A REVIEW OF HYBRID APPROACH FOR BIG DATA AND HADOOP IN SECURITY POINT OF VIEW

Suvrat Jain
Student, Neerja Modi School, Jaipur, Rajasthan, India

Abstract: Big data describes innovative technique to capture, store, distribute, manage and analysis large data sets with high velocity and different structures. Hadoop is an open source programming framework that enables the distributing processing of large data setsacross clusters and commodity servers with a very high degree of fault tolerance. This paper demonstrates the Hybrid Enormous information issues and concentrated on Security challenges emerges in Hadoop Design center layer called Hadoop Distributed File System (HDFS). The HDFS Security enhancement by using new theme approach like Secure Hash Algorithm SHA-3 for secure data transfer in between node to node.

Index Terms: Big Data, Hadoop, HDFS, SHA3 Security

I. INTRODUCTION

Big data [1] is a current technology which rule a world in future. The term "Big Data" end up well known in most recent couple of years, as it speaks to the diligent work of researchers to accomplish business insight by handling immensely substantial measure of information. To gather, store, oversee and investigations it is extremely troublesome for customary dataset programming devices. In 1944 Fremont Rider [2] specified that the American University library was multiplying in the measure at regular intervals. He spoke to in 2040 this library will hold in excess of 200,000,000 volumes of books which will possess 6000 miles of racks in library. Of course big data is too large to load into memory and store on a hard-drive and fit in a traditional database. The idea picked up force in the mid-2000s when industry investigator Doug Laney explained the now-mainstream meaning of big data as the three V's: **Volume**: Associations gather information from an assortment of sources, including business exchanges, online networking, and data from a sensor or machine-to-machine information. Before, putting away it would've been an issue – however new advancements, (for example, Hadoop) have facilitated the weight. **Velocity**: Information streams in at a phenomenal speed and should be managed in an auspicious way. RFID labels, sensors, and shrewd metering are driving the need to manage downpours of information in close ongoing. **Variety**: Information comes in a wide range of organizations – from organized, numeric information in customary databases to unstructured content records, email, video, sound, stock ticker information, and monetary exchanges. Big data in real time its rate of growth is increased from Gigabytes in 2005 to Exabyte in 2015(forecast) which is reported by IDC research.

Recently Wikibon extended its Big Data forecast that Big Data market to reach $50 Billion by 2018 to year 2020. Wikibons says that Big Data advertise proceeded with its development in 2014, encountering both noteworthy developments as estimated by seller income related with the offer of Huge Information items and benefits and expanded reception of Huge Information instruments and advances by substantial endeavors crosswise over vertical markets. For the calendar year 2014, the Big Data market - as estimated by income related with the offer of Big Data related equipment, programming, and expert administrations - came to $27.36 billion, up from $19.6 billion in 2013. While becoming fundamentally quicker than other endeavor IT markets. Wikibon said that Big Data market to top $61 billion in 2020, a 26% compound yearly development rate for 2011-2020. As there are more 1 billion individuals utilizing a versatile for exchanging data every month where these information are observed by media transmission enormous server farm and permits the server farm to store in excess of 621 petabytes of information for each year. The big data examination [3] permits rapidly recognizing the dangers and openings and furthermore expanding capacities of prescient investigation and Big Data Attributes [22]. Various components driving development of the Big Data market which include: data warehouse optimization, starting Big Data utilize case material crosswise over vertical markets and proceeded with development of big data items and administrations, especially as to information administration, information change and information pipeline creation abilities. The expanding foundation of Big Data driven basic leadership as a key need in board-rooms and C-suites, monetary administrations, retail, medicinal services, and broadcast communications ventures. Table 1 demonstrates the contrast between Traditional data and Big data.

Table-1: Difference between Traditional Data and Big Data

Dimensions	Traditional Data	Big Data
Data sources	Mainly Internal	Both inside or outside organization including traditional data warehouses
Data Structure	Pre defined Structure	Un structured in nature
Data Relationships	Stable and Interrelationships	Unknown Relationships
Data Location	Centralized	Physically Highly Distributed
Data Analysis	After complete build	Intermediate analysis as we go
Data Reporting	Limited and Predefined interaction paths	Reporting in all possible direction across the data in real time mode
Cost Factor	Specialized High end hardware and software	Inexpensive commodity boxes in cluster mode

1.1 Various Challenges associated with Big Data

Big data analytics face different challenges. They are several big data Challenges such as Management, Privacy and Security Processing Challenge [24], and Storage Challenges [25]. Each challenge has its own task of surviving in big data and mainly focusing on security Challenges.

1.1.1. Management Challenges

The greatest data administration [23] is the gathering of substantial volumes of Organized, Semi-organized and unstructured information from a various association, Government area and Private and Open Organization. The big data administration is guaranteeing a lot of valuable information and its possession, obligations, institutionalized documentation and availability of informational index to be handled. The information are broke down and it is imperative to catch the huge test. As per Gartner [4]"Big Data" Challenge Involves More than Just Managing Volumes of Data said in his Article.

1.1.2. Storage Challenges

The Storage is accomplished by utilizing virtualization in big data where it holds a substantial arrangement of Sensor data, media, recordings, E-business exchange records, Mobile phone Flag Directions. Numerous big data Stockpiling Organizations Like EMC [12], IBM, Netapp, Amazon Handles an information in a Substantial volume by utilizing a few devices like NoSQL, Apache Drill, Horton Works [13], SAMOA, IKANOW, Hadoop, Map reduce, Grid Gain.

1.1.3. Processing Challenges

The opportuneness there is numerous circumstances in which the aftereffect of the investigation is required instantly. Given a Vast informational index, usually important to discover the components huge informational collections that meet determined criteria to process, examine the time it will take. That is Information is becoming hard to outline a structure rapidly. The big data handling breaks down the huge information estimate in Petabyte, Exabyte or even in Zettabyte either in Bunch Preparing or Stream Preparing.

1.1.4. Diversity and Incompleteness

The machine algorithms expected homogeneous assays Data, and you cannot understand the nuances. Even after data Cleaning and bug fixes, some incomplete and the data is likely to be some mistakes.

1.1.5. Human cooperation and Skills

A large data analysis system many of entry should support human experts, and shared the results of exploration. Big data requires people with new skills. Management of large data effectively requires accurate people.

1.1.6. Privacy and Security Challenges

Another great challenge Protection of personal privacy such as health industry, the history of the person is Personal. But it may be available from many sources. Therefore, it is difficult to maintain privacy and security.

For dealing with a substantial informational index in secure way and wasteful devices, open and private database contain more dangers and vulnerabilities, volunteered and startling spillage of data, and lack of Open and Private Arrangement makes a programmer to gather their assets at whatever point required. In Appropriated programming structures, the security Difficulties begin working when huge measure of private information put away in a database which isn't encoded or in general arrangement. Anchoring information with nearness of untrusted individuals is more troublesome and while moving from homogeneous information to the Heterogeneous information certain apparatuses and innovations for gigantic informational index isn't frequently created with greater security and arrangement endorsements. Once in a while information programmers and framework programmers includes in gathering a freely accessible huge informational collection, duplicate it and store it in a gadgets like USB drives, hard circle or in PCs. They includes in assaulting the information stockpiling by sending a few assaults like Refusal of Administration [14], Snoofing assault and Animal Power assault [15]. At the point when the Capacity of information increments from single level to Multi stockpiling level the security level should likewise be expanded. Keeping in mind the end goal to diminish these Difficulties some cryptographic System procedures and hearty calculation must be created with a specific end goal to improve the security of information for future. So also a few devices are produced like Hadoop; NoSQL, Oozie technology can be utilized for big data storage. In our proposed work theme a few thoughts are given to overcome security Challenges in Hadoop environment.

II. HADOOP AND ITS FRAMEWORK

Hadoop (Highly Archived Distributed Object Oriented Programming) was created by Goug Cutting and Mike Cafarella in 2005 for supporting a distributed search Engine Project. Hadoop is utilized for putting away getting to and preparing the big data in a disseminated environment at less cost, a high level of adaptation to internal failure and high adaptability. Hadoop is an open source java – construct programming system which keeps running with respect to Product equipment. Hadoop [5] handles the vast number of data like Pictures, sounds, recordings Documents, Programming, Sensor Records, Correspondence information, Organized Question, unstructured information, Email from various framework at any configuration. Every one of these assets can be put away in a Hadoop group with no pattern portrayal as opposed to gathering from various frameworks. There are many components involved in Hadoop like Avro, Kafka, Flume, HBase, Hive, Oozie, Pig and Zookeeper including third party application services. The Hadoop Package provides Documentation, sourcecode, location awareness, Workflow scheduling.

A Hadoop cluster contains one Master node and Many Slave nodes. The master node consists of Name node which is responsible for indexing of data node when applications contact the name node it tells the application to go to the particular node to get required data. And Job Tracker which play the role to break the bigger task into smaller modules & send each module of computation to task tracker. At slave node acts as both a TaskTracker performs the small task & send result back to the job tracker and Data node manage the data that has been given to particular node. The Job Tracker manages the job scheduling.

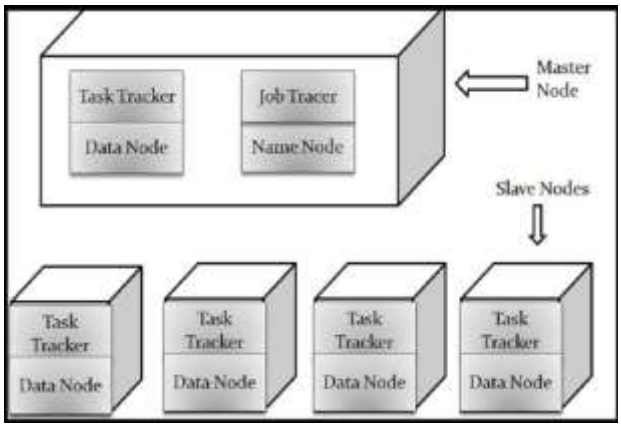

Fig.1. Structure of Hadoop

Hadoop classified into two parts one is Hadoop Distributed File system (HDFS) and other is Map Reduce[6]. HDFS provides Storage of data while Map Reduce provides analysis of data in clustered environment using maping and reducing functions. Hadoop is a programming paradigm that provide large scalability across thousands of server in a cluster that implements large data set. The main purpose developing this technique for a fast & efficient searching of data in the Map reduce paradigm of the hadoop distributed file system. In Feb. 2003 the hadoop technology introduce that first map reduce library by GOOGLE.In July 27th 2011, FACEBOOK has claimed that they had the largest hadoop cluster in the world with 21PB {1PB=1024TB} of storage. They have announced that the data grown up to 30PB. Firstly, YahooJoint at 2006 and deployed the large scale science clusters in 2007. Radoop is based on Rapid Minder while hadoop has attentions in data analytics.

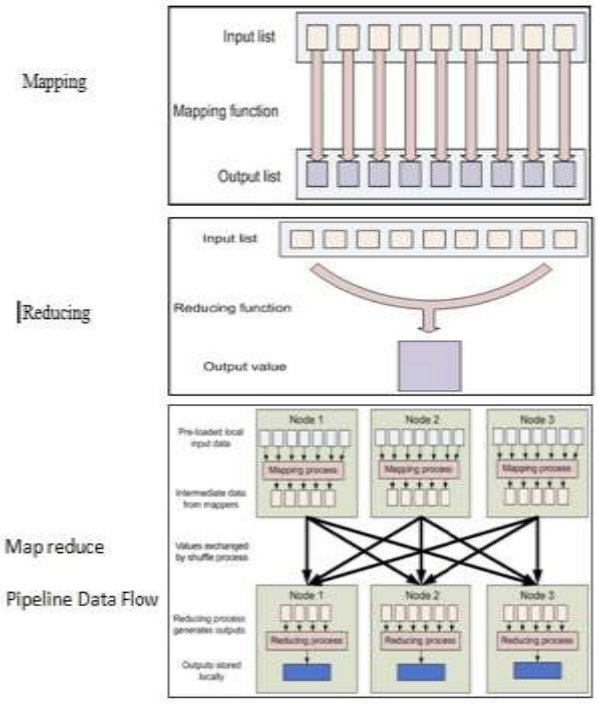

Fig.2. Map Reduce Architecture.

2.1 Hadoop Distributed File System [HDFS]

The HDFS is designed to run on clustered computing platform. One of the salient features of HDFS is that it is fault- tolerance to a very high degree and cost effective. It is based in the MASTER – SLAVE architecture. HDFS -Storage and replication of big data. HDFS is the Java portable file system which is more scalable, reliable, distributed in the Hadoop framework. A Hadoop cluster contains the combination of single Name node and group of Data nodes as shown in figure . The files are stored as default Block size of 64MB.HDFS provides redundant storage of large data with low latency when it perform operations such as "Write Once, Read Many Times". The Remote Procedure call based communication between the nodes. Name node stores metadata like the name, replicas, file attributes, locations of each block address and the quick query of metadata is put away in Arbitrary Access Memory by Metadata. It additionally diminishes the information misfortune and avoids debasement of the document framework. Name node just screens the quantity of block in data node and if any block lost or flopped in the imitation of a data node, the name node makes another reproduction of a similar block. Each block in the data node is kept up with timestamp to distinguish the present status. In the event that any disappointment happens in the node, it require not be repair instantly it tends to be repaired occasionally.

2.2 Related Work

2.2.1. HDFS Security Challenges

HDFS core layer of Hadoop Framework that contains different types of data and when its contain more sensitive data so that security issues arises. There is no appropriate access method to contollling security risk like stolen of data unauthorized data access in hadoop. The replicated data is additionally not anchor which needs greater security for shielding from ruptures and vulnerabilities. For the most part

Government Area and Associations never utilizing Hadoop condition for putting away significant information due to less security worries inside a Hadoop Technology. They are giving security in outside of Hadoop Condition like firewall and Interruption Discovery Framework. A few creators scrambled the data block and nodes utilizing encryption strategy however no flawless calculation is said to keep up the security in Hadoop Condition. So as to upgrade security a few methodologies are specified below.

2.2.2 HDFS Security Approach

A) Kerberos Mechanism

Kerberos [10] based on Master Slave ariticture and use as network authentication protocol from transfering any file by node over a non secure channel as a tool called ticket to providing unique identification between them. Kerberos[10] mechanism is used Remote Procedure Call [11] to enhance the security in HDFS. Token is used to authenticate a RPC connection between client(useing HTTP) and data node is Achieved using Block Transfer. .Ticket Granting Ticket (TGT) or Service Ticket are used to authenticate a name node by using Kerberos.Key Distribution Centre (KDC) issues advantage is even if the ticket is stolen by the attacker it can't be renewed.The data node also issues a token called Name Token where it allows the Name node to enforce permission for correct control access on its data blocks. Block Token allows the data node to identify whether the client is authorized access to access data blocks. These block token and Name Token is sent back to client who contains data block respective locations and you're the authorized person to access the location.

This paper in Proposed theme mentioned new Algorithm to enhance more security in the data nodes of HDFS.

III. PROPOSED HYBRID THEME

In addition to enhancement of security in hadoop framework based on HDFS file system this paper proposed theme mentioned new algorithm named as SECURE HASH ALGORITHM-3 [SHA-3].

3.1. Secure Hash Alogrithm [Sha-3] Approach

As bull eye algorithm[26] which are given for secure only sensitive information like credit card numbers, passwords, account numbers, personal details in hadoop storage but it consuming more time, effort and give narrow values. A bull eye algorithm is not fully trusted secure because it has own boundary criteria are not working for complex computation model. It's modify algorithm shrinking bull's eye algorithm were not simulated practically implemented. As Kerberos approaches [26] has their own limitations that when clients and data nodes are using HTTP connection and remote procedure call for communicating each other. Attackers must recording and tracking, snap capturing during unsecure communication in between client and data node. To prevent and more secure data transfer We have focus a new security approach SHA-3 which are deals any type of application releated to hadoop storage in more efficient and faster way than bull eye algorithm called as "SECURE HASH ALGORITHM-3 [SHA-3]". 64 candidates are participating the sha-3 build competition most from Europe and North America in 31 oct.2008. NIST [National Institute of Standard and Technology] are announced their 5 finalized which are:

BLAKE, Grøst1, JH, KECCAK, SKEIN

The SHA-3 is winning introduced in 2012. This Standard specifies the Secure Hash Algorithm-3 (SHA-3) family of functions on binary data. Each of the SHA-3 functions is based on an instance of the KECCAK algorithm that NIST selected as the winner of the SHA-3 Cryptographic Hash Algorithm Competition. This Standard also specifies the KECCAK-p family of mathematical permutations, including the permutation that underlies KECCAK. The SHA-3 family consists of four cryptographic hash functions, called SHA3-224, SHA3-256, SHA3-384, and SHA3-512, and two extendable-output functions (XOFs), called SHAKE128 and SHAKE256. Extendable output functions are different from hash functions but it is used in similar ways, with the flexibility to be adapted directly to the requirement of individual applications. The 5 finalized are provides 3 needs to give the best security:-

a) Introduce a formal adversarial model for a concrete security.

b) Define a security notion chosen by adversarial model.

c) Exhibit security reduction {break atomic primitive} for build a secure atomic primitive most known model is random oracle model".

To prove your system is highly secure 3 securities notion is must be firstly proof:-

1.) Collision resistance [coll]:-

It occurs because 2 distinct inputs M1 and M2 have generate same hash output

$$H(M1) = H(M2)$$

The result of the collision attack is a bit of difference due to "birthday paradox".

2.) Preimage resistance[pre]:-

It is hard to find because it's committed schema. The committed schema broken if the adversary succeeds to retrieve a massage M from a hash value

$$Y = H(M)$$

"The main problem behind this is its committed schema allows a prover to commit on data massage without revealing it verifier checks the correctness of it".

1.) Second preimage resistance[sec]:-

It is most hard thing to find because of their conditions.

When the collision event occurs the adversary trick an honest party by first asking him to sign a "harmless massage M". H (M) = H (M')

The adversary send a harmful massage M' with is massage which are break the security. Only SKEIN and Grøst 1 are work on second preimage resistance.

Grøst 1= ideal permutation model

skein= ideal cipher model

After the attack the 3 notion give the value about -

Collision resistance = n/2 bits (approximate)

Preimage resistance = n bits (approximate)

Second preimage resistance = n-L (approximate) [L is length of first preimage]

3.2. Merkle-Damgard Mode of Operation :-

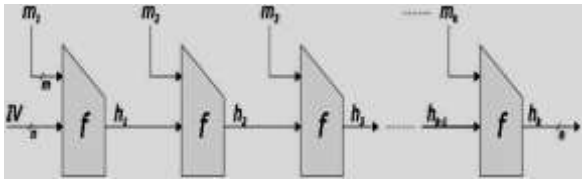

Fig.3. Using hash functions by markle – demgard.

The 5 finalist are given their security notions solution by reference of Merkle – Demagard mode. A message M is divided into message blocks M = m1jjm2jj : : : jjmk 1 jjmk of xed size. Further, the hash function is computed in the iterative manner: hi f(hi 1; mi) where f(hi 1; mi) = DESmi (hi 1) and h0 = IV . Finally, the hash function returns a hash value H(M) = hk. A colliding message is found if the rst input block is removed and for IV is selected h1. In addition, trivial preimage attacks are possible under the assumption that IV can be chosen by the adversary. Merkle and Damgard independently offered a solution to address these problems their idea was to x a default value for IV and to use a padding scheme with the message length appended at the end. Each of them offered a different padding scheme. Merkle's padding scheme emerged as standard due to its higher efficiency as the smaller number of padded bits is needed in the case of large messages.

3.3 Implementation by KECCAK:-

Keccak one of the finalized of competition use spong construction with markle damagard operations a combination of linear mixing operations and non linear operations.

1)SHA3-224(M) = KECCAK[448] (M || 01, 224);
2)SHA3-256(M) = KECCAK[512] (M || 01, 256);
3)SHA3-384(M) = KECCAK[768] (M || 01, 384);

4)SHA3-512(M) = KECCAK[1024] (M || 01, 512);
In each case the capacity is double the digest length I.e. **c=2d.**

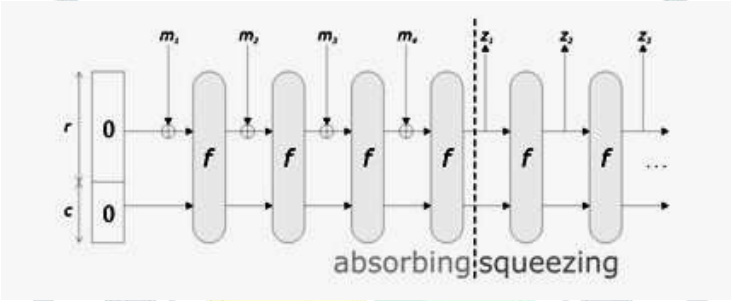

Fig.4. Spong Construction.

The sponge construction is a framework for specifying function on binary data with arbitrary output length.
Sponge function = SPONGE {f, pad, r}
The rate r is a positive integer that is strictly less than the width b. The capacity denoted by c, is a positive integer b-r. Thus r+c=b. The padding rule pad, is a function that produces padding. The padding rule of keccak is called “multi rate padding”. The parameters of keccak should be satisfy l=2n+m.
The Keccak hash function is proven indifferentiable from a random oracle up to bound ((Kq)2=2l m) if the underlying permutation is assumed to be ideal. Using this indifferentiability bound renders an optimal collision resistance bound for Keccak, AdvcolH = (q2=2n), as well as optimal preimage second preimage resistance bounds (q=2n).

	MODEL	Adv_f^{Coll}	Adv_f^{Pre}	Adv_f^{sec}	Adv_H^{coll}	Adv_H^{Pre}	Adv_H^{sec}
BLAKE	Ideal cipher E	$\Theta(q^2/2^n)$	$\Theta(q/2^n)$	$\Theta(q/2^n)$	$\Theta(q^2/2^n)$	$\Theta(q/2^n)$	$\Theta(q/2^n)$
Grøst1	Ideal permutations P,Q	$\Theta(q^4/2^l)$	$\Theta(q^2/2^l)$	$\Theta(q^2/2^l)$	$\Theta(q^2/2^n)$	$\Theta(q/2^n)$	$\Theta(q/2^{n-L})$
JH	Ideal permutation P	$\Theta(1)$	$\Theta(1)$	$\Theta(1)$	$\Theta(q^2/2^n)$	$O(q/2^n+q^2/2^{l-m})$	$O(q/2^n+q^2/2^{l-m})$
KECCAK	Ideal permutation P	$\Theta(1)$	$\Theta(1)$	$\Theta(1)$	$\Theta(q^2/2^n)$	$\Theta(q/2^n)$	$\Theta(q/2^n)$
SKEIN	Ideal block cipher E	$\Theta(q^2/2^l)$	$\Theta(q^2/2^l)$	$\Theta(q/2^l)$	$\Theta(q^2/2^n)$	$\Theta(q/2^n)$	$\Theta(q/2^n)$
NIST security requirement	Not specified	Not specified	Not specified	$O(q=2^n)$	$O(\lambda/m \cdot q/2^n)$	$O(q^2/2^n)$	Not specified

Fig.5. A Schematic summary of all security results of the SHA-3 finalists. The last row of the table gives a representation of the security requirements by NIST.

The use parameters n, l, m, and 2L denote the hash function output size the internal value size and the massage input size, the length of first preimage in massage block respectively. The first column indicates the name of hash function selected in final competition, while the second column describes the underlying assumptions. The next three column shows the security bounds on compression functions while the last three columns are summarize the security result on complete hash functions. JH last two columns indicates the existence of a non-trivial upper bound which is not yet optimal both 256 and 512bits variant. So we can say that JH compression function is insecure in ideal permutation model. Grøst1 last column show a security reduction while other result does not show any type of reduction in the table.

Security of SHA-3:-

For the security strength second preimage attack on a massage M, for function L{M } . the four SHA-3 hash functions are alternatives to the SHA-2 functions, and they are designed to provide resistance against preimage, second preimage and collision attack which equals or exceed the resistance that the corresponding SHA-2 function provide. The SHA-3 function are also designed to resist other attack such as length -extension attack that would be resisted by a random function of same output length ,providing security strength up to the Hash functions.

functions	Output size	Security strength in bits		
		collision	preimage	2nd preimage
SHA-1	160	<80	160	160-L{M}
SHA-224	224	112	224	MIN[224,256-L{M}]
SHA-512/224	224	112	224	224
SHA-256	256	128	256	256-L{M}
SHA-512/256	256	128	256	256
SHA-384	384	192	384	384
SHA-512	512	256	512	512
SHA3-224	224	112	224	224
SHA3-256	256	128	256	256
SHA3-384	384	192	384	384
SHA3-512	512	256	512	512
SHAKE128	d	MIN[d/2,128]	≥MIN{ d,128}	MIN{ d,128}
SHAKE256	d	MIN[d/2,256]	≥MIN{d,256}	MIN{ d,256}

Fig.6. Comparison between all SHA Standard for Security.

IV. Conclusions

This paper indicates Big Data security challenges specified to give thought an about the big data issues progressively. The security issue is directed more all together towards improving the security in big data. We can enhance security in big data by utilizing hybrid approach or by consolidating these methodologies in Hadoop Distributed File System which is the center part in Hadoop, where it contains a substantial number of blocks. These studied methodologies are overcome certain issues occur secure delivery of data between node to node. In Future these methodologies are likewise actualized in different layers of Hadoop Technology.

References

[1] Prof. Dr. Philippe Cudré-Mauroux, "An Introduction to BIG DATA", June 6, 2013 Alliance EPFL, http://exascale.info/
[2] Fremont Rider, "The future of the Research Library", http://crl.acrl.org/content/50/1/48.html
[3] RobPegler,"Introduction to big data, analytics knowledge and skill approach with various techniques,http://www.snia.org/sites/default/files2/ABDS2012/Tutorials/RobPeglar_IntroductionAnalytics%20_Big%20Data_Hadoop.pdf
[4] Gartner, http://www.gartner.com/newsroom/id/2848718, STAMFORD, Conn., September 17, 2014
[5] "Leveraging Massively Parallel Processing in an Oracle Environment for Big Data", An Oracle White Paper, November 2010.
[6] Jeffrey Dean and Sanjay Ghemawat, "Map Reduce: Simplified Data Processing on Large Clusters", Google, Inc.
[7] "Hadoop and HDFS:Storage for Next Generation Data Management", Cloudera, Inc, 2014.
[8] Data guise protect, http://www.dataguise.com/?q=dataguise-dgsecure-platform
[9] Parviz Deyhim, "Best Practices for Amazon EMR", August 2013.
[10] Al-Janabi, Rasheed, M.A.-S., "Public-Key Cryptography Enabled Kerberos Authentication", IEEE, Developments in E-systems Engineering (DeSE), 2011
[11] Heindel L.E, "Highly reliable synchronous and asynchronous remote procedure calls", Conference Proceedings of the IEEE Fifteenth Annual International Phoenix Conference on computers and communications,1996.
[12] The journey to big data, EMC2 Publications.
[13] Horton Technical Preview for Apache Spark, Horton works Inc.
[14] Shay Chen, "Application Denial of Service", Hack tics Ltd, 2007.
[15] Daniel J. Bernstein, "Understanding brute force", National Science Foundation, Chicago.
[16] Introduction to Pig, Cloud era, 2009.
[17] P. Victer Paul, N. Saravanan, S.K.V. Jayakumar, P. Dhavachelvan and R. Baskaran, "QoS enhancements for global replication management in peer to peer networks", Future Generation Computer Systems, Elsevier, Volume 28, Issue 3, March 2012, Pages 573–582. ISSN: 0167¬739X.
[18] Ashish Thusoo, Joydeep Sen Sarma, "Hive –A Petabyte Scale Data Warehouse Using Hadoop, Facebook Data Infrastructure Team.

[19] P. Victer Paul, D. Rajaguru, N. Saravanan, R. Baskaran and P. Dhavachelvan, "Efficient service cache management in mobile P2P networks", Future Generation Computer Systems, Elsevier, Volume 29, Issue 6, August 2013, Pages 1505–1521. ISSN: 0167-739X.
[20] N. Saravanan, R. Baskaran, M. Shanmugam, M.S. SaleemBasha and P. Victer Paul, "An Effective Model for QoS Assessment in Data Caching in MANET Environments", International Journal of Wireless and Mobile Computing, Inderscience, Vol.6, No.5, 2013, pp.515-527. ISSN: 1741-1092.
[21] R. Baskaran, P. Victer Paul and P. Dhavachelvan, "Ant Colony Optimization for Data Cache Technique in MANET", International Conference on Advances in Computing (ICADC 2012), Advances in Intelligent and Soft Computing" series, Volume 174, Springer, June 2012, pp 873-878, ISBN: 978-81-322-0739-9.
[22]https://www.ida.gov.sg/~/media/Files/Infocomm%20Landscape/Technology/TechnologyRoadmap/BigData.pdf
[23] Philip Russom, "Managing Big Data", TDWI research, Fourth Quarter 2013.
[24] Changqing, "Big Data Processing in Cloud Computing Environments", International Symposium on Pervasive Systems, Algorithms and Networks, 2012.
[25] Young-Sae Song, "Storing Big Data- The rise of the Storage Cloud", December, 2012.
[26] B. Sarala devi, N. Pazhaniraja, P. Victer Paul, M.S. Saleem Basha, "Big Data and Hadoop-A Study in Security Perspective", Procedia, Elsevier 2015.