Analytical Study of Big Data Security Issues and It's Tools and Techniques

[1]U. S. Junghare
Assistant Professor
Dept. of Computer Science
Shri Shivaji Science College,
Amravati, MS. India,

[2]Dr.H.M.Deshmukh
Associate Professor
Dept. of Computer Science
Shri Shivaji Science College
Amravati, MS. India,

[3]Mr..Y.V.Hushare
Assistant Professor
Dept. of Computer Science
Shri Shivaji Science College
Amravati, MS. India,

Abstract: Now a day big data is most crucial aspect for industries, education, hospitals, governments etc. Big data analytics is the complex process of examining large and diverse data sets orbig data to uncover information including hidden patterns, unknown correlations, market trends and customer preferences that can help organizations make informed business decisions. Traditional techniques are not adequate for the analysis of the big data. There are many analytical tools and techniques available for big data analysis. As big data is one of the critical aspects so security of big data is vigorous issue because detecting and preventing fraud in data, datalost, error prone data, monitoring of real time data etc. are challenging tasks. With this approachpaper comprises analytical study onbig data security tools and techniques which is need of current research and development.

Keywords: Bigdata, Data analytics, Security, Security Tools.

I. INTRODUCTION

Big data requires a set of techniques and technologies with new forms of integration to reveal insights from datasets that are diverse, complex, and of a massive scale. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source. Big data is in focus not only because of its demand but for its privacy and security issues. Attacking on big data tools are increasing now a day so security of big data is a challenging task. Day to day data quantity is increasing with various formats but due to limited bandwidth of network and less storage and computation pawer security of data is the challenging task.

II. SECURITY ISSUES FOR BIG DATA

Security issues of Bigdata belongs to following category,

2.1 Data while transmission on social network

- Limited bandwidth- As bigdata is big in volume, velocity and variety so limited bandwidth is a major issue for cloud data.
- Data Attacks - Big Data may be attacked by malicious users [4]
- Access Controls- To restrict the non-desirable access, it is significant to provide access control for big data [2,3].

2.2 Data on System

- Less storage Capacity of system is critical issue for big data.
- Less or insecure Computation Power: Due to less and insecure computation power there is a chance of losing data or data corruption [3].
- Data redundancy is also a serious issue regarding big data [1].
- Infrastructure security [3] It belongs to the security of tools and technology used for bigdata. Also, while dealing with bigdata authentication is also important. Only authentic users can process data.
- Anonymization of data- It is very vital to protect data from anonymizing it by removing sensitive data or hiding data. [3]
- Integrity of data- Data can be in structured and non-structured form so streaming of data also one issue [4].

III.BIG DATA SECURITY TOOLS

3.1 Hadoop: It is Java based programming framework. It uses client server structure, in which server process large amount of data and applications run on number of client nodes. It handles number of nodes and uses distributed network. It increases the data transfer rate and reduce the frequency of system failure[6]. Main task of Hadoop Distributed File System (HDFS) is to improve reliability by overcoming nodes failure.

Hadoop is highly scalable tool for bigdata. Itstores and processes the large and complex unstructured dataset. It performs the fast processing of bigdata. Its redirecting technique enables to response system in real time without failure [8].

3.2 NoSQL database: Non-relational data base is used to accumulate big data which handle the challenges of big data analytics security. NoSQL data base covers the security surrounded in middleware. It does not provide the explicit security. In NoSQL database complex integrity constrains cannot be instructed. In NoSQL authorization provides at higher level only. It means NoSQL database have security issue.

3.3 MapReduce: It is one of the software tools which process large unstructured dataset in cluster. It accomplishes two task one is mapping and other reducing. Map convert the input data into intermediate pair and reduce operation performs the sorting of intermediate pair with matching key. Reducing is done with three phase shuffles, sort and reduce [5].

MapReduce framework consist of mapping master node which divides input into smaller part and assign it to workers node. Then partitioning and mapping on input data is done and also store that data. Master node instructs the worker node to reduce data to get output [8].

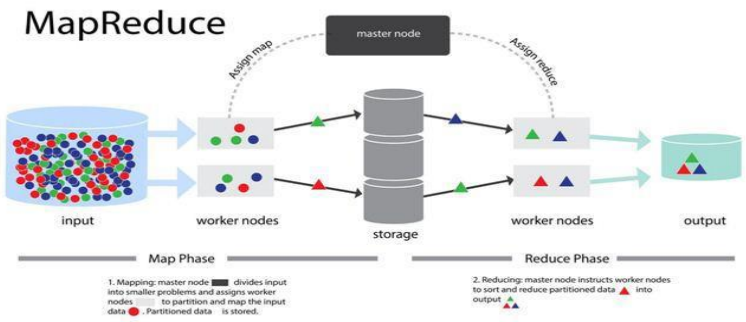

Fig.1 MapReduce

IV. PRIVACY AND SECURITY TECHNIQUES FOR BIG DATA

4.1 Data Filter and Compression: Maximum data produced and collected from smart productions are from various sensors which are linked to various machines and tools. Data can take more storage space and effect on data analysis process due to redundant, useless or error data. So, to avoid redundancy, to discard the useless data and to capture specific information, compression and filter techniques are used [1].

4.2 Data Cryptography / Encryption: Data encryption technique presents smaller well-define attacking surface. It is very dangerous to adapt side channel attacks and take out secrete key. It should be difficult for opponent to recognize plane text from cipher text. Cryptographic protocol makes possible searching and filtering of encrypted data. This protocol must ensure that opponent should not copy data that came from declared source [9].

4.3 Privacy Preserving Clustering: This technique is based on distance metrices. Clustering is carried out by considering different factors like size of data, types of data, shapes of the cluster, time, number of clusters etc. Various clustering methods are carried out for bigdata such as partitioning method, hierarchical method and privacy on clusters [5].

In Partitioning similar data objects are group into related partition and cluster is found using various clustering algorithms such as weighted k-mean, k-medoids, k median etc.

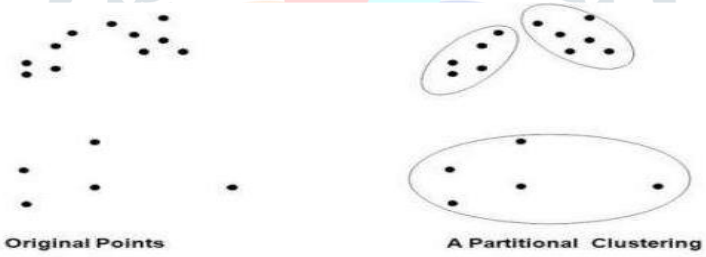

Fig.2 Cluster Partitioning

In hierarchical clustering tree like clusters are formed. Clusters are outline recursively or iteratively portioning tea dataset. Grouping of data objects are done by using top down and bottom up approach.

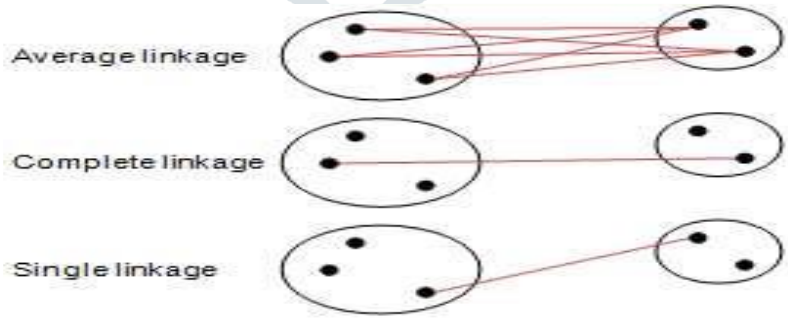

Fig. 3 Hierarchical Clustering

4.4 AuthorMohammed S. Al-Kahtani [7] focus on big data network security by introducing distributed architecture of big data network. It discovered big data network intrusion detection, network threat monitoring systems based on MapReduce machine-learning methods, and flow-based anomaly detection. These are specific security areas for network.

V. ANALYSIS

Table 5.1: Analysis of Security issue, its tools and techniques

Security Issues	Security Tools and Techniques	Benefits
Privacy	Clustering	Deal with missing data in dataset
	Map Reduce	Protect numerous amounts of sensitive data
Monitoring Real time data	Hadoop Clustering	Works for public cloud, secures the interconnected nodes
Authentication	Digital signature	Denied unauthorized access of data, protect data from hackers,
Data storage and system failure	Distributed network and distributed computing system	Increases the data transfer rate and reduce the frequency of system failure
Data redundancy and data storage	Data compression and filter	Avoid redundancy, discard useless and inconsistent values

VI. CONCLUSION

While gathering, storing and analyzing big data it requires extra efforts. In this paper, we studied big data security and privacy issues and discuss regarding its tools and techniques. As per study it is observed that, network traffic should be encrypted with suitable standards; access to devices should be checked; employees should be authorized to access systems; analysis should be done on anonymized data; communication should be made for the secure channel to prevent leakage, and network should be monitored for threats.

It is the future need to deliberate more on big data privacy, safety and security. So that more techniques, technologies and solutions need to be developed or existing technologies should be improved for accurate results.

VII. REFERENCES

1. Alice Joseph, Dr. Mathew Cherian,, March 18, "Big Data Security and Privacy in Smart Industry", International Journal of Computer Engineering and Applications, Volume XII, Issue III, ISSN 2321-3469.

2. Renu Bhandari, Vaibhav Hans and Neelu Jyothi Ahuja,"Big Data Security – Challenges and Recommendations", International Journal of Computer Sciences and Engineering Open Access, Volume-4 , Issue-1, E-ISSN: 2347-2693, pp 93-98.

3. Julio Moreno, Manuel A. Serrano and Eduardo Fernández-Medina, September 2016, "Main Issues in Big Data Security", Future Internet Article, Published: 1 , pp 1-16.

4. P.Joseph Charles, I.Carol, S.Mahalakshmi, Feb-2018, "Big Data Security an Overview", International Research Journal of Engineering and Technology (IRJET), Impact Factor value: 6.171, Volume: 05 Issue: 02, pp-130-134.

5. Anju Abraham ,Shyma Kareem , 2018, "Security and Clustering Of Big Data in Map Reduce Framework: A Survey", International Journal of Advance Research, Ideas and Innovations in Technology, ISSN: 2454-132X, Impact factor: 4.295, Impact factor: 4.295, pp 199-203.

6. "A Research on Big Data Analytics Security and Privacy in Cloud, Data Mining, Hadoop and Mapreduce", April 2018, Shreyas Satardekar Int. Journal of Engineering Research and Application, ISSN: 2248-9622, Vol. 8, Issue4 (Part -III), pp 65-78.

7. Mohammed S.Al-Kahtani, February 2017, "Security and Privacy in Big Data", International Journal of Computer Engineering and Information Technology, VOL. 9, NO. 2, E-ISSN 2412-8856.

8. K.Valli Madhavi, Dr.Y.Venkateswarlu, Varsha Sharma, Feb. 2018, "Big Data Analytics for Security to Obtain Actionable Intelligence in Real Time", Special Issue Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, pp 126-130.

9. GetanehBerieTarekegn, July–Aug 2016, "Big Data: Security Issues, Challenges And Future Scope", International Journal of Computer Engineering & Technology (IJCET), Volume 7, Issue 4, pp. 12–24.