguishable information portions and store them just once, the aftereffect of encryption is to make two indistinguishable information fragments undefined in the wake of being encrypted. A system which has been proposed to meet these two clashing necessities is concurrent encryption whereby the encryption key is generally the consequence of the hash of the information section.

Objectives

1. To spare distributed storage and save the protection of information holders by proposing a plan to oversee scrambled information stockpiling with deduplication.
2. To demonstrate the security and execution of the plan through examination and reproduction.
3. A Study on Data Auditing and Security in Cloud Computing

II. LITERATUE REVIEW:

Shobana, R et al (2016) Proposes Cloud Computing Secure Framework (CCSF). In this manner CCSF comprises of four sections: 1) Identity Management 2) Intrusion identification and counteractive action framework 3) Data deduplication 4) Secure Cloud Storage. Interruption discovery and aversion are performed physically by system administrators in the current framework. In our proposed design the interruption identification and anticipation is performed naturally by characterizing rules for the significant assaults and alarm the framework automatically. To guarantee information classification the information is put away in a scrambled sort utilizing Advanced Encryption Standard (AES) calculation.

Wu, T. Y (2015) proposed Index Name Servers (INS) to oversee not just document stockpiling, information deduplication, enhanced hub choice, and server burden adjusting, yet additionally record pressure, lump coordinating, ongoing input control, IP data, and occupied level list checking. The primary favourable position of this procedure is that Index Name Servers calculation help to decrease remaining tasks at hand of assets and improve the presentation of framework. INS additionally handles server burden adjusting. – The fundamental burden of this procedure is that scrambled information can't be de-duplicated.

Shweta D. Pochhi (2014) displayed that the information and the Private cloud where the token age will be performed for each document. Before transferring the information or document to open cloud, the customer will send the record to private cloud for token age which is exceptional for each document. Private mists at that point create a hash and a token and send the token to customer.

A framework which accomplishes privacy and empowers square level de-duplication simultaneously. Before transferring the information or record to open cloud, the customer will send the document to private cloud for token age which is extraordinary for each record.

Puzio. P et al (2013) have proposed ClouDedup security framework to give secure and proficient stockpiling administration which guarantees square level deduplication and information privacy simultaneously. The security of ClouDedup depends on its new engineering with metadata administrator and an extra server. The server adds an extra encryption layer to anticipate surely understood assaults against concurrent encryption and in this way ensure the classification of the information; then again, the metadata director is capable of the key administration

III. SYSTEMARCHITECTURE:

In late time, there are numerous issues of capacity puts in cloud. On the off chance that information holder store document in cloud which is as of now accessible in cloud. The security of ClouDedup depends on its new engineering whereby notwithstanding the essential stockpiling supplier, a metadata administrator and an extra server are characterized: the server adds an extra encryption layer to forestall surely understood assaults against focalized encryption and along these lines ensure the secrecy of the information; then again, the metadata director is mindful of the key administration task since square level deduplication requires the remembrance of an enormous number of keys. Along these lines, the fundamental deduplication is performed at square level and we characterize an effective key administration system to maintain a strategic distance from clients to store one key for each square.

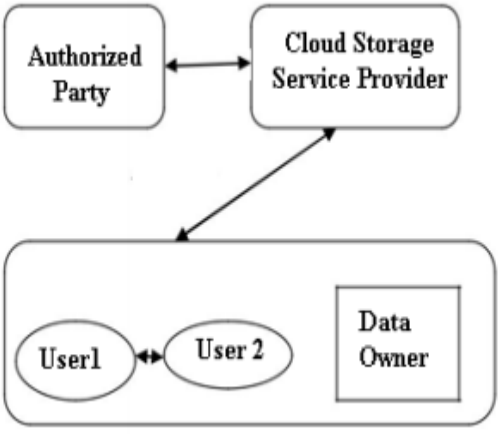

Figure 1: System architecture

Information Holder: The information holder can transfer and spares their information and documents in the CSP. In this framework is conceivable to number of information holders could spare their records in encoded crude information in the CSP. The document is viewed as information proprietor by the information holder that produces or makes the File. The information holder is in ordinary structure than the higher need of proprietor.

Cloud Service Provider: When the information holder erases information from CSP, CSP right off the bat deals with the records of copied information holders by expelling the duplication record of this client. In the event that the rest records are not void, the CSP won't erase the put away scrambled information, however square information access from the holder that solicitations information erasure. On the off chance that the rest records are unfilled, the encoded information ought to be expelled at CSP.

Encoded Data Update: on the off chance that that DEK is refreshed by an information proprietor with DEK 0 and the new scrambled crude information is given to CSP to substitute old stockpiling for the explanation of accomplishing better security, CSP issues the new re-scrambled DEK 0 to all information holders with the help of AP.

Information Owner Management: on the off chance that that a genuine information proprietor transfers the information later than the information holder, the CSP can figure out how to spare the information encoded by the genuine information proprietor at the cloud with the proprietor created DEK and later on, AP underpins re-encryption of DEK at CSP for qualified information holders.

Information DEDUPLICATION: Data deduplication or Single Instancing basically alludes to the disposal of repetitive information. As the measure of computerized data is expanding exponentially, there is a need to convey capacity frameworks that can deal with and deal with this data proficiently. Information deduplication is one of the developing strategies that can be utilized to advance the utilization of existing extra room to store a lot of information. Fundamentally, information deduplication is evacuation of repetitive information. Along these lines, diminishing the measure of information lessens a great deal of costs stockpiling prerequisites costs, framework the executives cost.

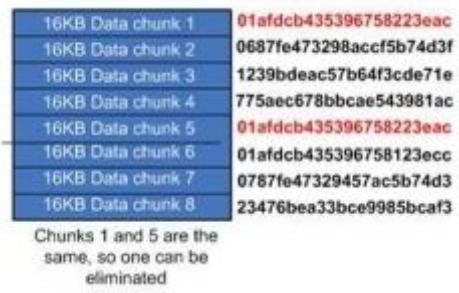

Figure 2: Deduplication in the cloud

IMPLEMENTATION:

For execution we favoured ASP.NET C# language, Visual studio system and Windows O.S. Stage as it gives inbuilt server called IIS. ASP.NET gives inbuilt MSDN oversaw code to help cryptographic hashing calculation expected to perform encryption and decryption. Data deduplication is alluded to as a procedure offered to distributed storage suppliers (CSPs) to dispose of the copy information and keep just a solitary exceptional duplicate of it for extra room sparing purpose. Data deduplication is one of the methods which used to unravel the redundancy of information. The deduplication systems are commonly utilized in the cloud server for diminishing the space of the server. Cloud Storage normally contains business-basic information and procedures; subsequently high security is the main answer for hold solid trust connection between the cloud clients and cloud specialist co-ops.

In this procedure we need to recognize the copy duplicate of the record any kind of document can be distinguish record .txt,.doc,.xls, .ppt, .pdf. so we need to begin with transferring the record when we transfer the document we need to separate initial 50 bytes from the document and last 50 bytes from the document

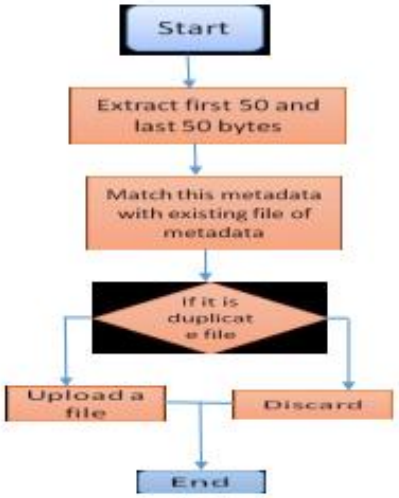

Figure 3: Flow Chart

IV. RESULTS:

Efficiency of information encryption and unscrambling. In this trial, we tried the activity time of information encryption and decoding with AES by applying diverse AES key sizes (128 bits, 196 bits and 256 bits) and different data size (from 10 megabytes to 600 megabytes). we saw that even when the information is as large as 600 MB, the encryption/unscrambling time is under 13 seconds if applying 256-piece AESkey. Applying symmetric encryption for information assurance is a reasonable and pragmatic decision. The time spent on AESencryption and decoding is expanded with the size of data. This is inescapable in any encryption plans. SinceAES is very efficient on information encryption and decryption, thus it is pragmatic to be applied for enormous information.

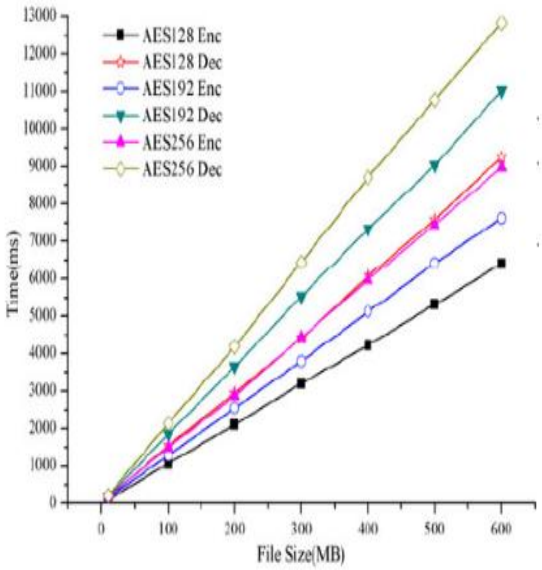

Figure 4: Operation time of file encryption and decryption with AES

Data Ownership Challenge:

In this test, we chose 192-piece field of elliptic curve (160-piece ECC has a security level practically identical to 1024-piece RSA), 256-piece AES, 1024-piece PRE and 10M transferred data. We can see that information transfer is the most tedious if the file is huge, yet it is unavoidable in all schemes. Therefore, our plan can spare a great deal of calculation load and correspondence cost for cloud clients. Moreover, the data proprietorship challenge in the proposed plan is very lightweight, which doesn't include a lot of weight to cloud users.

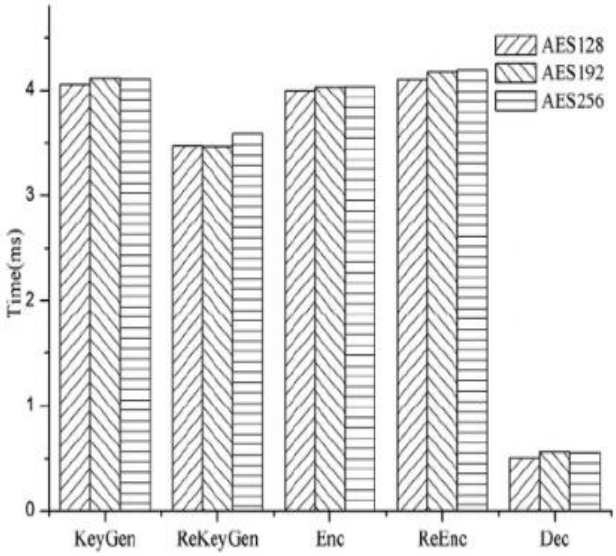

Figure 5: The execution time of PRE operations.

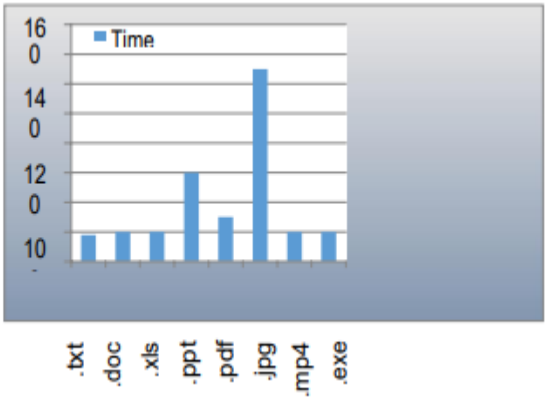

Figure 6: Deduplication time factor at Byte level

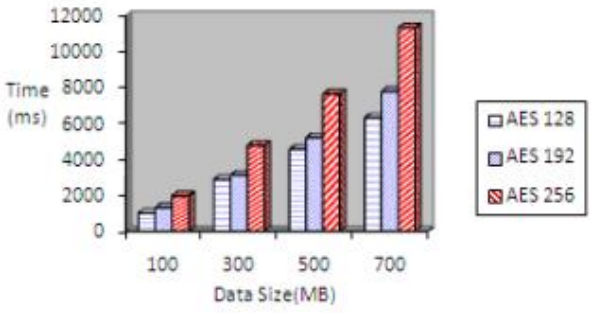

Figure 7: Data encryption and decryption efficiency (Period Analysis)

In this analysis, the time taken by different encryptions standard of AES is illustrated. It was discovered that more noteworthy the bit size the encryption sets aside considerably more effort to encode file. Achieving this measure of uniqueness and speed settled on AES our first decision significantly for getting lower encryption - decoding time when contrasted with others. This will exceptionally decrease the measure of time spent in the whole procedure when gigantic information documents are to be considered.

V. CONCLUSION:

We have proposed another framework called deduplication which spares extra room of cloud empowering to lessen the information duplication by sparing just single duplicate of information for various clients and furthermore furnishes security mechanism. Managing scrambled information with deduplication is significant and huge practically speaking for accomplishing a fruitful distributed storage administration, particularly for enormous information storage. Our plan can deftly bolster information update and imparting to deduplication in any event, when the information holders are disconnected. Scrambled information can be safely gotten to in light of the fact that solitary approved information holders can acquire the symmetric keys utilized for information decryption. To secure the secrecy of touchy information during deduplication, the united encryption procedure is utilized to encode the information before outsourcing. The aftereffects of our PC re-enactmentsfurther showed the practicability of our plan. Future work includes optimizing our plan and execution for practical deployment and considering verifiable calculation to ensure that CSP carries on true to form in deduplication the executives.

REFERENCES:

1. Shobana, R., K. ShanthaShalini, S. Leelavathy V. Sridevi (2016), “De-Duplication Of Data In Cloud”, Int. J. Chem. Sci., ISSN: 0972-768X, Volume: 14, Issue: 4, PP: 2933-2938
2. Wu, T. Y (2015) J. S. Pan, and C. F. Lin, Improving accessing efficiency of cloud storage using deduplication and feedback schemes, IEEE Syst. J., Volume: 8, Issue: 3, PP: 1-10
3. Shweta D. Pochhi, Prof.Pradnya V. Kasture (2014), “Encrypted Data Storage with De-Duplication Approach on Twin Cloud”, International Journal of Innovative Research in Computer and Communication Engineering, Volume: 3, Issue: 2, PP: 22-31
4. Puzio.P., R. Molva, M. Onen, and S. Loureiro (2013), “ClouDedup: Secure deduplication with encrypted data for cloud storage,” in Proc. IEEE Int. Cof. Cloud Comput. Technol. Sci., Volume: 5, Issue: 3, PP: 363–370.
5. Deepak Mishra, Dr. Sanjeev Sharma, “Comprehensive study of data de-duplication”, International Conference on Cloud, Big Data and Trust, Nov 2013.
6. Zhang, D., Liao, C., Yan, W., Tao, R., & Zheng, W. (2017, August). Data Deduplication Based on Hadoop. In Advanced Cloud and Big Data (CBD), 2017 Fifth International Conference, PP: 147-152.
7. Ajay Jangra, Vandna Bhatia, UpasanaLakhinazand, NiharikaSinghx (2015), “An Efficient Storage Framework design for Cloud Computing: Deploying Compression on De-duplicated No-SQL DB using HDFS” 2015 1st International Conference on Next Generation Computing Technologies, Dehradun, India, 4-5 September 2015
8. Wang, C., Z. Qin, J. Peng, and J. Wang (2010), “A novel encryption scheme for data deduplication system,” Proc. International Conference on Communications, Circuits and Systems (ICCCAS), PP: 265–269.
9. YAN, Zheng; DING, Wenxiu; YU, Xixun; ZHU, Haiqi; and DENG, Robert H (2016), “Deduplication on encrypted big data in cloud”, IEEE Transactions on Big Data, Volume: 2, Issue: 2, PP: 138-150.
10. Liu, C. Y., X. J. Liu, L. Wan (2013), “Policy-based deduplication insecure cloud storage,” in Proc. Trustworthy Comput. Serv., Volume: 8, Issue: 3, PP: 250–262

AUTHOR DETAILS:

My name is Mohd. Akbar. I am a Research Scholar from Shri Jagadishprasad Jhabharmal Tibrewala University, Rajasthan. I have completed my M. Tech in Computer Science and Engineering from Jawaharlal Nehru Technological University, Hyderabad. I am having more than 17 years of experience (including overseas) in Teaching field. My research area of interest is Big Data, Cloud Computing. Other areas of interest are Computer Networks, Artificial Intelligence, Machine Learning. I taught several subjects such as Databases, Programming Languages, Operating Systems, Computer Networks, etc.