Volume 6, Issue 5, May 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

De-Duplication Over Cloud Computing for Valuable Storage: A Review

Manreet Kaur
M.Tech (IS) Scholar
Department of Information Technology
CEC Landran, Mohali, Punjab, India

Jaspreet Singh
Assistant Professor
Department of Information Technology
CEC Landran, Mohali, Punjab, India

Abstract: Cloud computing is the emerging technology that helps in consolidation of resources. Many organizations have their public as well as private clouds. Private clouds can be built from unused resources to store data. But as private cloud has limited ways of storage, so they need to be utilized properly.De-duplication is the means of storing data in effective way over clouds. This paper will discuss about the use of De-duplication method din cloud computing for the storage of the data in effective way.

Keywords: Cloud computing, De-duplication and Key technique.

I. INTRODUCTION

Cloud is the large pool of resources. In cloud all resources are virtually connected to each other. These resources can be reconfigured according to storage needs. There are mainly five features that cloud has:

- On-demand self-service
- Broad network access
- Resource pooling
- Measured service
- Rapid elasticity

Cloud computing contains both hardware and application provided to users. Computing resources are limited and will saturate at some time. This scalability alarms the clouds to structure your data accordingly. It means its giving way to compress the data for further scalability. So here comes cloud storage.

Cloud storage is that model where data can be placed, managed, back up, stored and modified. Cloud storage makes available data to clients in any time, with high storage space [1] and also makes it user friendly so that availability of data increases. There are many companies that help to store data in cloud server online. There is a need of online interface when user wants to store data online. Cloud storage is done mainly to back up the data. Cloud storage is a SLA services i.e. it is Service Level Agreement. Cloud storage is of three types: Public, Private and hybrid.

Advantages of Cloud Storage

- Availability to access the data from any place/ location increases.
- No need to carry physical storage device.
- Other trusted people can also allow sharing the data.

II. LITERATURE SURVEY

Yinjin Fu, et.al [2] introduced a technique for de duplication that will optimize the performance of look up. This technique has been used for personal environment and to reduce overhead. Earlier methods are focusing only on removing redundancy. But this method focused on faster time retrieval.

AvaniWildani et. al [3] demonstrated the effectiveness of our approach using a simple neighborhood grouping that requires only timestamp and block number, making it suitable for a wide range of storage systems without the need to modify host file systems.

Don fang Zhao, et.al [4] proposed system that is based on Hy Cache. It provides the transparency to exchange of data, to modification of data. This caching advance shows 29X speedup over the conventional LRU algorithm. De - duplication on primary storage system.

Puzio et.al [5] proposes Clouded up, a secure and well-organized storage service which assures block-level de-duplication and data privacy at the same time. Although based on convergent encryption, Clouded up remains secure thanks to the definition of a component that execute an additional encryption operation and an access control mechanism. Furthermore, as the requirement for de-duplication at block-level raises an issue with respect to key management, we propose including a new component in order to execute the key management for each block together with the actual de-duplication operation. We show that the in the clouds introduced by these new components is minimal and does not impact the overall storage space and computational costs.

Dirk Meister, et.al [6] proposed a method in which earlier backup in sequence is used to predict the future backup. These methods enhance the lookup performance. It is better than BLC approach.

III. APPROACHES TO DEDUPLICATION

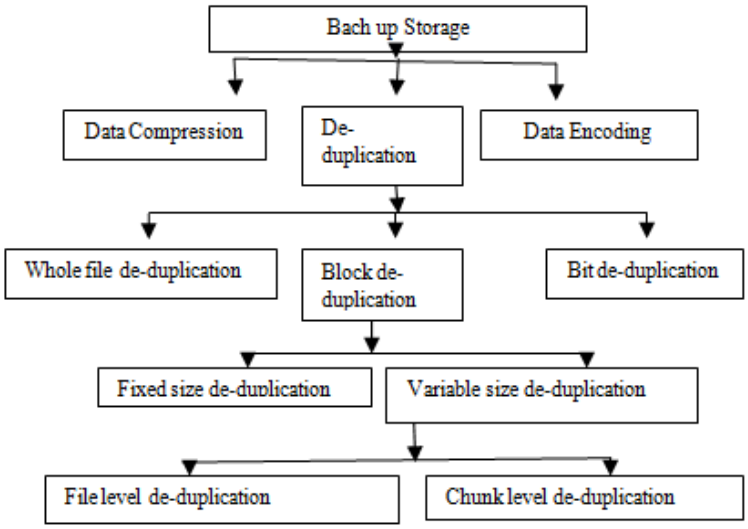

Fig. 1 Approaches to de-duplication

Data de replication describes a class of approaches that diminish the storage capacity needed to store data or the amount of data that has to be transfer over a network. These approaches detect coarse-grained redundancies within a data set, e.g. a file system; Data de duplication not only reduces the storage space requirements by eliminate redundant data but also minimizes the network transmission of duplicate data in the network storage systems. It splits files into numerous chunks that are each uniquely identified by a hash signature called a fingerprint. It removes duplicate chunks by checking their fingerprints, which avoids byte by byte comparisons. Mainly data de duplication listening carefully on different terms like throughput, advance chunking schemes, other type of storage capacity and cluster method and system workload.

IV. CONCLUSION AND FUTURE SCOPE

In this paper, i have discussed about storage issues in the cloud computing and how de-duplication method solves the difficulty of storage at cloud. One way of storage space is using hash functions but they are to effective approach so there is an enhancement needed in which use of external hard disk is required.

REFERENCES

[1] Rashid, Fatema, Ali Miri, and Isaac Woungang. "A secure data de-duplication framework for cloud environments." *Privacy, Security and Trust (PST), 2012 Tenth Annual International Conference on*. IEEE, 2012.

[2] Yinjin Fu, Hong Jiang, Nong Xiao, Lei Tian, Fang Liu,''A De-duplication: An Application-Aware Source De-duplication Approach for Cloud Backup Service "IEEE International Conference on Cluster Computing in the Personal Computing Environment (2011).

[3] A. Wildani, E. L. Miller, and O.Rodeh. HANDS: A heuristically arranged non-backup in-line de-duplication system. Technical Report UCSCSSRC- 12-03, University of California, Santa Cruz, March 2012.

[4] Dirk Meister, Jorgen Kaiser," Block Locality Caching for Data De-duplication". In Proceedings of the 11th USENIX Conference on File and Storage Technologies (FAST). USENIX, February 2013.

[5] Dongfang Zhao, KanQiao, IoanRaic, y, "HyCache: Towards Scalable High-Performance Caching Middleware for Parallel File Systems", Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357, 2014.

[6] Pasquale Puzio, Melek O¨ nen, Sergio Flourier, "Clouded up: Secure De-duplication with Encrypted Data for Cloud Storage", IEEE, 2013

