International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 05 | May-2018 www.irjet.net p-ISSN: 2395-0072

An optimization method to remove replication in cloud storage

NIKHIL TIWARI

Computer Science & Engineering Dr. Apj Abdul Kalam Technical University Uttar Pradesh, India

---

Abstract - *Clouds are the bigger pools of memory, easy to use and access virtualized properties. Cloud computing presents an on demand, self-service, fast, acceptable and universal access to several computing entities and resources. Data de - replication is information-compression techniques which reduces the storage consumed by eliminating duplicate copies of information.*

In the proposed study we implemented SHA-3 algorithm to generate the Hash Value in the form of string. K-mean clustering algorithm used for divide the data into the hash values. Genetic algorithm used for optimize to find the best solutions from the hash values for unique data. The parameters on which the comparison is done are memory consumption, hashing time, detection time and detection accuracy in the proposed method the times saving is less in comparison to previous approach. It also calculated on different file sizes and different types of operation like upload update and delete in cloud environment. The results are calculated as working with five and 10 replicas as different number of files. Accuracy of current approach is used to measure working probability of algorithm. The probability will be high means accuracy than it will be gives better results to their users. The accuracy will be less than it uploads files multiple times on the server and also overwrites some file during execution.

Key Words: virtualization, measure, Accuracy of current approach, overwriting, hashing time, memory consumption etc.

1. INTRODUCTION

Clouds are the bigger pools of memory easy to use and access imaginary properties. These properties can be vigorously re-configured to alter to a variable scale, permitting optimum resource utilization Cloud computing includes on demand deployment, virtualization, internet delivery of service and open source. Cloud computing is a technology that uses the internet and centre remote server to maintain applications and data. Cloud computing allows customers and businesses to use applications without access to their individual files at any computers with internet access.

Cloud computing could be an innovation that uses the web and focal remote servers watch out to keep up the information and functions. Cloud computing licenses buyers and in addition organizations to utilize functions while not establishment and contact their own documents at any PC by web contact. This innovation licenses for a lot of capable computing by integrative information stockpiling, procedure and information measure. A clear case of cloud computing is Yahoo email, Gmail, or Hotmail and so on. All you would like is essentially a web alliance and you'll have the capacity to start sending electronic mail. The principle server and in addition electronic mail administration programming framework is all on the cloud (web) and is totally overseen by the cloud administration dealer Yahoo, Google and so on the purchaser gets the chance to utilize the bundle alone and delight in the preferences.

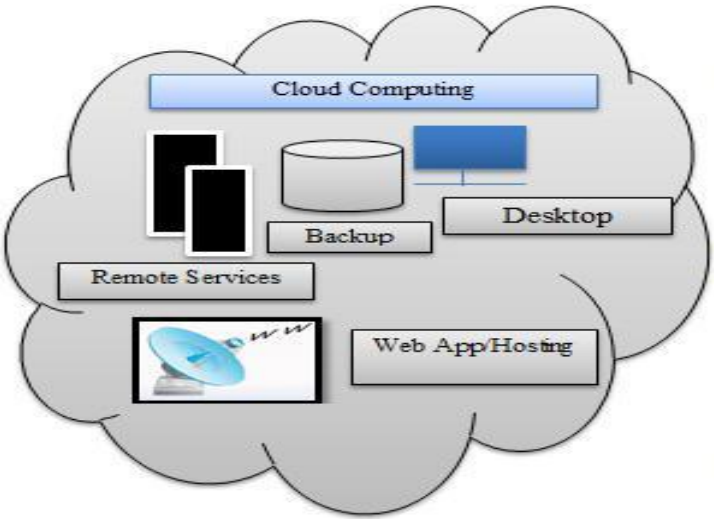

Fig -1: cloud computing use cloud

1.1 Service Models

Cloud administrations proposes that administrations made offered to clients on interest by means of the web from a cloud computing supplier's servers as against for being given from an organization's own particular on-premises servers. Cloud offices are planned to make basic, access to functions, resources and also benefits, and are totally overseen by a cloud services merchant. Delineations of cloud offices typify on-line information stockpiling and reinforcement clarifications, Net-based electronic mail courtesies, suited working environment gatherings and record organization administrations, information procedure, fulfilled methodological support administrations thus on. It is mainly accessed through service oriented architectures based on web service technologies and a web portal

1.2 Cloud Deployment Models

Cloud computing, frequently mentioned to as unjust ―the cloud,‖ is the release of on-demand computing possessions the lot from submission to information Centre in excess of the Internet resting on a pay-for-use basis .There are four

primary cloud computing deployment models which are available to service consumer.

1.2.1 The Private Cloud

This model doesn't acquire much terms of expense productivity: It is equivalent to purchasing, building and dealing with your own base. Still, it acquires enormous quality from a security perspective. These concerns are dealt with by this model, in which facilitating is constructed and kept up for a particular customer. For example, SOX, HIPAA, or SAS 70, which may oblige information to be overseen for protection and reviews that administer the corporation.

1.2.3 The Public Cloud

The public cloud deployment model speaks to genuine cloud facilitating. In this deployment model, administrations and framework are given to different customers. Google is a case of an open cloud. This administration can be given by a merchant gratis or on the premise of a pay-every client permit policy. This model is most appropriate for business prerequisites

1.2.3 The Hybrid Cloud

This deployment helps organizations to exploit secured applications and information facilitating on a private cloud, while as yet appreciating money saving advantages by keeping shared information.

2. DE-replication

Data DE-replication or data DE-duplication is information - compression techniques which reduces the storage capacity by eliminating duplicate copies of information or reduce the sum of information that has to be transfer over a complex.

Data DE-duplication – Also identified as Single Case Storage. Data DE-duplication not simply reduce the storeroom gap necessities by eliminating redundant information but minimizes the system broadcast of photocopy information in the system storeroom system. It is a means of dropping storage wants by eliminate disused statistics the optimization of storage is identified as DE –duplication storing space. Only one exclusive illustration of the information is really retain on storage medium

Data DE-duplication done at client side and server side. In client side DE-duplication is done previous to distribution the information to a storage space device. Only unique information is transferred to the mechanism with the minimum available band measurement and it requires less time. At server side DE-duplication is done after sending the information to storage scheme. D-duplication is as well used in back up services to reduce network bandwidth.

3. Techniques of DE-replication

3.1 Hashing Technique

In hashing technique hashing the information means creating a hash value or number of the file, block and byte which guarantee to be unique for all the above types. In hashing technique some hashing algorithms are used. These hashing algorithms have their own properties like their output size, block size, rounds and performance. Hashing

technique is used after uploading the file. When fingerprint of the file is generated then it is stored in the metadata and used for the comparison purpose. From the above block diagram the file to be uploaded is fed to the hashing method which generates the hash worth. The hash value is compared with already existing hash values. If a matching hash value is found the particular file will not be added to the cloud storage, else server will store the file.

If two files have same hash value then it is said to be a similar file otherwise it is considered as a different file. By using this technique storage space is reduced and time is also saved to find the duplicate files. Searching of files is also easy when we have customized information storage.

4. Hashing Algorithms

Hashing the information means creating a mathematical form of a specific information set that is unique for all information sets. These hashing algorithms generate hash value or fingerprint by using some steps. Some of hashing Algorithm discussed below in a table

Table -1: Sample Table format

Algorithms	O/P(bits)	Internal state size	Rounds
WHIRLPOOL	512	512	10
PANAMA	256	8736	64
Havel	160	256	160
MD2	128	384	864
MD5	128	128	64
SHA-0	160	160	80
SHA-1	160	160	80
SHA-2	256	256	64

2. **SHA(Secure Hash Algorithm)**

Different kinds of SHA:-

SHA 0

SHA 1

SHA 256

SHA 512

SHA 3

The safe Hash Algorithm is 1 of a figure of cryptographic hash purpose. There are at present 3 production of safe Hash method:

SHA-1 is the unique 160-bit hash purpose. The alike to the previous MD5 method. SHA-2 is a relative of 2 alike hash functions, with dissimilar obstruct sizes, recognized as SHA-256 and SHA-512. They are different in the express size; SHA-256 uses 32-bit words where SHA-512 uses 64-bit words.

SHA-3 is a prospect hash purpose criterion still in expansion.

3. CONCLUSIONS

Cloud is the costly storage provider, so the motivation is to use its storage area efficiently DE-replication has been proved to reduce memory consumption by removing the useless duplicate files. So far from the previous studies file level DE-replication is the better approach to be used, the focus of the proposed work will be on file level DE-duplication

In this research work, a dynamic information DE-duplication method for shade storage; in direct to fulfill stability between varying storage effectiveness and mistake tolerance desires, and also to pick up presentation in cloud storage systems. A lot of research has been carried out over this by means on hashing algorithm. From the previous hashing algorithms Hybrid optimization approach will perform better than MD5 and other SHA Techniques.

We achieved the best performance with Hybrid optimization by combining the SHA-3, Genetic algorithm and also with k-mean algorithm used for optimize the best solution in the string values. Then evaluate the performance parameters like time detection, accuracy and memory consumption.

ACKNOWLEDGEMENT

I express my sincere gratitude to the Dr.APJ Kalam Technical University, UP. I generally very grateful to my Senior Ms. Rohini sharma (Mtech) who provide her knowledge and support in completion of my thesis.

REFERENCES

[1] D. Kornack and P. Rakic, "Cell Proliferation without Neurogenesis in Adult Primate Neocortex," Science, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467.

[2] V. Andrikopoulos, Z. Song, and F. Leymann, —Supporting the migration of applications to the cloud through a decision support system,|| IEEE Int. Conf. Cloud Comput. CLOUD, pp. 565–572, 2013.

[3] Haitao Li, LiliZhong, Jiangchuan Li, , Bo Li, KeXu, — Cost-effective Partial Migration of

[4] VoD Services toContent Clouds||, 2011 IEEE 4th International Conference on Cloud Computing, pp. 203-110, 2011.

[5] C. Ward, N. Aravamudan, K. Bhattacharya, K. Cheng, R. Filepp, R. Kearney, B. Peterson,

[6] L. Shwartz, C. C. Young, —Workload Migration into Clouds – Challenges, Experiences, Opportunities||, 2010 IEEE 3rd International Conference on Cloud Computing, pp. 164-171, 2010.

[7] Kangkang Li, HuanyangZheng, & JieWu . —Migration-based Virtual Machine Placement in Cloud Systems||, 2013 IEEE 2nd International Conference on Cloud Networking (CloudNet, IEEE, pp. 83-90, 2013.

[8] Jianxin Li, Yu Jia a, Lu Liub, TianyuWoa, — CyberLiveApp: A secure sharing & migration approach for live virtual desktop functions in a cloud environment", Elsevier, Vol. 29, pp.334-340, 2013.

[9] R. Maggiani, "Cloud computing is changing how we communicate," 2009 IEEE International Professional Communication Conference, IPCC 2009,Waikiki, HI, United states ,pp 1, July 2009.

[10] Rohini Sharma and Amritpal Kaur, —ISSUES & BENEFITS OF DATA DE-DUPLICATION IN CLOUD COMPUTING,|| vol. 1, no. 8, pp. 46–50, 2016. International Journal of Engineering Applied Sciences and Technology, 2016