

Cloud Deduplication

D. T. Shreyaswi¹, Arhath Kumar²

¹Student, Department of MCA, NMAM Institute of Technology, Karkala, India

²Professor, Department of MCA, NMAM Institute of Technology, Karkala, India

Abstract: Cloud deduplication describes a new supplement, consumption and delivery model for IT services based on internet protocols. It typically involves provisioning of dynamically scalable and often virtualized resources. Cloud collects the computer resources. Moreover, it supplies a management process and at the same time, it provides services for trillions of users.

Keywords: deduplication, USENIX and CBIR,

1. Introduction

Storage and management are major reviews in this time of big data. The capacity for storing the devices to meet the rate at which it grows, magnify access time and transferring of data. Cloud Deduplication works to help the organization dealing with strong redundant operations that needs constant copying and storing of information for future reference. This technique is a part of backup and catastrophe recovery solution. It allows enterprises to store data continuously and gives fast, cost-effective and reliable data redemption.

2. Literature survey

[1] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010. System is currently based on client-side deduplication using whole file hashing. Hashing process is performed at the client, and connects to any one of Deduplicators according to their loads at that time. The deduplicator then identifies the duplication by comparing with the existing hash values in Metadata Server.

[2] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server aided Encryption for deduplicated storage. In USENIX Security Symposium, 2013. In this paper, they have CBIR, which is known as Content based Image Retrieval is the application of computer vision technique to the image retrieval problem from the large databases. In this paper de-duplication of electricity bills is done through CBIR. The technique for comparing the images is Block truncation coding; it is a lossy compression technique for grayscale image.

[3] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin Clouds: An architecture for secure cloud computing. In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011. Here they propose the image deduplication cloud storage system. To protect the confidentiality of sensitive image content, the convergent encryption has been used while supporting image deduplication. Owner could download the

cipher text again and retrieve the image with secret key, as the same time, image owner makes use of attribute-based encryption scheme to share images with friends by setting the access privileges.

3. Problem description

Cloud deduplication is used to eliminate the duplicate data from the different users. The problem occurs when the data confidentiality in the encryption is incompatible with cloud deduplication. So these similar data of different users may lead to cryptograph texts it makes deduplication impossible. The convergent encryption technique has been proposed to encrypt the data before outsourcing. The users having different privileges are examined in duplicate test beside the data itself.

4. Product description

Cloud collects the computer resources. Moreover, it supplies a management process and at the same time, it provides services for trillions of users. The problem is it stores the similar data of different user so this may effects the storage space. The solution for this problem is cloud deduplication it compares the data and eliminates the similar data. This reduces the storage space. Cloud deduplication is a by-product and consequence of approach to eliminate variety of computing sites which is provided by the internet according to the user needs. It enhances the security and protects the data confidentiality. Cloud deduplication helps in encrypting the text files.

5. Screenshots

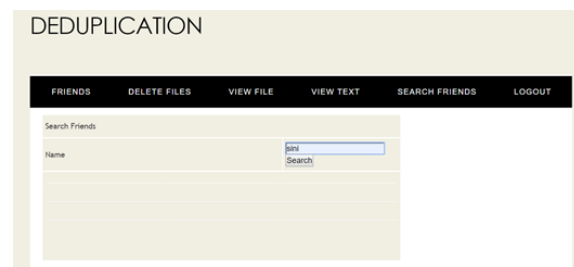




Fig. 1. Dashboard

6. Conclusion

Cloud deduplication technique compresses the data by removing the similar copies of identical data. Based on the selection of cloud deduplication and the architecture of proportion, the backup environment as a great deal can be impacted. Thus this project helps in benefiting us in storing the data and eliminating unwanted data. Mainly it minimizes the storage space.

References

- [1] P. Anderson and L. Zhang, "Fast and secure laptop backups with encrypted de-duplication," in Proc. of USENIX LISA, 2010.
- [2] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Server aided Encryption for deduplicated storage," in USENIX Security Symposium, 2013.
- [3] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider, "Twin Clouds: An architecture for secure cloud computing," in Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.