IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified

Vol. 6, Issue 5, May 2017

Data Deduplication on cloud

Manisha Galphade[1], Roshan Kunal[2], Mayuri Ingulkar[3], Safiya shaikh[4], Mamta Chavan[5]

Professor, Computer Dept, SIT Lonavala, Pune, India [1]

Student, Computer Dept., SIT Lonavala, Pune, India [2,3,4,5]

Abstract: De-duplication is the path toward choosing all classes of information inside an enlightening gathering that suggest a comparable certified life/world component. The data amassed from various resources may have quality issues in it. The thought to perceive duplicates by using windowing and blocking system. The objective is to finish better precision, extraordinary viability and besides to diminish the false positive rate all are according to the surveyed comparable qualities of records. De-duplication is a property which gives additional information of comparable qualities between the two substances. In this paper the basic focus is given on right ID of duplicates in the database by applying thought of windowing and blocking. The objective is to achieve better precision, awesome capability and moreover to reduce the false positive rate all are according to the assessed similarities of records.

Keywords: Access control, big data, cloud computing, data deduplication, proxy re-encryption.

I. INTRODUCTION

Present day society is a computerized universe. No data or, on the other hand industry applications can get by without this advanced universe. The measure of this computerized universe in 2007 is 281 exabytes and in 2011 [10], it progresses toward becoming 10 times bigger than it was in 2007. The most basic issue is that about a large portion of the advanced universe can't be put away appropriately in time. This is brought about by a few reasons: right off the bat, it is elusive such a major information compartment; furthermore, regardless of the possibility that a major holder can be discovered, it is as yet difficult to oversee such a huge dataset; lastly, for financial reasons, building and keeping up such a gigantic stockpiling framework will cost a great deal of cash. This is especially testing for non-IT areas, for instance, designing and bio-science ventures. This paper presents a deduplication on system, named “DeDu”, which runs on commodity hardware. Deduplication means that the number of the replicas of data that were traditionally duplicated on the cloud should be managed and controlled to decrease the real storage space requested for such duplications

II. RELATED WORK

There are many circulated record frameworks that have been proposed for huge scale data frameworks, which can be dispersed over the Internet and this incorporates variable and non-confided in associates. Every one of these frameworks need to endure visit arrangement changes. Venti [22] is a system stockpiling framework. It utilizes one of a kind hash qualities to distinguish piece substance so it diminishes the information control of storage room. Venti manufactures obstructs for mass stockpiling applications and authorizes a compose once strategy to maintain a strategic distance from decimation of information. This system stockpiling framework risen in the early phases of system stockpiling, so it is not reasonable to manage mass information, and the framework is not versatile

DeDe [2] is a piece level deduplication bunch document framework without unified coordination. In the DeDe framework, each host makes content synopses then the hosts trade content outlines with a specific end goal to share the file and recover duplications occasionally and autonomously. These deduplication exercises don't happen at the document level, and the aftereffects of deduplication are not exact.

HYDRAstor [7] is an adaptable, optional capacity arrangement, which incorporates a back-end comprising of a matrix of capacity hubs with a decentralized hash list, and a conventional record framework interface as a front-end. The back-end of HYDRAstor depends on Directed Acyclic Graph, which can sort out extensive scale, variable-measure, content tended to, permanent, and exceptionally versatile information squares. HYDRAstor identifies duplications as per the hash table. A definitive focus of this approach is to frame a reinforcement framework. It doesn't consider the circumstance when different clients need to share documents.

III. IMPLEMENTATION

We are providing registration and login to user. User will register by filling all details. User will get login credentials.. Login of data owner and key authorizer is also provided. File will be uploaded after login of key authorizer as shown in fig 1.

DOI10.17148/IJARCCE.2017.6560

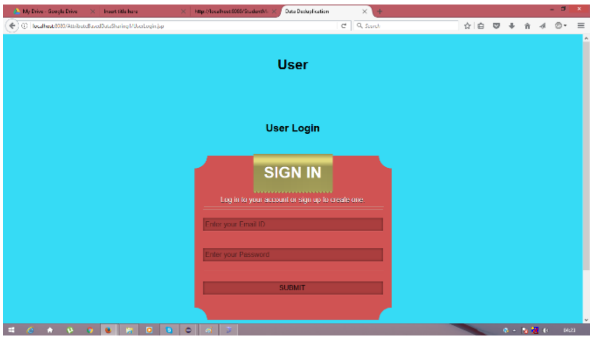
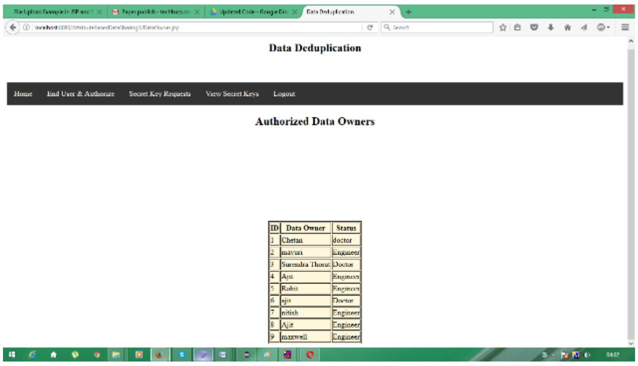


Fig. 1

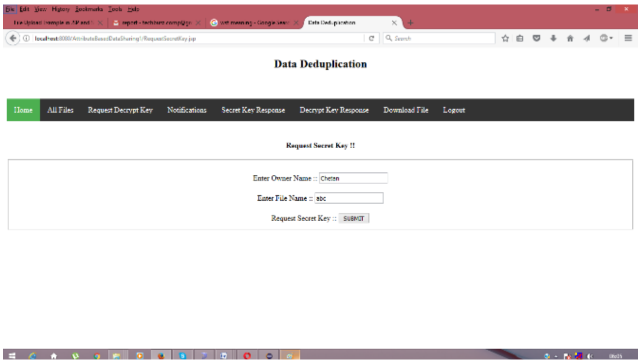
Data owner key view request for key as shown in fig. 2. User can view all files which selected for encryption. User will request for encrypt key and decrypt key as shown in fig 3(A) and 3(B).



Authorized Data Owners

ID	Data Owner	Status
1	Chetan	doctor
2	mayuri	Engineer
3	Surendra Thorat	Doctor
4	Ajit	Engineer
5	Rohit	Engineer
6	ajit	Doctor
7	nitish	Engineer
8	Ajit	Engineer
9	maxwell	Engineer

Fig. 2



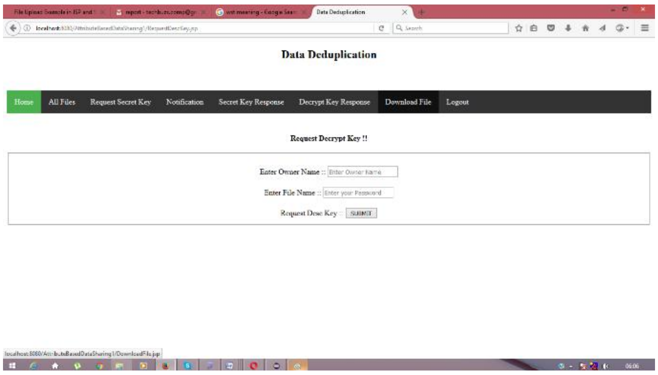
Request Secret Key !!

Enter Owner Name :: Chetan

Enter File Name :: abc

Request Secret Key :: SUBMIT

Fig 3(a)



Request Decrypt Key !!

Enter Owner Name :: Enter Owner Name

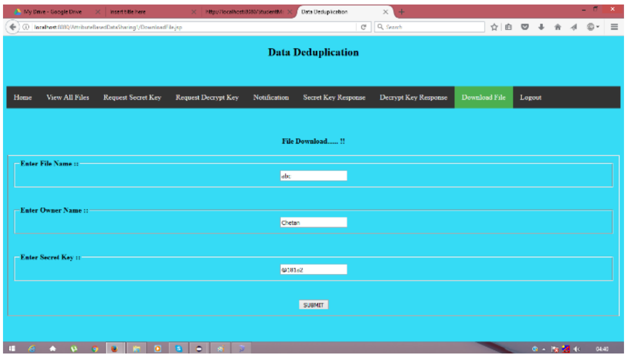
Enter File Name :: Enter your Password

Request Desc Key :: SUBMIT

Fig 3(b)

User can download file by entering file name, owner name and secret key as shown in fig 4

DOI10.17148/IJARCCE.2017.6560



Data Deduplication

File Download...... !!

Enter File Name :: abc

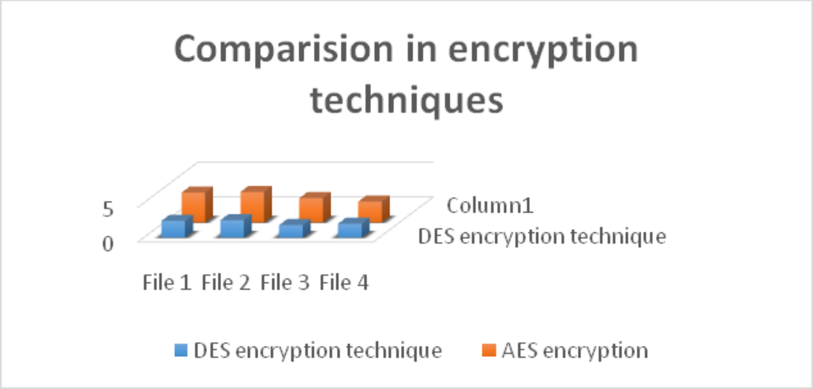
Enter Owner Name :: Chetan

Enter Secret Key :: @181a2

Fig. 4

IV. RESULT AND ANALYSIS

We are comparing encryption techniques. Files are encrypted using AES techniques. DES and AES encryption techniques are compared. Result is as shown in graphical presentation



V. CONCLUSION

Overseeing encrypted information with deduplication is essential and noteworthy practically speaking for accomplishing an effective distributed storage benefit, particularly for huge information stockpiling. In this paper, One of the component is, information is in encrypted shape so protection of client is kept up. we proposed a reasonable plan to deal with the encrypted huge information in cloud with deduplication in view of possession test and PRE. Our plan can adaptably bolster information refresh and offering to deduplication notwithstanding when the information holders are disconnected. Encrypted information can be safely gotten to on the grounds that lone approved information holders can get the symmetric keys utilized for information unscrambling. Broad execution investigation and test demonstrated that our plan is secure and proficient under the portrayed security show and extremely reasonable for enormous information deduplication.

REFERENCES

[1] M. Bellare, S. Keelveedhi, and T. Ristenpart, "DupLESS: Server aided encryption for deduplicated storage," in Proc. 22nd USENIX Conf. Secur., 2013, pp. 179–194.

[2] Dropbox, A file-storage and sharing service. (2016). [Online]. Available: http://www.dropbox.com

[3] Google Drive. (2016). [Online]. Available: http://drive.google.com

[4] Mozy, Mozy: A File-storage and Sharing Service. (2016). [Online]. Available: http://mozy.com/

[5] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system," in Proc. IEEE Int. Conf. Distrib. Comput. Syst., 2002, pp. 617–624, doi:10.1109/ICDCS.2002.1022312.

[6] G. Wallace, et al., "Characteristics of backup workloads in production systems," in Proc. USENIX Conf. File Storage Technol., 2012, pp. 1–16.

[7] P. Anderson and L. Zhang, ''Fast and Secure Laptop Backups with Encrypted De-Duplication,'' in Proc. USENIX LISA, 2010, pp. 1-8.

[8] J.R. Douceur, A. Adya, W.J. Bolosky, D. Simon, and M. Theimer, ''Reclaiming Space from Duplicate Files in a Serverless Distributed File System,'' in Proc. ICDCS, 2002, pp. 617-624

[9] D. Harnik, B. Pinkas, and A. Shulman-Peleg, ''Side Channels in Cloud Services: De-duplication in Cloud Storage,'' IEEE Security Privacy, vol. 8, no. 6, pp. 40-47, Nov./Dec. 2010.

[10] S. Halevi, D. Harnik, B. Pinkas, and A. ShulmanPeleg,''Proofs of Ownership in Remote Storage Systems,'' in Proc. ACM Conf. Comput. Commun. Security, Y. Chen, G. Danezis, and V. Shmatikov, Eds., 2011, pp. 491-500.