

# Load Balancing Duplicate Detection

Prof.Miss.S.A.Aher<sup>1</sup>, Bhagat Narendra N.<sup>2</sup>, Chavan Dnyaneshwar R.<sup>3</sup>, Kshirsagar Pankaj B.<sup>4</sup>,  
Shinde Dashrath N.<sup>5</sup>

<sup>1,2,3,4</sup> *Students, Information Technology Department, SVIT Chincholi, Nashik, Maharashtra, India.*

<sup>5</sup> *Professor, Information Technology Department, SVIT Chincholi, Nashik, Maharashtra, India.*

## ABSTRACT

*Cloud computing enables the on-demand network to retrieve to an assigned join of configurable computing basic material a well-known as servers, computerized information and applications. These shared resources can be shortly provisioned to the consumers on the core of paying only for along by all of the others they use. Cloud computerized information refers to the labor of computerized information resources to the consumers from one end to the other the Internet. Private cloud computerized information is isolated to a particular organization and data security risks are few and far between compared to the public cloud storage. Hence, private cloud computerized information is built by exploiting the production machines within the organization and the proper data is collected in it. When the utilization of a well-known private cloud computerized information increases, there will be a rebound in the computerized information demand. It advances the expansion of the cloud computerized information with additional storage nodes. During such expansion, storage nodes in the cloud storage require being normal in restriction of the load. In edict to strengthen the load adjacent several storage nodes, the data wish to be migrated across the storage nodes. This data migration consumer greater network bandwidth. The key tenor behind this complimentary is to materialize a dynamic load balancing algorithm based on deduplication to ensure the load across the storage nodes completely the expansion of private cloud storage.*

**Keyword :** - *Big Databases, Hadoop, Security,SHA Algorithm,AES Algorithm, Dynamic Load Balancing Algorithm.*

---

## 1. INTRODUCTION

Cloud computing enables an on-demand network to retrieve to an assigned hang out with of configurable computing basic material a well-known as servers, computerized information and applications. These shared resources can be instantly provisioned to the consumers on the core of paying only for along mutually others they use. Cloud computerized information refers to the propagation of computerized information resources to the consumers completely the Internet. Private cloud computerized information is isolated to a particular organization and data security risks are the minority compared to the public cloud storage. Hence, private cloud computerized information is built by exploiting the brand machines within the university and the germane data is collected in it. When the hard pull of a well-known private cloud computerized information increases, there will be a rebound in the computerized information demand. It advances the expansion of the cloud computerized information with additional storage nodes. During such high on the hog, storage nodes in the leave in the shade storage have a passion for to be normal in restriction of the load. In sending up the river to uphold the load adjacent several storage nodes, the data require being migrated across the storage nodes. This data migration consumer greater network bandwidth. The key kernel behind this Application is to develop a dynamic load balancing algorithm based on duplication detection to balance the load across the storage nodes completely the expansion of private cloud storage.

## 2. MOTIVATION:

The main motivation of the system is to remove a load on cloud base servers and avoiding data Duplications using some methodologies and algorithm. This system is basically performed on HashCode detection techniques which are used for avoiding multiple storages 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 the valid user key which is given by the admin.

### 3. LITERATURE SURVEY:

In the current cloud server storage techniques there is a less security for the update, Search, delete and download a file. There is less load balancing techniques and duplication detection. The current system has only provided the spaces on the server but not avoid the duplicate files. In this project we are using hash code for the content of the file, if this code is found in the database then system generates a duplicate file message for the users and file will be divided into three chunks which are stored in the three different location so the load will be divided and automatically load balancing happened. Propose a system architecture for inline duplication detection based on existing protocol of the hadoop distributed file system.

### 4. OBJECTIVE:

This System has a functionality to ask information for the customer to the login and send the username, password and private key to the user with the help of the admin. Those have a login credential as well as a private key for the login who can easily perform upload, delete, and download operations. Using the Advanced Encryption standards (AES) and Secure HashCode (SHA) algorithm the data security and load balancing will be managed. The HashCode is used to create code according to the file data and stored in the database if the code is same then Duplicate file message will arrive otherwise the code is unique then file split into three different chunks and stored it in three Different location. If the user tries to Delete or Download the file without Private Key and it is login credential it gets fails. If the Login credential gets match then all of the three chunks gets merged into a single file and Delete/Download Operations performed this makes the faster and more secure

### 5. ALGORITHM:

**Step 1:** Login.

**Step 2:** Client makes request for file uploading.

**Step 3:** After upload data checking duplication in database.

**Step 3:** Dedup Server encrypts file using AES/RSA algorithm and split the inputted file into different chunks and store on file server having low load balance on CPU.

**Step 4:** Dedup Server generates hashcode for every chunk of file.

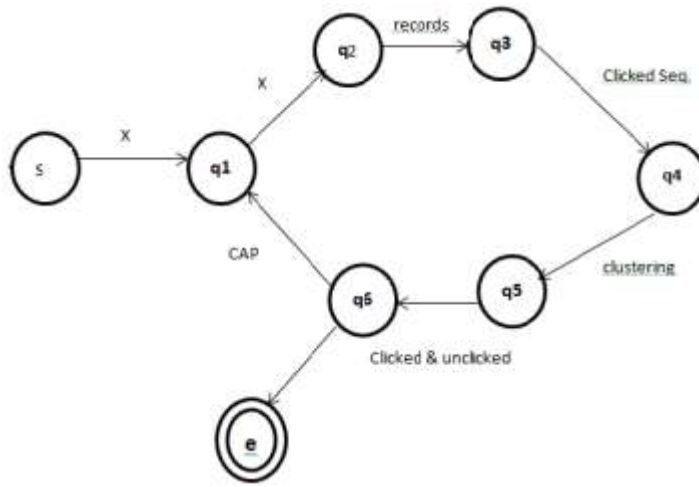
**Step 5:** If hashcode is already existed in database then chunks uploading avoided by dedup server as one copy is already on file server else stored into database.

**Step 6:** Client request for file downloading.

**Step 7:** Dedup Server merge the chunks of requested file and download the file by using AES/RSA algorithm for file decryption.

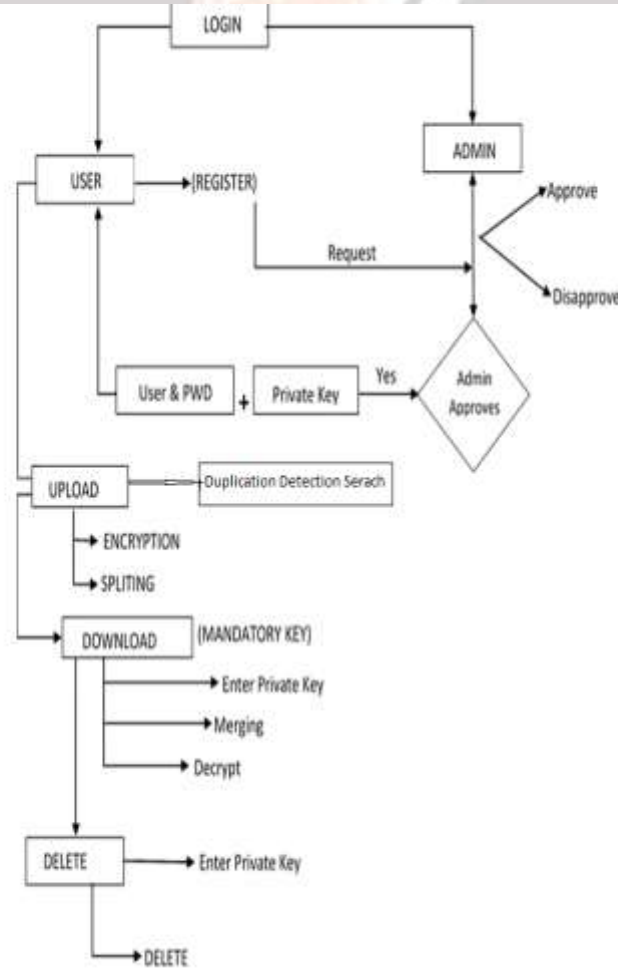
**Step 8:** Stop.

**6 .STATE TRANSITION DIAGRAM:**



**Fig 1.** state transition diagram

**7. PROPOSED ARCHITECTURE DIAGRAM:**



**Fig 2.** System Architecture

## 8. CONCLUSION:

This system proposes the architecture of duplication detection system for cloud storage environment and gives the process of avoiding deduplication in each stage. In Client, system employs the file-level and chunk-level duplication detection to avoid duplication. The algorithm also supports mutual inclusion and exclusion. Load sharing algorithm which is having a policy to partitions the system into various domains and also having a concept of cache manager and information dissemination for the various cloudlets.

## 9. 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. Intel. Compute. Cognitive Inform. (ICICCI), KualaLumpur2010,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 deduplication 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.

