International Journal of Engineering & Technology, 7 (2.33) (2018) 575-579

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Land forgery, black money rotation with genuine authentication using big data

Mrs. M. Kiruthiga Devi [1] *, Mrs. N. Ramya [1], Priyanka. K [2], Varsha. P [2]

[1] Assistant Professor,Dept Of CSE,,Sri Sairam Engineering College
[2] UG Scholar, Dept. of CSE, Sri Sairam Engineering College
**Corresponding author E-mail: kiruthiga.cse@sairam.edu.in,*

Abstract

Land forgery is predominant in India, the project aims to eliminate forgery activities while selling and buying the land with the guideline value. The money transaction made during the purchase of property majorly contributes to the black money problem in our country. To overcome this black money rotation and land forgery, both the Buyer and Seller has to register the land with their Aadhaar card to verify the genuineness of them. While verifying, if the system identifies any criminal cases filed against the buyer or seller or the land acquired by the seller is not authorized, it blocks the transaction. The process involves of document unique verification, legal activity verification and black money tracking, all the data are maintained in Hbase database and then integrated and stored in HDFS file storage. Map Reduce programming model is used to retrieve required information from big amount of data at a faster rate. If there are no such issues on the buyer and the seller. The overall transaction amount between the buyer and seller is finally referred to the Income Tax Department. If any illegal activity is detected in the transaction, the alert is sent to the Income Tax department and police officials so that the registration is blocked and black money rotation could be avoided.

Keywords: *Big Data; Hadoop; Hbase; HDFS; Forgery.*

1. Introduction

At the time of purchasing land, from the original owner by a buyer an amount is estimated by the government called the Guideline Value. This Guideline value is calculated based on the zone in which the property is located. When purchasing the land, a sale agreement between the seller and buyer is laid out in detail mentioning particulars of the property. The sale agreement consists of the agreed cost, time span for actual sale and down payment. Black money comes into play when an excess illicit amount other than the agreed cost on sale agreement is paid to the seller. Black money is one of the major problems in our country because of the disparity between the guideline value and market price. The other important issue involved is the land forgery. A non-genuine owner selling the same piece of land to multiple buyers. This occurs as the documents submitted in the sub registrar department are not properly verified for their genuity. In some cases even the government properties are sold through forged documents, since the buyer fails to acquire the previous behaviours of the seller such as any criminal cases or other illegal activities. The current process of land registration does not provide a tangible transparency. In this system, the prevention of black money rotation and land forgery using big data analytics is focused.

2. Existing system

The traditional method of land registration doesn't involve any document forgery detection. But there are different ways to perform document genuity verification by the managing authorities. If the officials involved in the process has any suspicion in the details provided in the particulars then they could perform a document check by performing cross-checks using databases and coordination with the judicial authorities. Since this type of authentication would result in the detection of the bogus documents and only the authenticated documents will be allowed to proceed further.

3. Overview of big data

The term big data is being increasingly used almost everywhere Big Data is a term used to mean a massive volume consisting of both structured and unstructured data that is very large it is difficult to process using traditional database systems and software techniques. In most enterprise situations the volume of data is very big, it moves too fast, or it exceeds current processing capacity. Big data primarily is analysing the present data and a result is acquired for prediction and other purposes. The term structured data usually refers to data stored in a database and unstructured data term usually refers to data that doesn't reside in a traditional row-column database. With the traditional processing methods, it is difficult to handle very large amount and complex data, which comes from different sources. Big data is often characterized by some specific attributes, in most big data terminologies; these are termed as the five V's: volume, value, velocity, variety and veracity. Volume: Volume indicates the amount of data produced every second across all online channels, including all areas like social media platforms, mobile devices, online transactions, etc. While the data is growing by leaps and bounds every minute, every day, it is impossible to store and analyze data using traditional database methods. Instead, organizations and companies have to shift to use distributed systems,

Copyright © 2018 Mrs. M. Kiruthiga Devi et. al This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

where logical parts of data is stored and brought together by software for analysis. Value: Value often here refers to the worth of the data being extracted. Not all the data is made equal. In fact, having enormous amount of data need not always translate into having high value data. Velocity: Velocity indicates the speed at which new data is being generated, collected and analyzed, at any given time. The number of emails, social media posts, video clips, or even new text is being added everyday is in excess of several billion entries. And this is continuing to increase with lightning speed as tablets and mobile devices are gives us more access to add content online. As new data is added, it is always needed to be analyzed in real-time. Big data technology always provides the ability to instantly analyze the data generated.
Variety: With increasing volume and velocity comes the variety. Variety refers the different types of data that we use. As the internet and technology is evolving, data quickly becomes obsolete. Often nowadays the data are not easily categorized into tables. Veracity: Veracity is often referred to as the quality or trustworthiness of the data you collect. Considering the accuracy of the data you collect and to analyze it is important. In this context, quality is always preferred over quantity. To focus on quality, it is important to set metrics around what type of data is collected and from what sources.

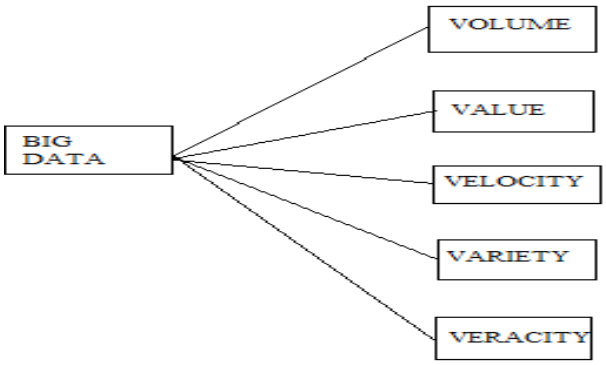


Fig. 1: V is of Big Data.

4. Hadoop overview

The Hadoop Distributed File System (HDFS) is a distributed file system built to run on various platforms. It has various resemblances with existing distributed file systems, still hadoop has its own uniqueness. One of the important characteristics of HDFS is its fault tolerance and it is not an expensive system. It provides high throughput and is useful for applications that have large data sets. With HDFS a single cluster could be made to scale up to a thousands of nodes. HDFS has the potential to work under various physical circumstances. It distributes storage and computation across many servers, this thus enables the combined storage resource to grow linearly with demand.
HDFS has shown the production scalability, of data up to many petabytes, that level of quantity and quality of enterprise data is available in HDFS. Name node mainly controls the namespace of the file system and balances the client's access to the desired files. It also performs the file system operations such like renaming, closing, and opening files and directories. Data node manages the read-write operations on file systems, based on the client request. Along with this, it also enables operations such as block creation, deletion, and replication based on the instructions given by the name node. The file in a file system will be divided into one or more segments (files are the user data of HDFS) and stored in individual data nodes. These file segments are also referred as blocks.

5. Hbase overview

Hadoop database or HBase runs on the top of Hadoop and is a No Sql database. Hbase runs on the Hadoop Distributed File System/HDFS. It uses a key - value pair and also makes use of the Map Reduce for the profound analysis of the data to be retrieved. Thus with the use of HDFS and the Map and Reduce tasks the HBase is employed efficiently. This makes HBase a dynamic tool to process billions of records of data, each with number of rows and columns. Its potential are not only limited to processing billions of amounts of data but also to the extent of enabling many real time queries to be implemented within any given particular time precisely, all of this is done with the help of MapReduce task. HBase also allows the retrieval of the specific singular data and also for any number of batches of data available within a large dataset.

- Row Key: Every row will have a specifically identifiable key called the row key.
- Column Family: Column Family is the data aligned inside of the row; every row will have the exact number of column family. Now the HBase will store these column family in a database maintained by its own.
- Column Qualifier: Column Qualifier is the columns described by the column family.
- Version: There are different modifiable types of columns called the version. Thus making it easier to access the record of a particular type.

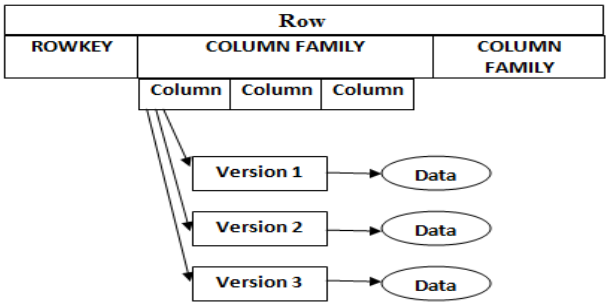


Fig. 2: Hbase Four-Dimensional Data Model.

For any specific row to be accessed the particular row key has to be used, which could contain a single or more number of columns. To make access to an individual piece of data, its row key, column family, column qualifier, and version has to be known. HBase data could be accessed in two ways: By making use of the row key or using map-reduce tasks.
Therefore this possibility of the Hadoop database to acquire data access makes it a very approachable tool. HBase enables both the dynamic analysis and the on demand batch-by-batch analysis possible for processing. By allowing the row key and column family to be the key and value pair, it is attainable.

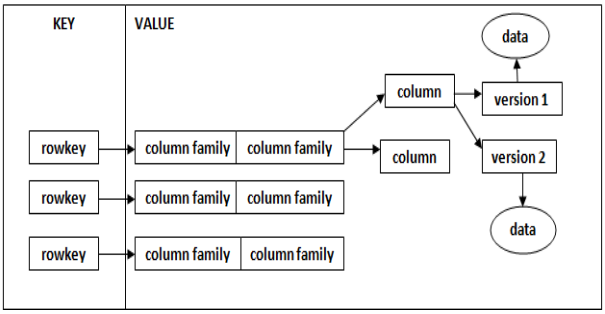


Fig. 3: Hbase as A Key/Value Store.

Each key will have a value related to it and this could be fetched by the Figure-3 shown above. This is made possible by providing the initial and final row key, or the table scan. Values contained in columns cannot be queried in a real-time query, which leads to: row key design. The design of the row key is important for two reasons:

- Table scans that operate not in favour of the row key, hence design of row key supervise how real-time access can be done against HBase.
- Running HBase in a production environment, enables to run it on top of HDFS also data is however distributed across HDFS according to row key. To administer the row key to different deployments it has to be unique.

Therefore the order by which row keys are allocated is important to acquire the row for further access. In HBase, when data are stored on single user basis, then the alternate approach is to rank the fact that row keys are a eventually stored as byte array.

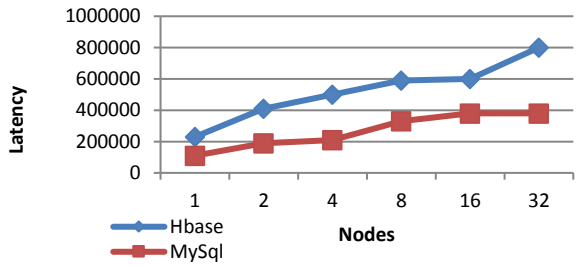


Fig. 4: Graph Showing Comparison of Hbase and Mysql Database Latency.

The above graph shows that the time of retrieval of data from hbase is much faster than mysql for huge data sets.

6. Implementation of hadoop using H base

The primary underlying task of Hadoop is to analyze the data and then to process those data which is easier to retrieve. Hadoop divides the work into two steps:

- Map: An initial input step and transformation step, in which individual input records can be processed in parallel.
- Reduce: An aggregation step or the summarization step, in which all associated records is processed together by a single entity.

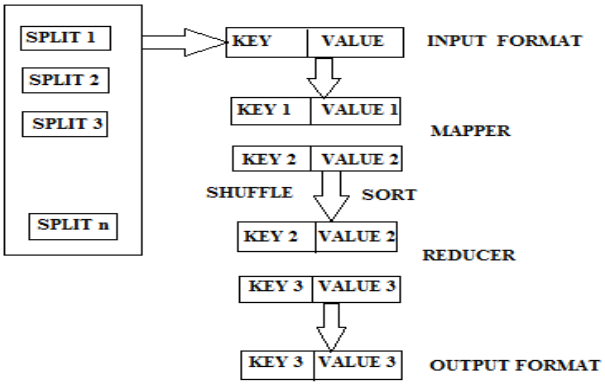


Fig. 5: Hadoop Map Reduce Process.

1) Hadoop splits the analysis data into logical chunks using as an Input Format.
2) Each logical chunk is to be sent to a MapReduce process in the form of value mapped to a key.
3) Each key/value pair is then sent to the map process and the map process generates a new key/value pair.
4) Hadoop then performs a shuffle and sort operation, to group all mapped keys together. A new key/value pair is now will be sent to the reduce process, which generates another key/value pair.
5) Finally, all key/value pairs generated by the reduce process are written to a database, such as a HBase table.

The first step is that the application has to gain access to all the Page Views table. For a connecting to the Page Views table in HBase can be done in one of the two ways:

1) Create an HTable instance.
2) Use a connection pool.

Methods used for operations in database:

- To Get an object instance from the database, Create a Get object and pass it to the HTableInterface get () method.
- To Put an object instance into the database, Create a Put object and pass it to the HTableInterface put () method.
- To Delete an object instance from the database, Create a Delete object and pass it to the HTableInstance delete() method.
- To Scan the database for rows within start and end row values, Create a Scan object and pass it to the HTableInstance getScanner() method.

For insertion of data: The put() method creates a new Put object, passing it the row key. all objects are stored as an array of bytes, then to convert the row key String value to a byte[]. We can do that with the Bytes.to Bytes() method. The constructor for the Put objects requires the row key, but then need to add our fields. Then column family has to be created with column qualifiers. Then column family, column qualifiers and associated values are to be converted to byte arrays then add them to the Put object using the add() method. Finally, by calling the HTableInterface put()method with the Put object to insert the record in the database.

For retrieval of data: The get() method creates a Get object, passing it a byte array representation of the row key, and then executes the HTableInterface get() method, passing it our Get instance. This method will return a Result instance that can be used to retrieve the values that are needed. Also it is possible to pass the Resultclass getValue() method, desired the column family and the column qualifier, as byte arrays, and it will return byte array that will contain the requested value.

For deletion: The delete() method is the simplest: It will create a Delete object with a byte array format of the row key to delete, and then it implements the HTableInterface delete() method.

For scan: The scan() method is the most complex method. In Hbase no query is used but can retrieve all records between two row keys. Firstly create a Scan object, and then pass it to the start and end row keys (as byte arrays). Then pass this Scan object to the HTableInterface getScanner()method, which returns a ResultScanner. Then it is necessary to iterate over the ResultScanner and handle the results just as we handle the Result in the get() method.

7. Proposed model

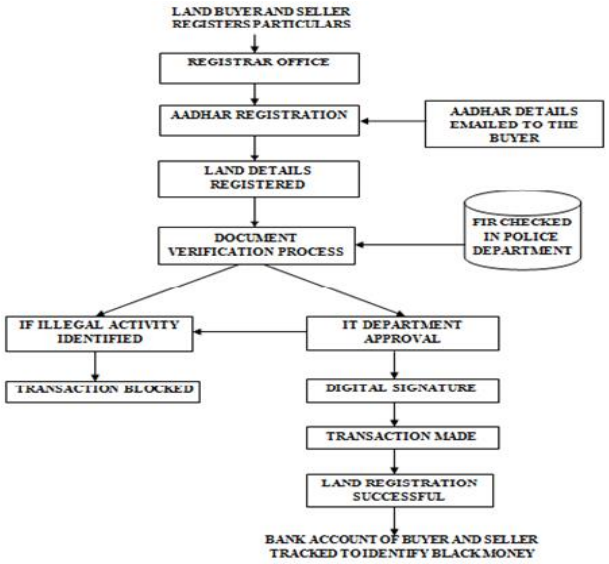


Fig. 6: Overall Block Diagram.

7.1. Buyer and seller registration

The buyer and seller involved in the land registration process is allowed to register and upload all the relevant land related documents along with the seller and buyer's identification. The identification will be the aadhar card held by them. These documents are uploaded in the registrar office of the particular area and the user credentials basically includes the address proof of the buyer and seller, encumbrance certificate to ensure that the seller has a complete ownership over the land without any legal liability, the details about the total square feet of the land, the draft agreement (including purchasing

cost) made according to the guide line value of that particular zone. All these particulars are enrolled in the presence of the registrar in the registrar office and once this enrollment is done, the softcopy of these documents are stored and maintained in the databases. These documents are further used and sent to various managing authorities in the distinct departments to verify the genuity of the buyer and seller.

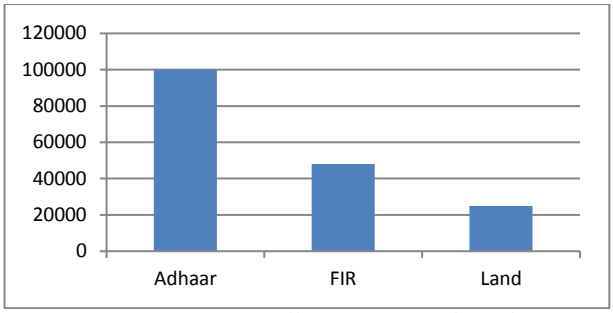


Fig. 7: Bar Graph Indicating the Difference in Range of Data for the Adhaar Holders, FIR Cases and Land Registered People Databases.

The above graph shows that number of people having the adhaar card is more than those people who have fir case lodged on them and the number of people who are involved in land registration.

7.2. Document genuity verification

The buyer and seller's details are sent to the corporation department's server where the document genuity verification of both of these parties begins. The corporation department of the state government will cross check the birth certificate of the proprietor of the land, if the property is transferred to the present proprietor from a previous proprietor, then the previous proprietor's birth and death certificates are verified. It also examines the legal status of the document by comparing encumbrance certificate and legal heir of the original owner. Since the corporation department has large volumes of records of the people, to fetch the details of the buyer and seller could be very complex. So to carry out this process of fetching records big data analytics is used, with the help of map reduce the particulars of the buyer and seller are retrieved. During inspection of the documents if the corporation department identifies any illegitimate activity of the buyer or seller, then they will block the land registration process. Further process will be proceeded only if the lawbreaker reports to the registrar and provides obligatory acknowledgement. Otherwise if the documents are established to be bona fide then these are redirected to the police department.

7.3. Legal activity verification

In the police department, the legal activity verification of the buyer and seller's previous behaviour and crime activities are analyzed by the law enforcement authorities by comparing the records with the police station. The police officials will maintain the database of all criminals who has First Information Report filed on them. This database will have personal details of the criminals including their aadhar number, name, residence, age and the type of crime committed by them. Using map reduce the police officials will cross check in their databases if there are any First Information Report cases filed on both the parties to know information about any criminal offence done by them in the past. This process is done by choosing the aadhar number of the buyer and seller and searching for it in the database held by the police department, if any match is found with the FIR then the land registration process will be blocked by the managing authorities in the police department. However if there are no such criminal activity found on both the parties then approval for the further process will be granted by the police officials.

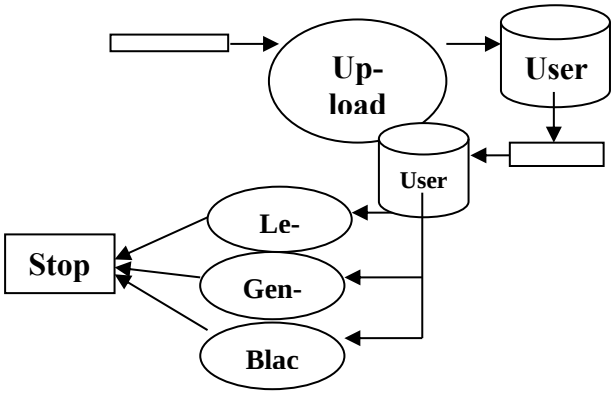


Fig. 8: Dataflow Diagram.

7.4. Black money tracking

Buyers are purchasing the land by paying white money towards guide line value and balance money is paid through black money to the seller. This black money is unaccounted money transaction done in the banks. So in order to avoid this unaccounted money transaction a digital cheque for the entire money specified in the draft agreement is given to the seller by the buyer. Because of this, all the money transaction will become white. Once the police officials grants permission for proceeding the land registration then, Income Tax department will scan all the land deeds to ensure that the purchasing amount specified in the draft agreement is the same amount as mentioned in the digital cheque. If the Income Tax department finds any suspicious activity in this process then they have the right to block the land registration process. Otherwise if there are no such felonious behavior found in both buyer and seller, then with the IT department's consent the money transaction will be allowed based on digital cheque. No cash transaction is entertained here in the land registration process. As all the bank accounts are integrated with Aadhaar card and PAN card. Any amount transaction beyond certain value is immediately communicated to the Income tax server. Even after all the money transaction in the land registration process is completed, the IT department will keep under the observation both the buyer and seller's bank account details, to ensure no black money transaction is made in the future.

8. Conclusion

The proposed system ensures fraud free land registration. If any fraudulent act is found, it could be identified by using their aadhar card details of both seller and buyer. It also monitors the entire money transaction through aadhar number and if any illegal money transaction or unaccounted cash transaction is made the system will block the land registration process. The proposed system will also prevent cheating of dishonest owners who sell the same piece of land to multiple buyers. By this method the government properties cannot be sold through fabricated documents. Overall, the proposed system will provide transparency in the land registration process.

References

[1] Understanding Big Data Analytics Workloads on Modern Processors Zhen Jia, Jianfeng Zhan, Lei Wang, Chunjie Luo, Wanling Gao, Yi Jin, Rui Han, and Lixin Zhang, June 2017

[2] Record Linkage Approaches in Big Data: A State Of Art Study Randa M. Abd El-Ghafar, Mervat H.Gheith Computer Science Department Institute of Statistical Studies and Research, Cairo University, Cairo, Ali H. El-Bastawissy Faculty of Computer Science Modern Sciences and Arts University Cairo, Eman S. Nasr Independent Researcher Cairo, Egypt,2017

[3] Survey on Big Data Analytics for Digital World Sayali Gaikwad, Department of Computer Engineering, JSCOE, Hadapsar, Pune, India. Pranali Nale, Department of Computer Engineering, JSCOE,

Hadapsar, Pune, India. Ravindra Bachate Assistant Professor, Department of Computer Engineering, JSCOE, Hadapsar, Pune, India, Dec 2016.

[4] A review of research on MapReduce scheduling algorithms in Hadoop, Namrata Singh Dept. of Computer Engineering & Applications, National Institute of Technical Teachers' Training and Research, Bhopal, India Sanjay Agrawal Dept. of Computer Engineering & Applications, National Institute of Technical Teachers' Training and Research, Bhopal, India, July 2015.

[5] A survey on security of Hadoop Masoumeh Rezaei Jam Department of Computer, Engineering University of Tabriz Tabriz, Ira Leili Mohammad Khanli,Department of Computer, Engineering University of Tabriz Tabriz, Iran Morteza Sargolzaei Javan,Department of Computer Engineering and IT, Amirkabir University of Technology, Tehran, Iran, Dec 2014