An un-even Key Cryptosystem for Providing the Privacy-Preserving of Data Security

M. Naveen Kumar[1] Mr. J. S. Ananda Kumar[2]
[1]Student [2]Assistant Professor
[1,2]Department of Computer Applications
[1,2]KMM Institute of PG Studies, Tirupati, India

Abstract— The reality used by different organizations is expanding at a snappy cost. In this current day's reality, organizations need to framework Zeta bytes of comprehension. The ordinary database the executive's framework neglects to way such amazing arrangement of data. Along these lines, we would need to search out a proficient procedure for overseeing and procedure such monstrous amount The reality used by different organizations is expanding at of understanding which bears ascend to immense insights downside. Size or volume of data isn't exclusively the best criteria to order expansive records, we must remain as a primary concern the sort of data that is whether or no longer actualities is based or semi-organized or unstructured. To beat this monstrous ability detriment Apache made a product gadget named Hadoop that utilizes Map Reduce format to the way the tremendous measure of data. Amid this, AN uneven key cryptosystem is anticipated for the mystery composing of archives spared in Hadoop Distributed record framework. The impacts got from various trials infer positive results of the above strategy to manage security drawback identified with immense ability.

Keywords: Bigdata, Hadoop, Map Reduce, Encryption, Security

I. INTRODUCTION

Huge information is a term used to allude to informational collections that are excessively extensive or complex for customary information preparing application programming to enough arrangement with. Information with numerous cases (lines) offers more prominent measurable power, while information with higher unpredictability (more qualities or segments) may prompt a higher false disclosure rate. Huge information challenges incorporate catching information, information stockpiling, information examination, seek, sharing, exchange, perception, questioning, refreshing, data protection and information source. Enormous information was initially connected with three key ideas: volume, assortment, and speed. Different ideas later credited with huge information are veracity (i.e., how much clamor is in the information) and esteem. Currently used expression "enormous information" will, in general, allude to the utilization of prescient investigation, client conduct examination, or certain other propelled information examination techniques that separate an incentive from information, and sometimes to a specific size of the informational index.

Huge information, as the name demonstrates, is voluminous information with different highlights, for example, speed and assortment when information is a social database, it is known as organized information. At the point when information is as archives of any sort, it is known as unstructured information. At the point when information is as XML, it is known as semi-organized information. While the size used to decide if a specific informational collection is viewed as large information isn't immovably characterized and keeps on changing after some time, most experts at present allude to informational indexes from 30-50 terabytes to various Zeta bytes as large information. The proposed strategy is likewise a halter kilter key cryptosystem which utilizes two keys for encryption and decoding. Open key utilized for encryption and relating private key [1] is utilized for unscrambling. Proposed another variant of DSA Reverse the plain content then after encoding it. Our proposed plan has drawn thought from these plans. In DSA, just two huge prime numbers are considered while in our proposed encryption framework, we have taken four extensive prime numbers.

As in halter kilter key cryptography, the direct lies in how toward make it troublesome for the aggressor to factorize n which is the duplication of those four prime numbers. Our proposed framework isn't just about expanding prime numbers yet additionally, we have connected security as far as open key and private key. In DSA, the open key comprises of e and n yet we have incorporated another parameter f. What's more, in the private key, we have included three different parameters. The social database the executive's frameworks, work area measurements, and programming bundles used to envision information frequently experience issues taking care of huge information. The work may require "greatly parallel programming running on tens, hundreds, or even a large number of servers". What qualifies as being "huge information" changes relying upon the abilities of the clients and their apparatuses, and extending capacities make huge information a moving target? "For certain associations, confronting several gigabytes of information out of the blue may trigger a need to rethink information the board choices. For other people, it might take tens or many terabytes before information estimate turns into a noteworthy though.

Hadoop is a Programming structure used to help the preparing of huge informational indexes in a conveyed processing condition. Hadoop was created by Google's Map Reduce that is a production system where an application separates into different parts. The Current Apache Hadoop environment comprises of the Hadoop Kernel, Map Reduce, HDFS and quantities of different segments like Apache Hive, Base and Zookeeper, HDFS and Map Reduce.

II. RELATED WORK

A. A Survey on the Security of Hadoop:

Trusted computing and security of the several services is one of the most challenging topics today and is the cloud computing's core technology that is currently the focus of international IT universe. Hadoop, as an open-source cloud

computing & big data framework, is increasingly used in the business world, while the weakness of security mechanism now becomes one of the main problems obstructing its development. This paper first describes the Hadoop project and its present security [2] mechanisms, then analyzes the security problems and risks of it, pondering some methods to enhance its trust, security and also finally based on previous descriptions concludes Hadoop's security challenges.

B. Authentication Service in Hadoop Using one Time Pad:

"Big Data" - voluminous and a variety of data from various sources, which demands innovative processing and analysis for decision - making an analysis. The data can be either in the form of structured (or) unstructured data. Processing big data with the traditional processing tools, also the present relational database management systems tend to be a difficult task. Parallel execution environment, like Hadoop, is most needed for processing voluminous data. For processing the data in an open framework like Hadoop we need a highly secure authentication system for restricting the access to the confidential business data that are processed. Here, a novel and a simple authentication model using one-time pad algorithm that removes the communication of passwords between the servers is proposed. This model tends to enhance the security in Hadoop environment.

C. Map Reduce:

1) Simplified data processing on large clusters:

Map Reduce is a programming model and it is also an associated implementation for processing and generating large data sets. Users specify a map functions of processes a key or value pair to generate a set of intermediate key or value pairs. A reduce functions merge all the intermediate values associated with the same intermediate key. Programs are written in the functional style are the automatically parallelized and executed on a large cluster of commodity machines. The run-time system takes care of the details of partitioning the input data, scheduling the program's execution across a set of machines, handling machine failures, [3] and managing the required inter-machine communication. The inter-machine communication is allowed programmers without any experience with the parallel, distributed systems utilize the resources of a large distributed system. Our implementation of Map-Reduce runs on a large cluster of commodity machines and is highly scalable: a typical Map-Reduce computation processes many terabytes of data on thousands of machines. Programmers find the system easy to use: hundreds of Map-Reduce programs have been implemented and upwards of one thousand Map Reduce jobs are executed on Google's clusters every day.

D. Proposed Algorithm:

In awry key cryptography, the direct lies in how toward make it troublesome for the aggressor to factorize n which is the duplication of those four prime numbers. In our proposed encryption framework, we have taken four expansive prime numbers. Our proposed framework isn't just about expanding prime numbers yet, in addition, we have connected security as far as open key and private key. In RSA, the open key comprises of e and n yet we have incorporated another parameter And in private key. The key age calculation takes a little more noteworthy time than DSA calculation. In the event that we review encryption of reports, at that point, our proposed procedure stands out as it makes utilization of four specific best [5] numbers in inclination to two as in DSA calculation. As the assortment of best number is quickened, it will make the enemy difficult to factorize it. Aside from that, there is each other homegrown assortment each covered in both open key and individual key. So the general insurance will increment.

E. Algorithm

1) DSA Algorithm:

The Digital Signature algorithmic rule (DSA) could be a Federal science normal for digital signatures, These are supported the mathematical idea standard acceptations. The DSA algorithmic rule works supported the framework of public-key cryptosystems and relies on the algebraical properties of the standard exponentiations, besides the separate power drawback (which is taken into account to be computationally intractable). Messages are written by the signer's personal key and also the signatures are confirmed by the signer's corresponding public key. The digital signature provides message authentication, integrity, and repudiation in contrast to DSA, most digital signature sorts are generated by sign language message digests with the personal key of the mastermind. This creates a digital fingerprint of the info. Since simply the message digest is signed, the signature is usually abundant smaller compared to the info that was signed. As a result, digital signatures impose fewer masses on processors at the time of sign language execution, use little volumes of information measure, and generate little volumes of ciphertext meant for cryptanalytics.

DSA, on the opposite hand, doesn't code message digests exploitation personal key or rewrite message digests exploitation public key. Instead, it uses distinctive mathematical functions to form a digital signature consisting of 2 160-bit numbers that are originated from the message digests and also the personal key. DSAs build use of the general public key for authenticating the signature, however, the authentication method is a lot of sophisticated compared with RSA. The digital signature procedures for RSA and DSA are sometimes thought to be being equal in strength. As a result of DSAs are completely used for digital signatures and build [6] no provisions for encrypting knowledge, it's generally not subject to import or export restrictions that are usually implemented on RSA cryptography. The primary a part of the DSA algorithmic rule is that the public key and personal key generation, which can be described as:

- Select a prime number q, which is called the prime divisor.
- Select another primer number p, such that p-1 mod q = 0. p is called the prime modulus.
- Select an integer g, such that 1 < g < p, g**q mod p = 1 and g = h** ((p–1)/q) mod p. q is also called g's multiplicative order modulo p.
- Select an integer, such that 0 < x < q.
- Compute y as g**x mod p.
- Package the public key as {p,q,g,y}.
- Package the private key as {p,q,g,x}.

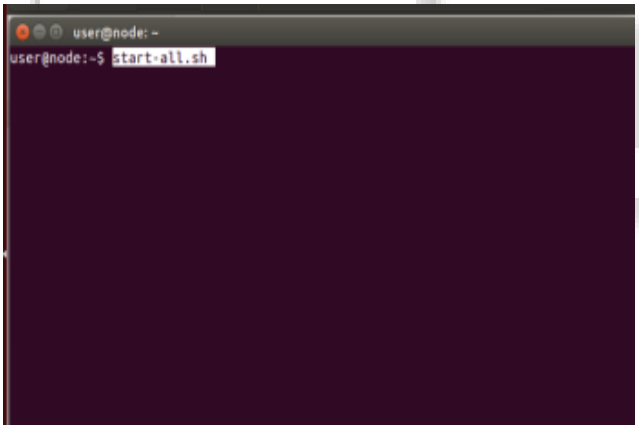
The second part of the DSA algorithm [7] is the signature generation and signature verification, which can be described as: To generate a message signature, the sender can follow these steps:

- Generate the message digest h, using a hash algorithm like SHA1.
- Generate a random number k, such that 0 < k < q.
- The computer as (g**k mod p) mod q. If r = 0, select a different k.
- Compute I, such that k*i mod q = 1. i is called the modular multiplicative inverse of k modulo q.
- Compute s = i*(h+r*x) mod q. If s = 0, select a different k.
- Package the digital signature as {r,s}.

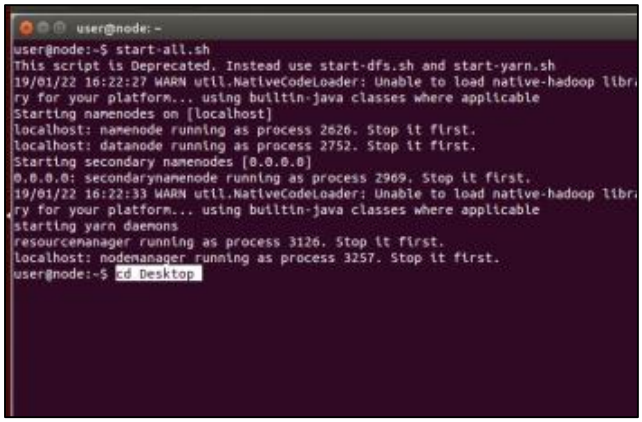
To verify a message signature, the receiver of the message and the digital signature can follow these steps:

- Generate the message digest h, using the same hash algorithm.
- Computer w, such that s*w mod q = 1. W is called the modular multiplicative inverse of s modulo q [8].
- Compute u1 = h*w mod q.
- Compute u2 = r*w mod q.
- Compute v = (((g**u1)*(y**u2)) mod p) mod q.
- If vs == r, the digital signature is valid.

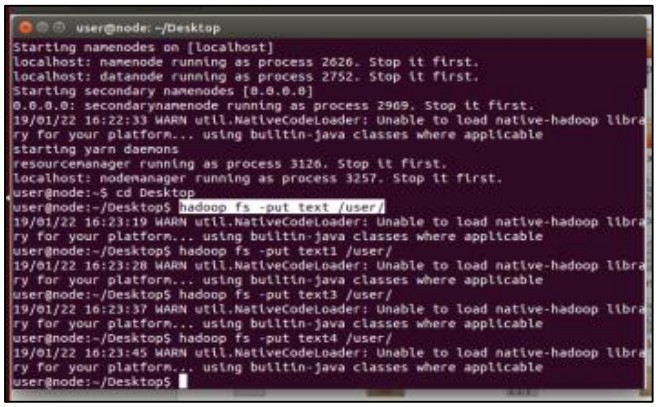
III. RESULT



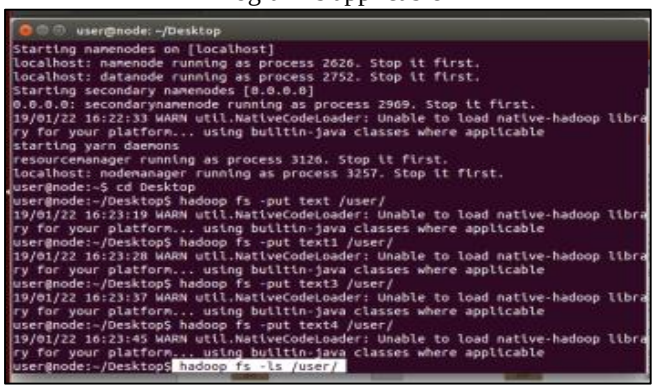
Start User Node



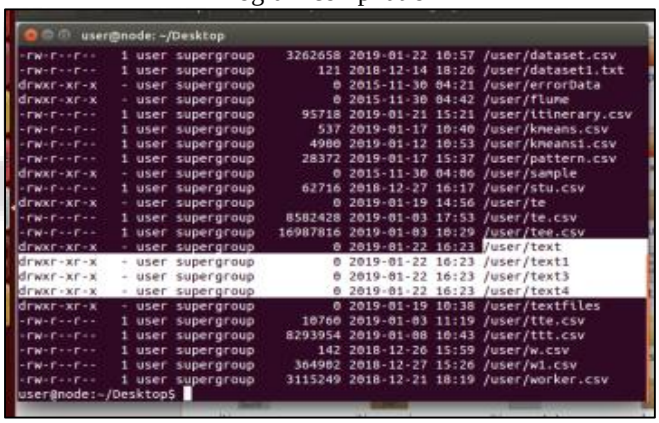
Create a localhost



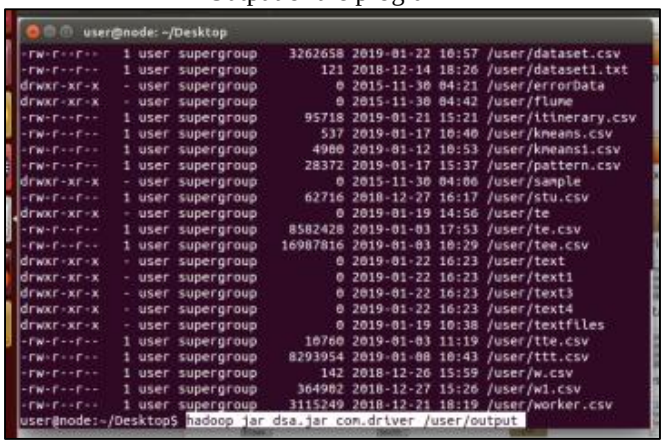
Program is applicable



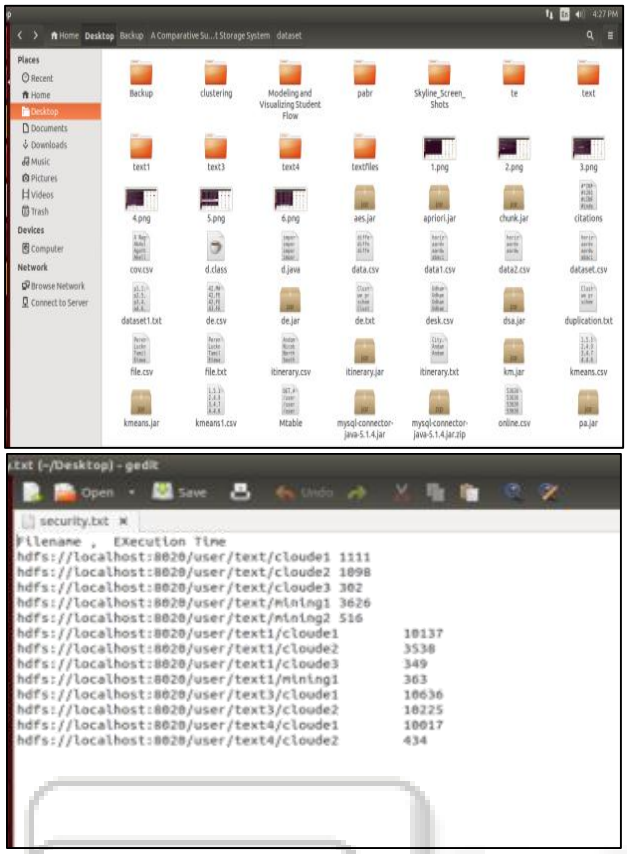
Program compilation



Output of the program



output



Security generated data

IV. Conclusion and Future Scope

In spite of the fact that Hadoop enables us to triumph over requesting circumstances faced in businesses and establishments because of extensive actualities, it has now no assurance system. The measurements put away in Hadoop can be undermined by an aggressor or meddler. As Hadoop does not give any insurance system, the genuineness of information is for the most part in question. The proposed halter kilter key arrangement of tenets encodes the substance material of records before putting away it into HDFS along these lines by methods for verifying it from the various assaults on the system. Consequently, the realities or reports presently can be put away in Hadoop without requesting around security issues by utilizing making utilization of the encryption calculation on the records sooner than putting away it in Hadoop.

References

[1] Turkington G. Hadoop Beginner's Guide. PACKT Publishing; 2013.
[2] White T. Hadoop Definitive guide OReilly, 2009.
[3] Owen O Malley. Integrating Kerberos into Apache Hadoop Kerberos. Conference 2010, 2010 26–27 Oct, MIT, USA.
[4] Hwang K, Kulkarni S, Hu Y. Cloud security with virtualized defense and reputation-based trust management. Eighth IEEE International Conference on Pervasive Intelligence and Computing, (PICom2009). 2009; Chengdu, China. p. 717–22.
[5] SudhaSadasivam G, Anitha Kumar K, Rubika S. A novel authentication service for Hadoop in cloud environment. IEEE International Conference on Cloud Computing in Emerging Markets (CCEM). 2012. p.16.
[6] Hamlen K, Kantarcioglu M, Khan L, Thuraisingham B. Security issues for cloud computing. International Journal of Information Security and Privacy. 2010 Apr-Jun; 4(2): 39-51.
[7] Kousiouris G, Vafiadis G, Varvarigou T. A frontend, Hadoop-based data management service for efficient federated clouds. IEEE Third International Conference on Cloud Computing Technology and Science (CloudCom). 2011; p. 511–16.
[8] Lin CL, Hwang T. A password authentication scheme with secure password updating. Comput Secur. 2003; 22(1): 68-72.

