International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com Volume 4, Issue 2 (March-April, 2016), PP. 302-304

PROVIDING CLASSIFICATION AND SECURITY OF BIG DATA IN CLOUD COMPUTING

Shantanu Kalbhor[1], Hiteshkumar Jain[2], Kaushiki Upadhyay[3]
Department of Computer Engineering
K.C. College Of Engineering Management Studies & Research
Thane, India
Shanty.kalbhor@gmail.com
hiteshjn94@gmail.com

Abstract— **This tutorial gives an overview on state-of-the-art methods for the automatic construction of large knowledge bases and harnessing them for data and text analytics. It covers both big-data methods for building knowledge bases and knowledge bases being assets for big-data applications. The tutorial also points out challenges and research opportunities.**

Index terms- **Big data, apriori algorithm, AES algorithm,**

I. INTRODUCTION

The usage of data has been increasing day by day due to which there is a large cluster of data is available for management which is pretty much unacceptable currently the cloud system is used to manage the data and make it available to the user according to his requirement

II. PROPOSED SYSTEM

The central idea of the project is to store data over internet using cloud computing. As the data stored over the cloud is not safe these days due to increased number of cyber-attacks, we secure the data using cryptographic security measures. We use the encryption and the decryption process to secure the files so that they are not accessible to everybody except the authorized user. They need to segregated and allotted as per their respective formats to avoid any further confusion. Thus we intend to store data over cloud providing security to it and segregating it according to its respective requirement.

III. SCOPE OF PROJECT

The main system is basically designed on certain principles where the user first logs himself and the then search's according to his requirement the main advantage being as we are using the AES algorithm along with the apriori algorithm the data which he wants is secured.

IV. LITERATURE SURVEY

Big data being one of the upcoming technologies there are multiple different types of system that are been experimented to incorporate in the Big data technology. A distributed cryptographic system that allows a set of servers to prove to a client that a stored file is intact and retrievable. HAIL strengthens formally unifies, and streamlines distinct approaches from the cryptographic and distributed-systems communities. HAIL cryptographically verifies and reactively reallocates file shares. [1]

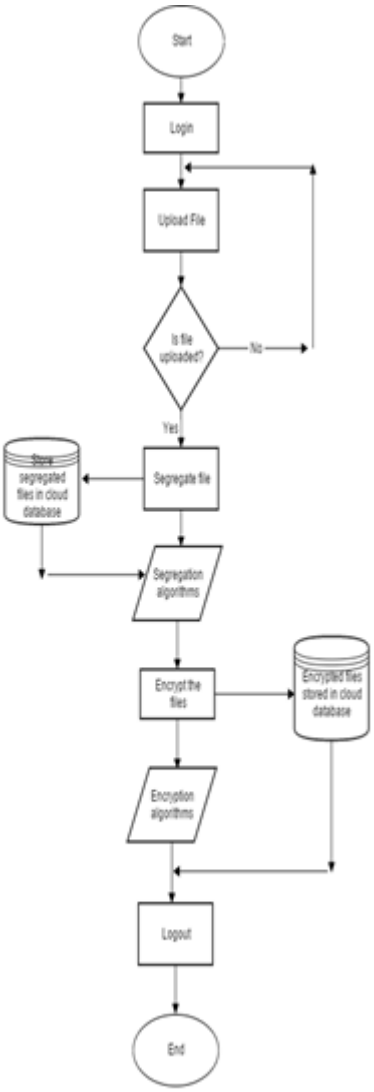
The homomorphism token with distributed verification of erasure-coded data. The integration of storage correctness insurance and data error localization, i.e., The identification of misbehaving server. [2] It ensures integrity of data is maintained due to use of distributed verification of erasure-coded data. It is highly efficient and resilient against Byzantine failure. [3] The main focus of the system is to basically analyze the complete data with the help the B-DAD framework which maps the big data tools and technique into the decision making process. [4-7] The system uses the BI&A architecture to improve the implementation of the Big data analysis on the large volume of data. It is one of the most significant type of methodology used in business practice. [8,9]

V. PROPOSED METHODOLOGY

The methodology which is to be used is to create a cloud set up using the localhost server which is the wamp server. The cloud set up is to use to store files over the cloud by creating an authenticated user login. For the user's data security, cryptographic techniques are using which are encryption and decryption. The encryption technique includes algorithm such as AES which is used to encrypt the plain text file and create its cipher text which is not

understood by common people. The big data segregation technique is used to create the segregations for the data which is stored in the cloud.

Flowchart

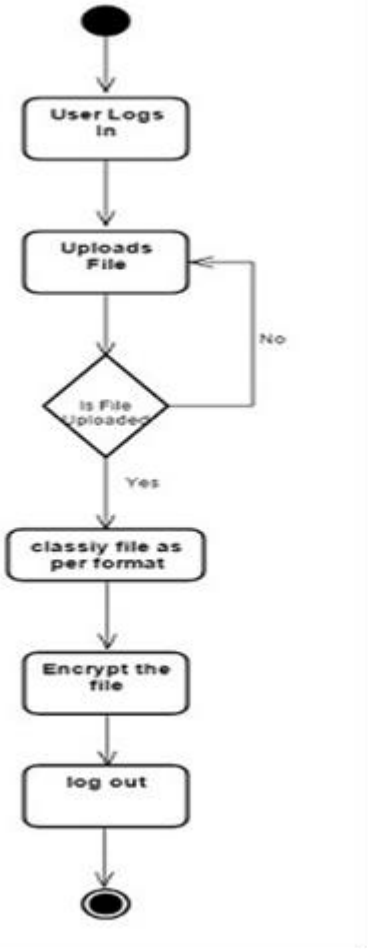


VI. FURTHUR SELECTION

1. **User Login:** The user logins using its username and password
2. **Upload file:** User uploads file on cloud
3. **Segregating file:** Here the file is classified as per its format technology
4. **Encrypting file:** Once the file is classified it is then encrypted using cryptographic AES algorithm & apriori algorithm
5. **Storing file on cloud:** Once the file is encrypted it is stored on cloud

VII. IMPLEMENTING THE SYSTEM IN REAL LIFE

The main difficulties occur when the project like big data is implemented in real life but the main advantage of big data is that we can make their data sets and implement the system



The approach of combining both the algorithm of apriori and AES for the encryption as well as decryption of data. The Big data technology is used in small data sets as we are mainly focusing on medical grounds and collecting and maintainig smaller data set in this we are incorporating the encryption and decryption algorithms for data to provide more security for the system in this the complete data is classified depending on the users choice and is also avaliable to the user to read in encrypted form if not a valid user.

VIII. CONCLUSION

We have addressed the two main issues of efficiency and accuracy in this system. The efficiency of the system can be further improved by incorporating more algorithms for the big data technology. We can classify any dataset depending on the user requirements and provide security to the system in the form of encryption as well as decryption algorithms

REFERENCES

[1] HAIL: A High-Availability and Integrity Layer for Cloud Storage. Kevin D. Bowers (RSA Laboratories) Ari Juels(RSA Laboratories), Alina Oprea (RSA Laboratories).

[2] Data Storage Security in Cloud Computing Cong Wang,Qian Wang, Kui Ren, and Wenjing Louy Dept. of ECE, Illinois Institute of Technology

[3] Data Security in Cloud Computing with Elliptic Curve Cryptography. Veerraju Gampala, Srilakshmi Inuganti, Satish Muppidi

[4] Big Data Analytics, SAS

[5] Big Data Statergies, Australian Government

[6] Big Data Technologies and Infrastructures, Cory Abbort, Arundatta Verma, Jason Roy

[7] Formulating an Executive Strategy for Big Data Analytics, Gopikrishna Palem

[8] Business intelligence and analytics from big data to big impact, Jacob Saldana, Raven James

[9] Security issues related with big data, Venkata Narasimha Inukollu, Sailaja Arsi and Srinivasa Rao Ravuri