

A Survey Paper on Authorized De-duplication of Encrypted Data in Cloud

Nikumbh Bhagyashri¹, Ghuge Jayashri², Ubale Shubhangi³, Gursal Prabodhini⁴,

Prof. Chandgude A.S⁵

¹⁻⁴Student, Dept. of computer Engineering, SND College of Engineering And Research Centre, Nashik, Maharashtra, India

Abstract - The cloud computing is a on-demand network access service to a shared pool of configurable computing resources like servers, applications and storage. These shared resources can be rapidly provisioned to the consumers on the basis of only paying for whatever they use. Cloud storage refers to the de- livery of storage resources to the consumers over the Internet. Private cloud storage is restricted to a particular organization and data security risks are high in cloud storage. Therefore, private cloud storage is built by exploiting the commodity machines within the organization and It stored the important data in it. When the utilization of such private cloud storage increases, there will be an increase in the performance storage demand, privacy security, data integrity. Increase in storage demand leads to the expansion of the cloud storage with additional storage nodes. The storage nodes in the cloud storage need to be balanced in terms of load, during such a expansion.

The key idea behind this is to develop a dynamic load balancing algorithm based on de-duplication to balance the load across the storage nodes during the expansion of private cloud storage. For maintaining privacy, security data integrity we will use AES algorithm SHA algorithm. SHA algorithm also used to avoid duplication.

Key Words: Cloud computing, cloud solution, reliability, load balancing, encryption secure de-duplication, data integrity.

1. INTRODUCTION

Cloud computing is a on-demand network access service to a shared pool of configurable computing resources like servers, applications and storage. These shared resources can be rapidly provisioned to the consumers on the basis of only paying for whatever they use. Cloud storage refers to the de- livery of storage resources to the consumers over the Internet. Private cloud storage is restricted to a particular organization and data security risks are high in cloud storage. Therefore, the private cloud storage is built by exploiting the commodity machines within the

organization and it stored the important data in it. When the utilization of such private cloud storage increases, there will be an increase in the connection problem, performance storage demand, privacy security, data integrity. For connection problem we will implement offline data store sync mechanism. Increase in storage demand leads to the expansion of the cloud storage with additional storage nodes.

1.1. Problem Statement

When Cloud is capable of handling 3 request at a time so if any external request is arise then there will be connection reliability problem arise the Storage of cloud is full then the Data Management problem Occurs Then the data on cloud is visible/access by any other user then Privacy And Security related problem When the Performance of system is Decreased then Performance related problem occurs.

1.2. Motivation

Main motivation of the system is to remove a load on cloud base servers and avoiding data Duplication's using the some methodologies and algorithm. This system is basically perform on Hash Code detection techniques which is used for avoiding multiple storage of the files on the Cloud Server .For the load balancing techniques system split the file into three chunks and stored into the three different location and the access is only for the valid person's or authorized persons only who has login credentials with valid user key which is provided by the admin at the time of registration.

1.3 .Goals & Objective

1. This system is to store data offline when unavailability of network & sync on reconnection.
- 2 .The system is to removing a load on cloud base servers using the load balancing methodologies and algorithm.
3. This system is basically perform on Hash Code detection techniques which is avoiding the multiple storage of the same files on the Cloud Server
4. The Login credential gets match then the all of three chunks gets merged into a single

file and Delete/Download Operations performed this makes the faster and more secure.

If connection is unavailable then use offline store & sync mechanism.

Algorithm/ Methodology

Advanced Encryption Standard (AES):

AES calculation is utilized to encode the information. AES includes three square figures, AES-128, AES-192 and AES-256. Each figure encodes and unscrambles information in squares of 128 bits utilizing cryptographic keys of 128-, 192-and 256-bits, separately.(Rijndael was intended to deal with extra square sizes and key lengths, however the usefulness was not embraced in AES.) Symmetric or mystery key figures utilize a similar key for scrambling and unscrambling, so both the sender and the recipient must know and utilize a similar mystery key. All key lengths are regarded adequate to secure grouped data up to the "Mystery" level with "Top Secret" data requiring either 192- or 256-piece key lengths. There are 10 rounds for 128-piece keys, 12 rounds for 192-piece keys, and 14 rounds for 256-piece keys { a round comprises of a few handling steps that include substitution, transposition and mixing of the input plain text and transform it into the final output of cipher text.

Secure Hash Algorithm (SHA):

SHA calculation creates a hash code based on document content. Cryptographic hash capacities are numerical tasks kept running on computerized information; by contrasting the processed "hash" (the yield from execution of the calculation) to a known and expected hash esteem, an individual can decide the information trustworthiness. For instance, figuring the hash of a downloaded document and contrasting the outcome with a recently distributed hash result can demonstrate whether the download has been adjusted or altered. A key part of cryptographic hash capacities is their impact obstruction: no one ought to have the capacity to and two diverse info esteems that outcome in a similar hash yield.

Advantages

1. Efficient storage Allocation
2. Cost Saving
3. Network Optimization
4. Fat Recovery & Continuity
5. Data Center Efficiency

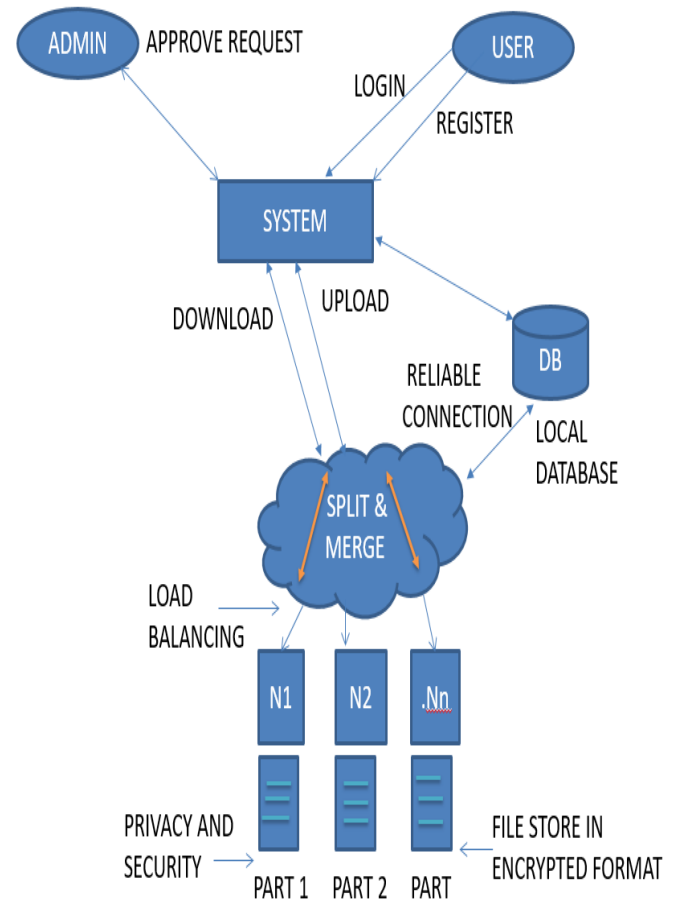


Fig 1: Architectural Design

Application

1. Allows enterprises to manage applications or workload demands by allocating resources among multiple computers or networks.
2. Provide single internet service from multiple servers, sometimes known as server farm.

CONCLUSION

To avoid cloud disaster such as problems in connection, performance, privacy & security, data management we are implementing this framework To overcome this problems.

REFERENCES

[1] J. Wu, L. Ping, X. Ge, Y. Wang, and J. Fu, Cloud storage as the infrastructure of cloud computing, in Proc. 2010 Int. Conf. Intell. Comput. Cognitive Inform. (ICICCI), KualaLumpur, 2010, pp.380- 383.

[2] J. Gantz and D. Reinsel, The digital universe decade-Are you ready, IDC White Paper, <http://www.emc.com/collateral/analyst-reports/idc-digitaluniverse-are-you-ready.pdf>, 2010.

[3] P. Xie, Survey on de-duplication techniques for storage systems, *Comput. Sci.*, vol. 41, no. 1, pp. 22-30, Jan. 2014.

[4] R. Hu, Y. Li, and Y. Zhang, Adaptive Resource Management in PaaS Platform Using Feedback Control LRU Algorithm, *International Conference on Cloud and Service Computing*, 2011.

[5] C. S. Pawar, and R. B. Wagh, Priority Based Dynamic Resource Allocation in Cloud Computing with Modified Waiting Queue, *2013 International Conference on Intelligent Systems and Signal Processing (ISSP)*, 2013.

[6] Buyya.R.et al., *Market-Oriented Cloud Computing: Vision, Hype and Reality for Delivering it Services as Computing Utilities*, c 2008.

[7] Ghalem Belalem, Said Limam, Fault Tolerant Architecture to Cloud Computing using Adaptive Checkpoint, *International Journal of Cloud Applications and Computing*, 1(4), pp 60-69, 2011.

[8] Malte Schwarzkopf, Derek G. Murray, Steven Hand, *The Seven Deadly Sins of Cloud Computing*, Research University of Cambridge Computer Laboratory.

[9] Sandeep Sharma, Sarabjit Singh, and Meenakshi Sharma, *Performance Analysis of Load Balancing Algorithms*, World Academy of Science, Engineering and Technology, 2008.

[10] R. Buyya et al, *Cloud Computing Principles and Paradigms* Published by John Wiley Sons, Inc., Hoboken, New Jersey.