IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727
PP 37-40
www.iosrjournals.org

A Review On Security Approach In Big Data

Ms. Chetana Girish Gorakh[1] Dr. Kishor M. Dhole[2]
[1]*Department of Computer science, VMV comm.JMT arts JJP science College Nagpur,(M.S.),India,*
Chetanagorakh@gmail.com
[2]*Assistant Professor, Department of Computer Science, S. K. Porwal College Kamptee, Nagpur,(M.S.),India.Km_phd108@yahoo.co.in*

Abstract: *Today we are living in the digital world, so everyday substantial amount of data generated and it is very tedious job to manage, acquire, access, deploy and store large scale of data. Traditional databases are not compatible to manage, store and analyze heterogeneous data generated every day. Hence it needs to study concepts which identify newest techniques, methods to provide high security of data configuration and extraction in big data environment.*
This paper presents an overview of big data concepts and characteristics. It discuss introductory scenario about tools used in big data environment. It also covers security issues for big data.
The term big data refers to the huge data configuration, distribution and analysis that overcome the drawbacks of traditional data processing technology. Big data which manage, store and acquire data very speedily and cost effective, involves various tools, technique and framework. Our main focus is on security issue in big data.
Keywords - *Big data, map reduce, Hadoop, security and privacy, big data analytics.*

I. INTRODUCTION

Big data refers to the databases whose size is beyond the normal size of traditional databases. It handles and measures in terabytes and zeta bytes. Big data has hardly ever generated by human being, whereas machine and sensors provides data.

Different technologies designed, used by human being itself to produce large amount of data every day. All data generated may be structured form (relational data), semi structured form (xml data) and unstructured form (document,pdf,image,audio,video,media logs, mri scan report, x-rays, etc.,) which is not managed by traditional databases (i.e. in rows and column). Big data is heterogeneous data. Social media like facebook, twitter, Google generates huge amount of data daily which is complex and unstructured in nature to handle by traditional databases [1].

II. CHARACTERISTICS

The term big data used very extensively to analyze wide range of data and characteristics of big data is incomplete without explaining 4 V's of big data. The 4 v's that define big data are volume, velocity, variety, veracity are discussed as follows [1].

a. **Volume** (Large scale of data): The volume of data concern about homogeneous size of data. It is increasing rapidly since last decade. User uploaded data on social media daily, financial transaction, and sensors data produces large volume of data.
b. **Variety** (Different form of data): Data comes to the system in different format i.e. variety of data generated like images,audio,video,daily post, bank transaction ,emails, graphics data. Data triggered in unstructured format.
c. **Velocity** (Analysis of streaming of data): Velocity refers "the speed at which data is generated" to deal with response to sensor, machine which generates data rapidly. The quick response to match the speed is new challenge for most of the organization. This speed requires enforces speedy analysis and analytics on top of the data.
d. **Veracity** (Uncertainty of data): The data whatever is generated, how would be user comes to know it is genuine. If huge amount of data considered in this context then there is no guarantee that the resultant data is genuine, reliable and usable.

III. TOOLS IN BIG DATA

Big data analytics refers to the acquisition, managing, extraction and analyzing huge range of data to process further. This helps to measures its performance and analysis of scaled data in it. It improves its efficiency and performance of the proposed system in user requirement area. For management of data Oozie,

EMR, Chukwa ,Flime ,Zookeeper in Hadoop is used and for data access Hive , Pig , Avro , Mahout ,Sqoop in Hadoop can be used. So new tools requires to study in big data environment like as Hadoop and map reduce [2].

a. HADOOP

In 2002 Google's was facing the problem of managing, storing, large scale of data, i.e., web addresses provide to save on server. Then it uses distributed file system to save data but problem was that google has to do it manually. Thus to save data automatically and overcome relevant problems in this regard , Google's wrote white paper based on this concept , which further used by yahoo to find the best solution on this issue

Hadoop is open source apache project donated by yahoo. Hadoop is a framework written in java which allows distributed processing of huge scale of data sets using programming model. Hadoop is distributed file system called Hadoop distributed file system (HDFS) and that programming model called Map reduce (MR).It basically help us to store large data sets of file in distributed file system and then with the help of programming model do analysis[2].

b. MAP REDUCE

Map reduce is a programming model and used to implement large data sets. The main task is to analyze, map, reduce, split, sort, shuffle, combine, spill, partition, merge of data sets. In map reduce function first takes the input key/value pair and produces the list of intermediate key/value pair and then group together all key/value pair and finally process to produce result. Map Reduce architecture depicts as shown in following figure 1.

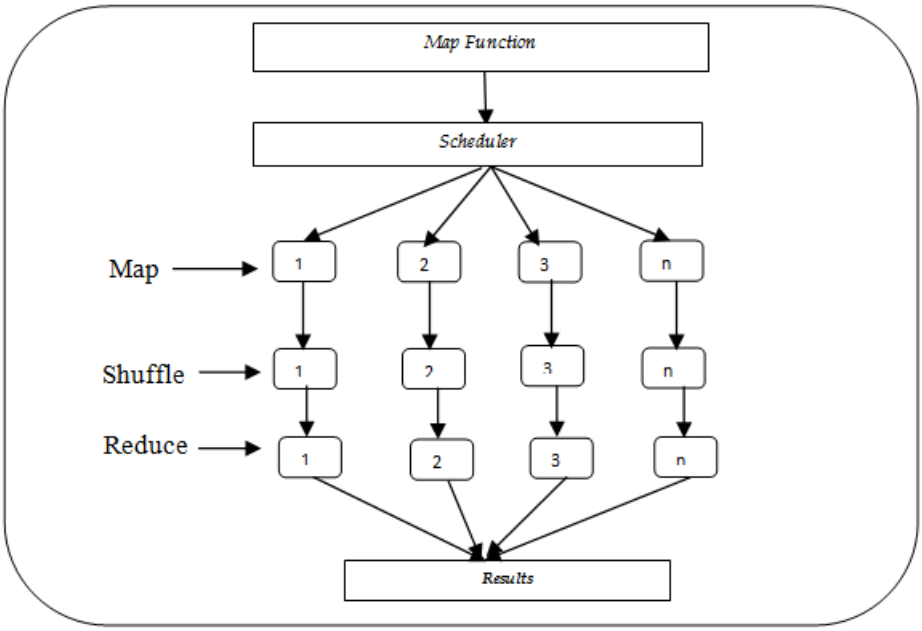

Figure 1: Map Reduce Function [Compiled by Researcher]

IV. SECURITY ISSUES IN BIG DATA

Security is a big challenge for big data. New technology, smart devices, sensors, mobile data and networks lead to produce extensive amount of data on large scale. In companies, business the term big data is using and most of that data may be confidential data or sensitive data. And it is necessary to provide security and restriction to access data. Hacking of data is most dangerous due to publically availability of large volume of data.

Privacy:

In large amount of data, there is also some personal or confidential data. So this data can be reuse by another sector. There should be limitation of extract, analyze and correlate potential sensitive data to shared one. This is new technology and many users are unaware of its uses and it introduces new vulnerability. Though implementation of big data increased but still user authentication and access to sensitive data is out of controlled. Most of the user are not comfortable with the idea about service providers are able to gather information like detail information, identification and purpose about the users likewise credit card, log files, location based data etc.,

Bigger dimension of data generation, deployment and control mechanisms in big data environment creates many more possibility to target that data by hackers, criminals to misuse it. It might be steals this personal data and sell it by these peoples. Henceforth, there is a need of research in this big data technology for securing data. Presently technologies for securing data are slow when applied to vast amount of data.

The four important security issues of big data are authentication level, data level, network level and generic issues [6].

Big data analysis implements advanced techniques using analytics and visualization mechanism for large data sets to control all hidden patterns and unknown correlations for accurate decision making situations.

V. CHALLENGES

Big data analytics needs various types of phases which include data acqusion, information extraction and cleaning, data integration, aggregation and representation, query processsing, data modeling and analysis, interpretation. Each of the above had challenges relevant with heterogeneity, scale, timeliness, complexity and privacy.

Unauthorized release of information, modification of information and denial of resources depicts the security violations. Security of big data can be improved by using techniques of authentication, authorization, encryption and audit trials. Following methods used for protecting big data to avoid security violations.

- Authentication method
- File encryption method
- Access control
- Key management
- Logging method
- Secure communication method

VI. ALGORITHM

As per literature study it has been observed that most of the researcher worked on the different algorithms in big data for security concerned. Following algorithmic comparative analysis covered with some factors. These factors are key length, block size, security rate and execution time as shown in following table 1.

1. **RSA (Rivest-Shamir-Adleman) algorithm**: Suppose any individual A wants to receive message M secretly will use pair of integers {e,n} as his public key also this A use {d,n} as his private keys. Another individual who wants to send message M secretly to A will use A's public key to encrypt a message and it will create cipher text C. Now only A can decrypt message M using his private keys. Where, cipher text C= (Me)*|n|.[6]

2. **ECC (Elliptic Curve Cryptography) algorithm**: Elliptic curve cryptography (ECC) is an approach to public key cryptography based on the algebraic structure of elliptic curves over finite fields. Elliptic curves are also used in several integer factorization algorithms that have applications in cryptography. The primary benefit promised by ECC is a smaller key size, reducing storage and transmission requirements, i.e. that an elliptic curve group could provide the same level of security afforded by an RSA-based system with a large modulus and correspondingly larger [6].

3. **DES (Data encryption standard) algorithm**: DES algorithm uses cipher key known as Feistel block cipher. DES expects two inputs - the plaintext to be encrypted and the secret key. The manner in which the plaintext is accepted, and the key arrangement used for encryption and decryption, both determine the type of cipher it is. DES is therefore a symmetric, 64 bit block cipher as it uses the same key for both encryption and decryption and only operates on 64 bit blocks of data at a time [6].

4. **AES (Advanced Encryption Standard) algorithm :** AES is new cryptographic algorithm that can be used to protect electronic data. It uses 10, 12 or fourteen rounds. Depending on the number of rounds, the key size may be 128, 192, or 256 bits. AES operates on a 4×4 column-major order matrix of bytes, known as the state. When encrypting data with a symmetric block cipher, which use block of n bits. With AES, n=128(AES-128, AES-192 and AES- 256 all use 128-bit blocks).This means a limit of more than 250 millions of terabytes. When encrypting data with a symmetric block cipher, which uses block of n bits. With AES, n=128(AES-128, AES-192 and AES-256 all use 128-bit blocks). This means a limit of more than 250 millions of terabytes [6].

Factors	DES	AES	RSA	ECC
Key length	56 bits	128,198,256 bits	Based on no of bits	135 bits
Block size	64 bits	128 bits	varies	varies
Security Rate	Not enough	Excellent	Good	Less
Execution Time	Slow	More Fast	Slowest	Fastest

Table 1: Comparative Study [6]

VII. ANLYSIS OF ALGORITM

AES algorithm is better than DES, RSA and ECC. But disadvantage of AES algorithm is sharing of key. There is no safe way to share the key. And there is also loss of data when we compressed large file so we are doing research work on security issue. These algorithms had some security issue related with key length, block size, security rate and execution time.

VIII. PROPOSED WORK

To overcome these issues there is need to identify new algorithmic techniques to provide high secure data and control mechanism in big data services. This motivates to work on security algorithms in big data. Security analytics in big data may give more authentic and reliable solutions for user's point of view. A big data application demands more efficient and cost based solution on security point of view.

In this context presently we started work to identify high privilege security algorithm for big data system. It proposed on big data analytics which imposes on the following three factors.

- Source of Data,
- Security based system using algorithm and
- Deployment of corrected data to end users.

IX. CONCLUSIONS

This paper introduced big data concepts and tools of big data such as hadoop and map reduce. In this era vast amount of data produced on daily basis and it becomes difficult to handle large data sets and analyze the pattern. In organization continue to collect large data and it is necessary to have security to that data. In previous work, algorithm has some drawbacks and there is huge scope to overcome these security issues and for future research work. This paper mainly covers the analysis of security algorithms used in big data services and security issue in big data. This imposes and motivates to identify new algorithms and techniques for security provisions in big data analytics for betterment of mankind.

ACKNOWLEDGEMENTS

I am very much thankful to Dr. Kishor M. Dhole, Department of computer science, S. K. Porwal College Kamptee, Nagpur University, for his encouragement and motivation in research work in big data.

REFERENCES

[1]. Sangita Bansal , Dr. Ajay Rana , Department of Computer Science and engineering Amity university , Noida (U. P.) India , transitioning From relational databases to big data , *International journal of advanced research in computer science and software engineering valume 4 , Issue 1, January 2014*

[2]. Punit singh Duggal and sanchita paul , Department of computer science and engineering , Birla institute of technology ,mesra , Ranchi ,India , Big data analysis : Challenge and solution ,*International conference on cloud , big data and trust 2013 , nov 13-15 ,RGPV.*

[3]. Raghav Toshniwal , kanishka Ghosh Dastidar ,Ashok Nath ,department of computer science ,st. xaviers college (autonomous) kollata,india , Big Data Security issue and challenge, *International Journal of Innovative in Advanced Engineering (IJIRAE) ISSN:2349-2163 ISSUE 2, Volume 2 (February2015).*

[4]. Venkata Narasimha Inukolu , sailaja Arsi and Srinivassa Rao Ravuri, Department of computer Engineering ,texas tech university ,USA Department of banking and financial services cognizant technology solution ,india ,*International journal of network security and its application (IJNSA), vol 6 no. 3 May 2014*

[5]. Big_Data_Analytics_for_Security_Intelligence.pdf

[6]. *Vinit Gopal Savant ,Department of computer engineering, ,Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India. ,vinitsawant06@gmail.com, Approaches to Solve Big Data Security Issues and Comparative Study of Cryptographic Algorithms for Data Encryption, International Journal of Engineering Research and General Science Volume 3, Issue 3, May-June, 2015, ISSN 2091-2730.*

[7]. Jainendra Singh ,Department of Computer Science, Maharaja Surajmal Institute C-4, Janakpuri, New Delhi, INDIA, Big Data Analytic and Mining with Machine Learning Algorithm , International Research Publications House http://www. irphouse.com /ijict.htm International Journal of Information and Computation Technology. ISSN 0974-2239 Volume 4, Number 1 (2014), pp. 33-40.