

# Secure Access to Personal Health Record Data Using Hadoop Framework

Pratibha Sajjad<sup>1</sup>, Manoj Challa<sup>2</sup>

<sup>1</sup>Computer Science and Engineering, CMR Institute of Technology, Bangalore, Karnataka, India

<sup>2</sup>HOD & Associate Professor, Information Science and Engineering, CMR Institute of Technology, Bangalore, Karnataka, India

**Abstract:** Hadoop is a free software framework utilized for appropriated storage and circulated handling of the exceptionally tremendous measure of information. HDFS is a Hadoop Distributed File System which gives various key components like adaptable, capable and dependable. It gives the information duplication highlight to keep away from the single purpose of disappointment. The MapReduce system is utilized to handle the information and this information will be partition into blocks these blocks will be prepared parallel and each block will be repeated inside the cluster to maintain a strategic distance from the disappointment. The MD5 encryption is used to store the data in encrypted format. These methods are used to upload/download the PHR data.

**Keywords:** Big Data, Hadoop, Hadoop Distributed File System (HDFS), MD5(Message Digest), MapReduce.

## I. INTRODUCTION

Personal health record (PHR) is a concept that emerges recently. With the PHR concept, an individual can store any of his/her health related information of his/her entire life. The PHR data contains variety of health related information such as patient data, family history, and some laboratory test results[1]. Thus, the PHR data requires a big storage for storing the massive PHR data and the data in the storage must be correct and accessible. In other words, PHR is another application that requires a big data management. Since the PHR belongs to the PHR owner, the access control on the PHR must ensure the protection of data. Moreover, the PHR may contain some sensitive data. Thus, a security mechanism to prevent any data leakage must be implemented on the PHR data.

### A. Related Work

Due to the high cost of building and maintaining specialized data centers, many PHR services are outsourced. The main concern is about whether the patients could really control the sharing of their personal health related data, particularly when they are put away on an outsider server which individuals may not completely trust. The PHR data is stored on a cloud storage. Today healthcare technology can improve quality of lives. Daily life activity information can be a source for predicting a disease or preventing one. Personal health record (PHR) is a concept that emerges recently. The PHR owner can store any health related information into the

PHR storage and the PHR system must ensure that the PHR owner has a full control over his/her data[2].

### B. Proposed System

Hadoop gives a successful and adaptable dispersed plan with express powerful information support to guarantee the rightness of clients' information. Here depend on eradication redressing code in the document conveyance planning to give redundancies and assurance the information reliability. This development definitely lessens the correspondence and capacity overhead when contrasted with the conventional replication-based record appropriation methods. The system focuses on the multiple data owner scenario, and divides the users in the PHR system that greatly reduces the key management complexity for owners and users[2]. It bridges the gaps by proposing a unified security framework for patient-centric sharing of PHRs in a multi-authority PHR system with many users

- In proposed system MapRedce algorithm is used
- MD5 encryption algorithm is used for encrypting the data.

### C. Implementation

Implementation is the phase of the venture when the hypothetical outline is transformed out into a working framework.

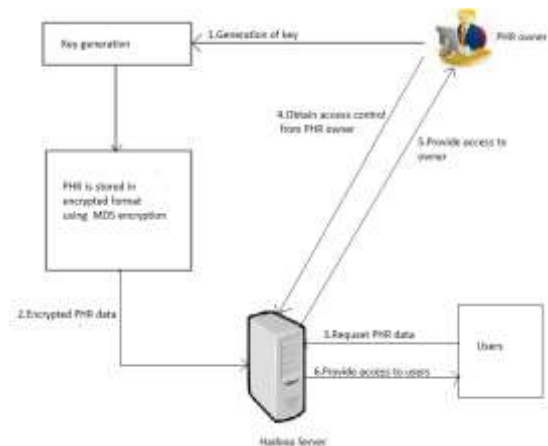


Fig. 1 System architecture

The figure 1 shows the architecture of the of the system. PHR owner is the one who creates their own data. For security purpose generating the key and PHR data is stored in encrypted format using MD5 algorithm. Encrypted data is stored on the hadoop storage. If the user request data then PHR owner should provide access control then users access the data. It contains the following modules.

- 1) *Registration*: In this module ordinary enlistment for the numerous clients. It includes the clients related information.
- 2) *Upload Records*: In this module, clients transfer their records with secure key probabilities. The proprietors transfer scrambled PHR records to the server.
- 3) *Key Generation*: To create secure secret key arbitrary key is utilized. It gives the way to scramble the information. Irregular numbers are valuable for an assortment of purposes, for example, creating information encryption keys.
- 4) *Download Files*: In this module, clients download their records. The client can download decoded PHR documents from the server. In the event that the client fulfilled the get to strategy then just gets the decoded information, generally the encoded information is accessible.

#### D. Hadoop Distributed File System (HDFS)

HDFS Hadoop is splitting the data into smaller pieces. From this it is able to deal with the Big data. So after breaking the data into smaller pieces the Hadoop is doing computation in the following way.[4] Hadoop divides the computation also into smaller pieces and then its forward each slice of computation into each slice of data. So this is called the Map Reduce. Data in a Hadoop cluster is split into the smaller unit called as block and distributed in the entire cluster. HDFS consists of two types of nodes:

**HDFS Namenode** :It stores all metadata such as a location of blocks in HDFS and a number of copies.

**HDFS Datanode** :Datanode holds the actual blocks of data. When the data comes, it will be divided into blocks and that blocks of data will be stored on Datanode, while the information associated with these blocks are stored in the Name node. Where each block of data has 3 copies placed on different Datanode.

#### E. Map and Reduce Algorithm

Map/Reduce is well-liked because it process large volume of data. Map/Reduce boosts to modernize and convert the existing algorithms to Map/Reduce algorithms for the big data. The algorithm is to sort the information set and to change over that into key, value sets to fit with Map/Reduce. It Resolves the challenges to examine tremendous volumes of information in distributed computing

environment. Map/Reduce programming platform is performed in the Apache Hadoop that creates open-source programming for dependable, adaptable, prudent, productive and conveyed processing.

#### F. MD5 Algorithm

The Message Digest it is a restricted cryptographic hash work. MD5 performs numerous binary operations on the "message" to process a 128-bit "hash". To be considered cryptographically secure, the hash capacity ought to meet two prerequisites: to begin with, that it is outlandish for an assailant to produce a message coordinating a particular hash value; and second, that it is unthinkable for an aggressor to make two messages that create a similar hash value[3].

#### G. Results

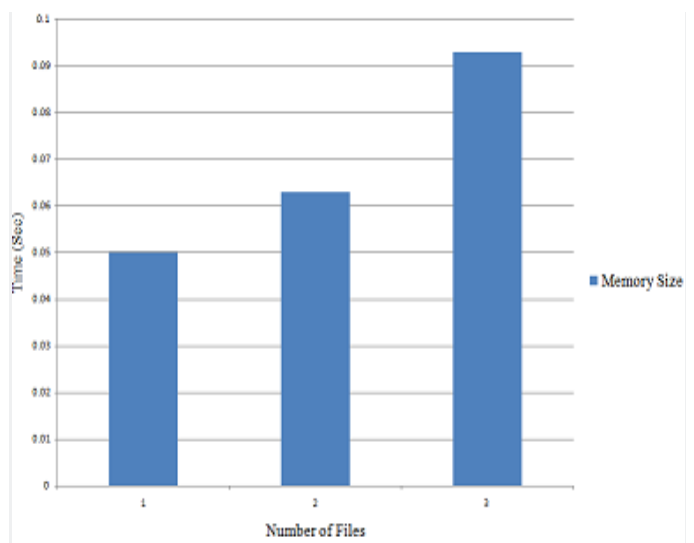


Fig. 2 Uploading time

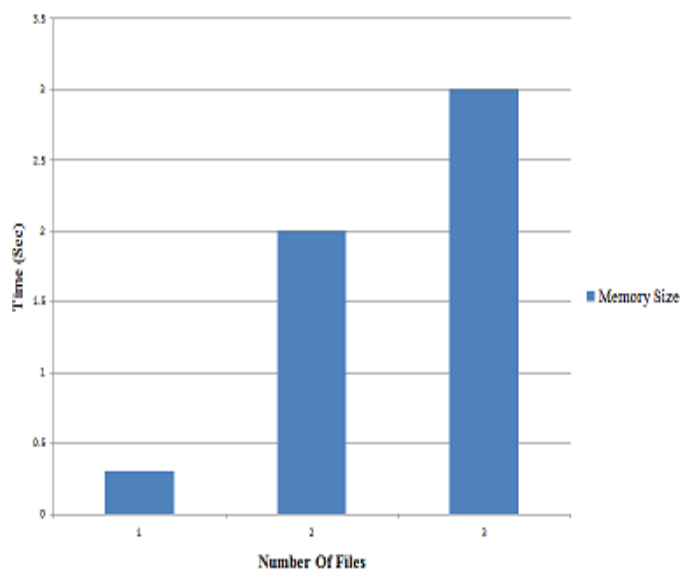


Fig. 3 Retrieving time (existing system)

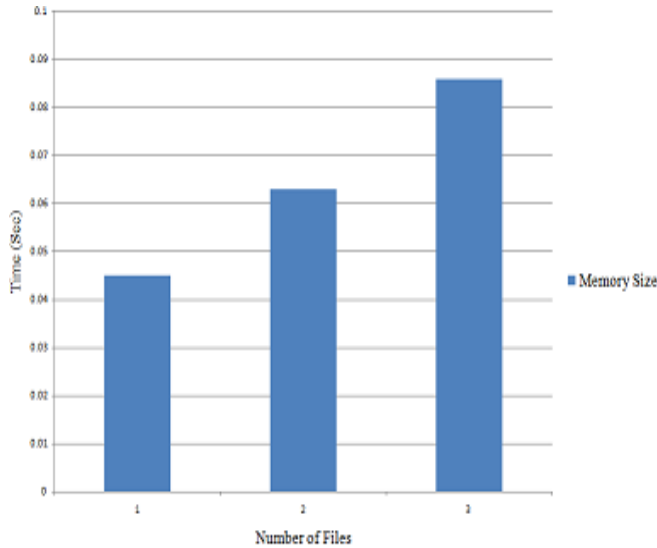


Fig. 4 Retrieval time (proposed system)

## II. CONCLUSION AND FUTURE WORK

Distributed storage design for encrypted personal health record data is presented in this project. The patient health related information is stored in hadoop. HDFS is used for storing the files. The map reduce algorithm is used for

processing the huge amount of data. The data is encrypted before being stored on hadoop storage. The personal health data is accessed based on the PHR owner access policy. In the proposed system performance is improved when multiple users are access the data simultaneously.

In future the spark can be used, it has excellent performance and is highly cost-effective in-memory data processing. Using spark the programs run up to 100x faster than Hadoop MapReduce in memory, or 10x faster on disk.

## ACKNOWLEDGMENT

I would like to express my special gratitude and thanks to my guide for giving me such attention and time.

## REFERENCES

- [1]. Kim, K. Nahm, (2012). Benefits of and barriers to the use of personal health records (PHR) for health management among adults: Online Journal of Nursing Informatics.
- [2]. V.Indhumathi, V.Prakasham, (2015). On Demand Security For Personal Health Record In Cloud Computing: IEEE Sponsored 2nd International .
- [3]. M.Rajasekar ,S. Pathur Nisha , V.S.Thangarasu, (2014). Scalable and Secure Sharing of Personal Health Record Using Enhanced Attribute Based: Indian Journal Of Research.
- [4]. Konstantin Shvachko, Hairong Kuang, Sanjay Radia, Robert Chansler, (2014). Apache Hadoop: California USA.