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

Avoiding Duplication of Encrypted Data using Cloud

Ashweta Magar[1], Trupti Jagtap[2], Pradnya Gaikwad[3], Rashmi Singh[4]

Student, Computer Dept, ISB&M school of Technology, Pune, India [1,2,3,4]

Abstract: De-duplication is the way toward deciding all classes of data inside an informational collection that imply a similar genuine life/world element. The information accumulated from different assets may have quality issues in it. The idea to recognize copies by utilizing windowing and blocking procedure. The goal is to accomplish better exactness, great effectiveness and furthermore to decrease the false positive rate all are as per the assessed similitudes of records. De-duplication is a property which gives extra data of similitudes between the two substances. In this paper the essential concentrate is given on correct ID of copies in the database by applying idea of windowing and blocking. The goal is to accomplish better exactness, great proficiency and furthermore to diminish the false positive rate all are as per the evaluated likenesses of records.

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

I. INTRODUCTION

Data deduplication facilitates capacity necessities and upgrades maintenance Data is flooding the venture. Capacity executives are attempting to deal with a spiraling volume of records, sound, video and pictures, alongside a disturbing expansion of substantial email connections. More stockpiling is regularly not the best answer—stockpiling costs cash and the sheer number of records in the long run troubles the organization's reinforcement and calamity recuperation (DR) arranges. As opposed to discovering approaches to store more information, organizations are swinging to information decrease advances that can store less information. Data deduplication has as of late risen as an imperative piece of any information decrease plot. This article clarifies the essential standards and usage issues of data deduplication and covers a few cases of the innovation at work today. Data deduplication is fundamentally a methods for diminishing storage room. It works by dispensing with repetitive information and guaranteeing that just a single exceptional example of the information is really held on capacity media, for example, plate or tape. Excess information is supplanted with a pointer to the special information duplicate. Data deduplication, now and again called wise pressure or single-occasion stockpiling, is frequently utilized as a part of conjunction with different types of data diminishment.

II. RELATED WORK

Numerous associations gather a lot of information to bolster their business and basic leadership forms [1]. The information gathered from different sources may have information quality issues in it. These sorts of issues end up noticeably unmistakable when different databases are incorporated. The coordinated databases acquire the information quality issues that were available in the source database. The information in the incorporated frameworks should be cleaned for legitimate basic leadership. Purifying of information is a standout amongst the most critical strides. In this examination, concentrate is on one of the real issue of information purging i.e. "copy record recognition" which emerges when the information is gathered from different sources. Thus of this exploration think about, examination among standard copy end calculation (SDE), sorted neighborhood calculation (SNA), copy end sorted neighborhood calculation (DE-SNA), and versatile copy location calculation (ADD) is given. A model is additionally created which demonstrates that versatile copy location calculation is the ideal answer for the issue of copy record discovery. For rough coordinating of information records, string coordinating calculations (recursive calculation with word base and recursive calculation with character base) have been executed and it is presumed that the outcomes are greatly improved with recursive calculation with word base.

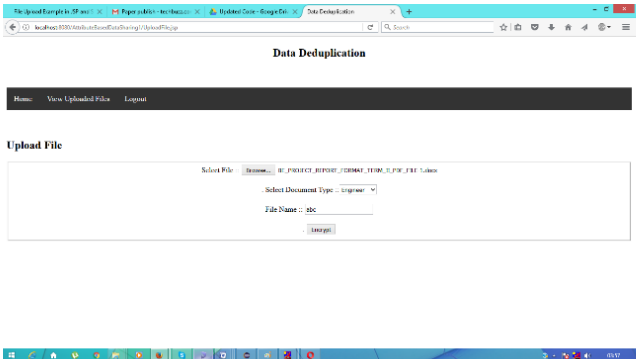
Regularly, in this present reality, substances have at least two portrayals in databases. Copy records don't share a typical key as well as they contain blunders that make copy coordinating a troublesome undertaking.[2] Blunders are presented as the consequence of translation mistakes, deficient data, absence of standard configurations, or any mix of these variables. In this paper, we display an exhaustive investigation of the writing on copy record identification. We cover likeness measurements that are ordinarily used to identify comparative field sections, and we exhibit a broad arrangement of copy identification calculations that can recognize roughly copy records in a database.

DOI10.17148/IJARCCE.2017.6559

Despite the fact that there is a long profession on recognizing copies in social information, just a couple of arrangements concentrate on copy identification in more unpredictable progressive structures, as XML information. In this paper, a novel strategy is exhibited for XML copy location, called XMLDup. XMLDup utilizes a Bayesian system to decide the likelihood of two XML components being copies, considering the data inside the components, as well as the way that data is organized. Furthermore, to enhance the proficiency of the system assessment, a novel pruning procedure, fit for critical increases over the unoptimized variant of the calculation, is displayed. Through examinations, we demonstrate that our calculation can accomplish high exactness and review scores in a few informational indexes. XMLDup is additionally ready to beat another cutting edge copy identification arrangement, both as far as productivity and of viability.

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.


File Upload Example in JSP and S
Paper publish - techbuzz.co
Updated Code - Google Dri
Data Deduplication
localhost:8080/AttributeBasedDataSharing1/UploadFile.jsp
Search
Home
View Uploaded Files
Logout
Browse...
Encrypt

Data Deduplication

Upload File

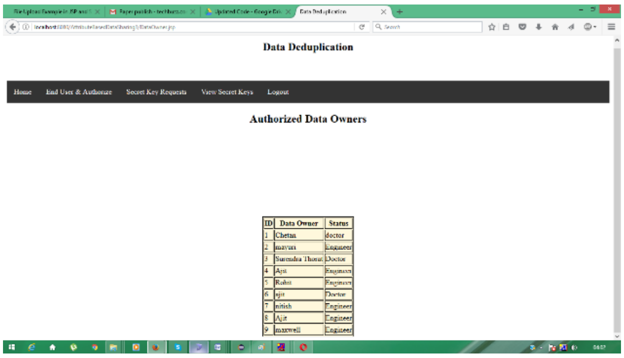
Select File :: BE_PROJECT_REPORT_FORMAT_TERM_II_PDF_FILE 1.docx

Select Document Type :: Engineer

File Name :: abc

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).

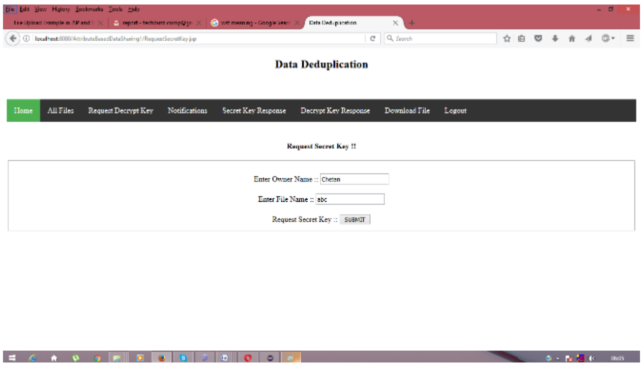

File Upload Example in JSP and S
Paper publish - techbuzz.co
Updated Code - Google Dri
Data Deduplication
localhost:8080/AttributeBasedDataSharing1/DataOwner.jsp
Search
Home
End User & Authorize
Secret Key Requests
View Secret Keys
Logout

Data Deduplication

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



Data Deduplication

Request Secret Key !!

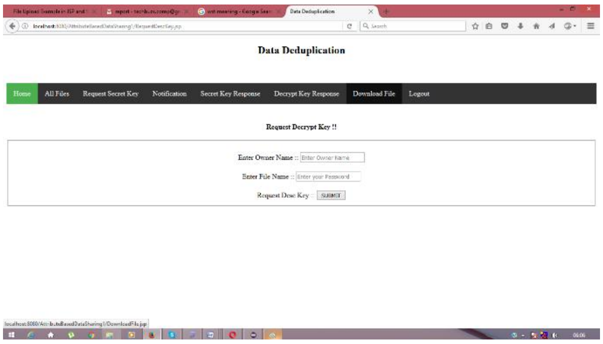
Enter Owner Name :: Chetan

Enter File Name :: abc

Request Secret Key ::

Fig 3(a)

DOI10.17148/IJARCCE.2017.6559



Data Deduplication

Request Decrypt Key !!

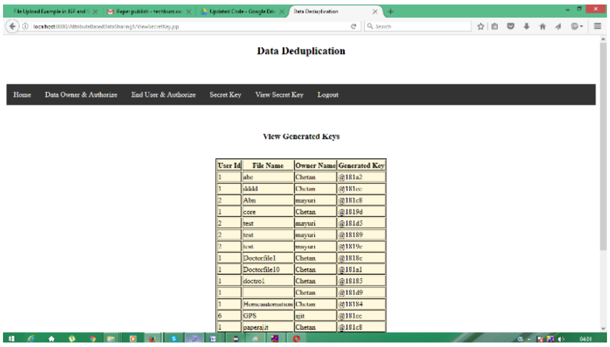
Enter Owner Name :: Enter Owner Name

Enter File Name :: Enter your Password

Request Decr Key :: SUBMIT

Fig 3(b)

Data owner can view request of received from users as shown in fig 4.



Data Deduplication

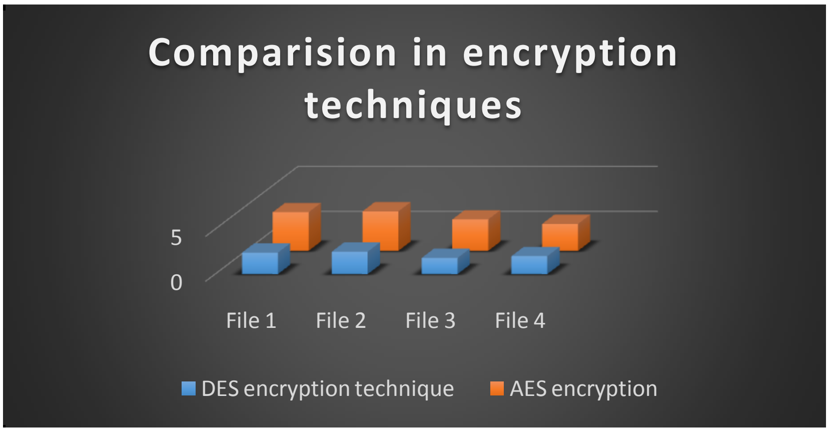
View Generated Keys

User Id	File Name	Owner Name	Generated Key
1	abc	Chetan	@181a2
1	dddd	Chetan	@181cc
2	Abn	mayuri	@181c8
1	core	Chetan	@1819d
2	test	mayuri	@181d5
2	test	mayuri	@18189
2	test	mayuri	@1819c
1	Doctorfile1	Chetan	@1818c
1	Doctorfile10	Chetan	@181a1
1	doctro1	Chetan	@18185
1		Chetan	@181d9
1	Homeautomation	Chetan	@18184
6	GPS	ajit	@181cc
1	paperajit	Chetan	@181c8

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.

DOI10.17148/IJARCCE.2017.6559

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.
[11] M. Mulazzani, S. Schrittwieser, M. Leithner, M. Huber, and E. Weippl, ''Dark Clouds on the Horizon: Using Cloud Storage as Attack Vector and Online Slack Space,'' in Proc. USENIX Security, 2011, p. 5.

DOI10.17148/IJARCCE.2017.6559