iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 5 **Issue:** IV **Month of publication:** April 2017

**DOI:** http://doi.org/10.22214/ijraset.2017.4245

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

# Brief Review of De Duplication over Cloud for Optimizing Storage

Anju Bala Malhotra[1], Er. Jasbeer Narwal[2]
[1]M.TECH 4th Sem, [2]Lecturer, Department of Computer Science
Haryana Engineering College Jagadhri, Kurukshetra University, Haryana, India

***Abstract:*** ***Cloud storage is one of the services provide in cloud computing which has been increasing in reputation. With the growing data size of cloud computing, a decrease in data volumes could help provider reducing the costs of running large storage system and saving energy use. So data de-duplication techniques have been brought to recover storage competence in cloud storages. In this paper, we suggest a dynamic de-duplication system for cloud storage, which aim to advance storage competence and maintain redundancy for fault acceptance.***
***Keywords: Cloud computing, De-duplication, Cloud Storage,***

## I. INTRODUCTION

As data is increasing exponentially, Local storage space has proved to be inappropriately wasted in some manner. The most important aspect of this insufficiency is data redundancy. Data is so redundant that a big percentage of storage space has been dedicated to it. Data is duplicated to such extent that even a single megabyte file sometimes takes gigabytes of space due to multiple duplication. This redundancy or duplication needs to be minimized. Duplication is present in all the levels which are file level, block level etc. Local storage is that model where data can be placed, managed, back up, stored and modified. Local storage makes available data to clients' quick time, with high storage space and also makes it user friendly so that availability of data increases. Local storage is done mainly to backup the data. Local storage is that model where data can be placed, managed, back up, stored and modified. Local storage makes available data to clients in any time, with high storage space and also makes it user friendly so that availability of data increases. Data de duplication in local Storage leads to saving storage space. The optimization of backup storage by removing redundant data is known as de–duplication. Data de-duplication not only reduces the storage space requirements by eliminating redundant data but also minimizes the network transmission of duplicate data in the network storage systems. Deduplication can be done at many levels viz file level, chunk level and block level.

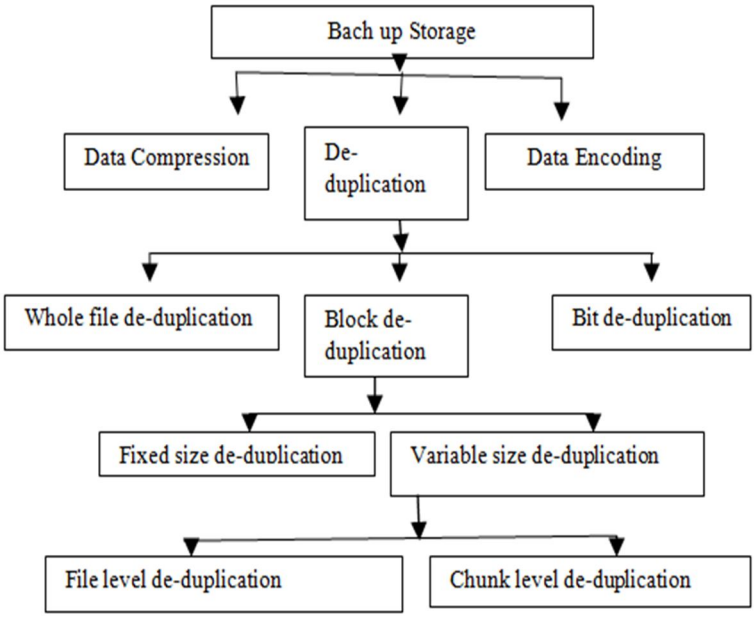



Fig 1. Approached to de-duplication


©IJRASET: All Rights are Reserved

*A. Benefits Of Deduplication*

1) *Increased Storage*: The exponential increase in volume of data with time has lead the cloud providers to present massive amount of infrastructure. The maintenance of that gigantic amount of data is also a very important aspect that cloud services does for the users. The deduplicated data, thus gets reduced and can be easily managed.
2) *Cost Reduction*: The cost associated always stays at lower side with lesser amount of data to maintain. This helps in cutting heavy costs for cloud service providers. The cloud providers only have to maintain that is unique and not duplicate.

## II. LITERATURE REVIEW

Yinjin Fu, et. al [1] introduced a method for de duplication that optimizes the performance of search. This method used in personal environment and to reduce overhead. Earlier methods were focusing solely on removing redundancy. But this technique targeted on quicker retrieval. Fatema Rashid et al. [2] proposed a framework that enables the file to be divided first into smaller units and then those blocks gets encrypted using an effective encryption technique however their framework wasn't tested in real cloud environment, AvaniWildaniet. al [3] demonstrated the effectiveness of their approach employing a straightforward neighborhood grouping that needs solely timestamp and block number, that makes it better to be used with various kind of storage systems without the need to modify host file systems.Dirk Meister, et.al [4] projected a way within which earlier backup information was used to predict the future backup. This technique increase the lookup performance. Pasqualepuzioet. al [5] proposedClouDedup, a secure and efficient storage service that assured block-level deduplication and data confidentiality at an equivalent time. Even whenbased on convergent encryption, ClouDedup remained secure because of a component that implemented an additional encryption operation and an access control mechanism. Also as the requirement for deduplication at block-level raises an issue with respect to key management, they advised to include a brand new part so as to implement the key management for every block together with the actual deduplication operation. They showed that the overhead introduced by these new elements was minimal and failed to impact the overall storage and computational costs.Dongfang Zhao, et. al [6] proposed system that is based on Hy Cache+. This caching approach showed 29X speedup over the standard traditional LRU algorithmic rule. De duplication on primary storage system.

## III. CONCLUSION

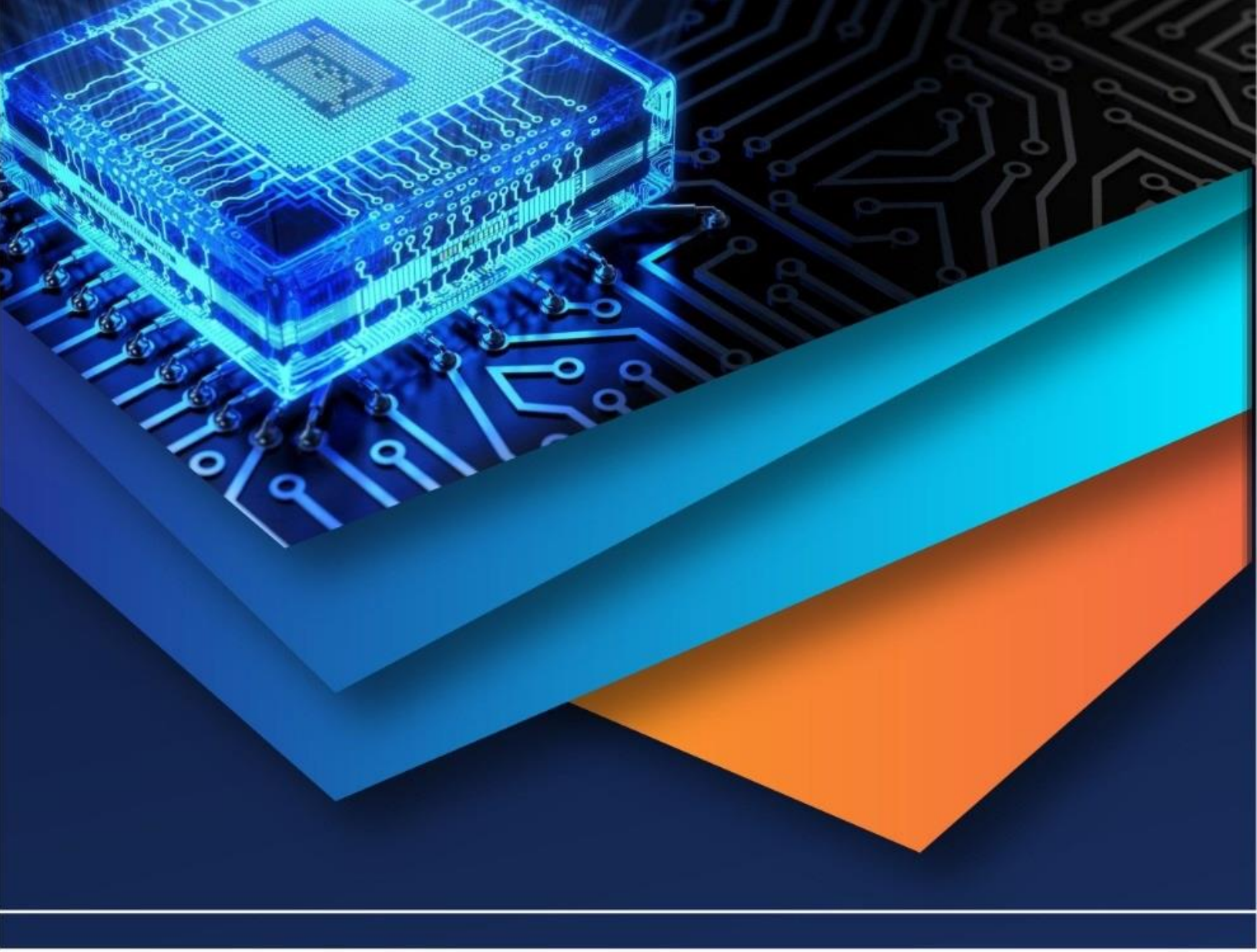
Improvement in MD5 algorithm used by hash function to generate hash value or unique ID is the main focus of this research proposal. The accuracy is measured in terms of matching errors that consist of false rejection rate and false acceptance rate. In this way, Accuracy to find duplicate data is increased. The implementation shall involve offline development on Microsoft visual studio. The platform shall be .net. This research work shall improve the accuracy of fingerprint generation algorithm for hash function. The storage space must be used in the best possible way. The optimized use of cloud storage needs to have some interface for deduplication of data before storing it to the actual database somewhere. The futuristic scope of this paper focuses on the implementation of a better way of recognition of unique data and thus the overhead of storage and maintenance of the duplicate data is minimized.

## REFRENCES

[1] Yinjin Fu, Hong Jiang, Nong Xiao, Lei Tian, Fang Liu, AA-Dedupe: An Application-Aware Source Deduplication Approach for Cloud Backup Service, IEEE International Conference on Cluster Computinges in the Personal Computing Environment (2011).
[2] Rashid, Fatema, Ali Miri, and Isaac Woungang. A secure data deduplication framework for cloud environments.Privacy, Security and Trust (PST), 2012 Tenth Annual International Conference on.IEEE, 2012.
[3] A. Wildani, E. L. Miller, and O.Rodeh. HANDS: A heuristically arranged non-backup in-line deduplication system. Technical Report UCSCSSRC- 12-03, University of California, Santa Cruz, March 2012.
[4] Dirk Meister, Jürgen Kaiser, Block Locality Caching for Data Deduplication. In Proceedings of the 11th USENIX Conference on File and Storage Technologies (FAST). USENIX, February 2013.
[5] Pasquale Puzio, Melek O¨ nen, Sergio Loureiro, ClouDedup: Secure Deduplication with Encrypted Data for Cloud Storage, IEEE, 2013.



©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)