A Review of Big Data Security Issues and Approaches

Jagruti Parmar[1] Prof. Jalpa Patel[2]
[1]Department of Information Technology [2]Department of Computer Information Technology
[1,2]Shri S'ad Vidya Mandal Institute Of Technology, Bharuch, India

Abstract— With the arrival of technology, there has been exploding increase in the generation of data. The data exists in different formats like text, audio, video, image referred as big data which is a large set of data. Due to this constant increment and habit on web requires security towards the private data. When this large amount of data travels through the internet then there is the problem of securing, managing, storing and analyzing the data. The major issue then arises is the privacy and security of big data. The importance of Big Data does not revolve around how much is the amount of data, but what to do with it. Data can be taken from various sources and analyze in order to find answers that enables cost reduction ,time reduction, new product development and optimizing offering, smart decision making. Then, a big question arises is- 'what security and privacy technology is adequate for controlled assured sharing for efficient direct access to big data'. This paper presents a quick review of the security issues and various approaches to overcome them Utility Miner) Using estimated utility Co-occurrence Pruning.

Key words: Big Data; security and privacy; value of big data

I. INTRODUCTION

Big data is the collection of data sets which is very large in size as well as complex. It contains the data which may be structured, semi-structured or unstructured. Generally, size of data is up to Petabytes or Exabytes. Big data have three important properties like volume, velocity and variety. Recently, there has been increasing interest in Big Data. However, the term Big Data remains unclear [1].

Recently, there has been an increasing interest in Big Data. However, the term Big Data remains vague. In Wikipedia, Big Data is an all-inclusive term for any collection of data sets so large and complex that it becomes difficult to procedure using traditional data handling applications. A widely recognized definition belongs to IDC: "big data technologies define a new age group of technologies and architectures, designed to economically extract value from huge volumes of a wide-ranging variety of data, by enabling the high-velocity capture, discovery, and/or analysis" [2].

Big data refers to collection of massive data with processing and data retrieval. As big data collects important and sensitive data from social sites and from government sites as a result of which security issues must be concerned. This collected data have to encrypt by using appropriate algorithms to secure the data. If the Internet age threatened security and privacy, the age of Big Data endanger them even more. Big Data security usually is to the use of the Big Data to implement solutions increasing security, reliability, and safety of a distributed system. Big Data privacy focuses on the protection of Big Data from unauthorized use and unwanted inference [3].

The current era is the "Age of Big Data" [4]. In the past few years, the total amount of data created by human has exploded [4]. From 2005 to 2020, the amount of data is predicted to increase 300 times, from 130 hexabytes to 40,000 hexabytes [5]. These data are generated from scientific research, finance and business informatics, government, Internet search, social networks, document, photography, audio, video, logs, click streams, mobile phones, sensor networks and so on. Big Data is the result of the dramatic increase of data [1].

Section II presents a brief summary of big data. Section III contains categorization of big data concerning security and privacy. The results obtained with security and privacy issues in big data are discussed in Section IV, and section V explains how to use big data to maintain security. Section VI considered big data analytics and section VII shows techniques for big data. Section VIII presents literature survey. Finally the conclusion highlights the importance and requirements to secure big data communication.

II. DEFINITION AND CHARACTERISTICS OF BIG DATA

Big data refers to large and complex datasets that typical software is insufficient for managing [2]. There are various explanations of big data via V's. Figure 1 illustrated the 5'Vs are typically used to characterize of Big Data as volume, velocity, variety, veracity and value [4]. These includes-Volume is the size of data; velocity is the high speed of data; variety indicates heterogeneous data types and sources; veracity describes stability and reliable of data; and value provides outputs for gains from large data sets.

Volume refers to the vast amounts of data generated each and every second. Just think through of all the emails, twitter messages, Facebook posts, photos, video clips, sensor data etc. we produce and share alternate. It's not only Terabytes but its Zettabytes or Brontobytes. On Facebook only per day we send 10 billion messages, click the like button in 4.5 billion times and upload 350 million new pictures daily. If take all the data generated in the world between the beginning of time and 2008, the similar volume of data will soon be made every minute! This gradually makes datasets too large to collection and study using traditional database technology. With big data technology we can now collection and usage these datasets with the help of distributed systems, where portions of the data is kept in different places and carried out by software [3].

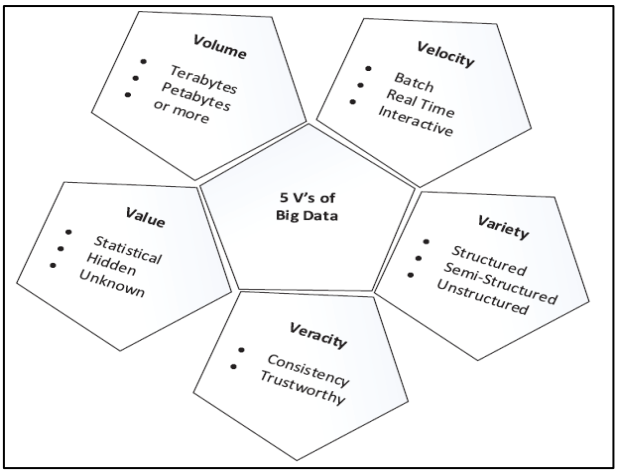

Fig. 1: 5V's of Big Data [2]

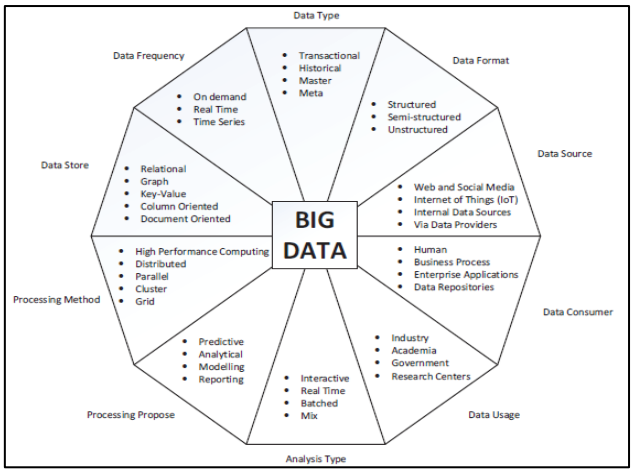

Fig. 2: Big data classification [2]

Velocity refers to the speed at which new data is generated and the speed at which data changes everywhere. Just reflect of social media messages successful viral in seconds, the speed at which credit card transactions are tested for fraudulent actions, or the milliseconds it takes transaction systems to analyze social media networks to collect signals that activate choices to buy or sell shares. Big data technology allows us now to analyze the data while it is produced, without constantly setting it into databases [3].

Variety refers to the different kinds of data that can use now. In the past, we concentrated on structured data that carefully turns into tables or relational databases, such as financial data. In fact, now 80% of the world's data is unstructured, and so can't simply be set into tables (reasons of photos, video orders or social media bring up to date). With big data technology we can now attach differed types of data (structured and unstructured) including messages, social media discussions, photos, device data, video or voice records and take them together with more traditional, structured data [3].

Veracity refers to the disorderliness or reliability of the data. With many arrangements of big data, feature and correctness are less manageable (e.g., Twitter posts with hash tags, abbreviations, typos and colloquial speech as well as the consistency and correctness of content) but now big data and analytics technology permits us to work with these type of data. The volumes frequently structure for the lack of feature or correctness [3].

Value is all well and good having access to big data but without we can crack it addicted to value it is unusable. So you can securely discuss that value is the main V of Big Data. It is significant that productions create a commercial case for any challenge to save and influence big data. It is so easy to decrease the high setup and board on big data edges without a clear kind of rates and profits [3].

Identifying characteristics of the data is helpful in extracting its hidden patterns. Big data is classified into ten categories in terms of data type, data format, data source, data consumer, data usage, data analysis, data store, data frequency, data processing propose, and data processing method[2].Figure 2 illustrated the classification of Big Data.

III. Security and Privacy in Big Data

Probably the most challenging and concerned problem in Big Data is security and privacy. Managerial agencies, healthcare industry, biomedical researchers and private businesses invest enormous resources into the group, combination and allocation of huge quantities of personal data for the vast benefit of Big Data [21].

A. Challenges of Big Data Security and Privacy

The security and privacy issues which should be concerned in Big Data context include [1]:

1) The personal data of a person when combined with external large data sets leads to the suggestion of new details about that person and it is possible that these types of details about the person are private. The person might not need the Data Owner to know or any person to know about them;
2) Information about the users is collected and used that add significance to the business of the organization. This is done by creating visions in their lives which they are uninformed of;
3) Another important significance arising would be social stratification where a well-educated person would be taking advantages of the Big data predictive analysis and on the other hand unfortunate will be definitely identified and treated not as good as;
4) Big Data used by law application will increase the chances of assured tagged people to suffer from adverse moments without the ability to fight back or even having knowledge that they are being distinguished [1].

The arena of privacy in big data which contains a group of challenges involves interaction with individuals, reidentification attacks, probable and provable results, and economic effects [1]. Interaction with individuals includes providing transparency, getting consent, revocation of consent and deletion of personal data. Re-identification attacks which have three sub-categories named correlation attacks, arbitrary identification attacks and targeted identification attacks mean that a huge dataset available is clearly scanned for associations that top to a unique fingerprint of a single individual. Probable and provable results refer the validity of the results gathered in big data. Economic effects of the big data paradigm are direct results of the exchange of datasets among business partners in advance [1].

B. Big Data Security

Whenever big data is discussed, the first thing to be considered is the data security. As Big data has all types of data whether relevant or irrelevant, private or public. There is no need to secure public data; the security concern arises when there is the involvement of private data. Big Data often contains sensitive information that wants to be secure from unauthorized access and release. The initiation of Big Data presented challenge in terms of security of data. There is a need of research and approaches that can handle this huge amount of data securely and efficiently. However many new technologies and methods have been established, but to some range they get slowed down when there is a participation of large amount of data. Table 1 depicts the comparison of various security algorithms.

IV. BIG DATA SECURITY ISSUES

Big Data security issues exist at different levels [1]:

1) Authentication level issues: In a cluster dissimilar nodes with different importance and rights are present. Accessing any kind of data is permitted only to the nodes with administrative rights. If in any case, malicious node gets administrative property then it may alter or steal the secured data. So as to avoid unauthorized access from malicious node, logging is used. Logging acting an essential role in Big Data which is used to have an appreciation on all the activities and record them. If there is absence of logging then no activities will be recorded and data may be modified or deleted.

Algo ri-thms	CONTRIBU TOR	BLO CK SIZE	KEY LENGT H	SECU -RITY RATE	EXEC U-TION RATE
ECC	Neil, Victor	Varie s	135 bits	Less	Fastes t
RSA	Rivest, Shamir, Adleman	Varie s	Based on no. of bits	Good	Slowe st
DES	IBM75	64 bits	56 bits	Not enoug h	Slow
3-DES	Derived fom DES	64 bits	168, 112, 56 bits	Weak	Slow
AES	Rijman, Joan	128 bits	128,192, 256 bits	Excell ent	More fast
BLO W FIS H	Bruce Schneier	64 bits	32 bits to 448 bits	Good	Fast

Table 1: Comparison Of Security Algorithms

2) Data level issues: Data acting a very important role in Big Data. Data is no extra than the information with inconstant priorities. Information found from the social networking sites, data from various government sites is growing fast. These issues basically deal with data integrity and availability like data protection and distributed data. So that improves efficiency, big data environments such as Hadoop, HDFS store the data as it is without encryption. If unauthorised access is probable then it may alter or modify the data. Data is kept in lots of nodes with replicas for quick access in distributed data storage.
3) Network level issues: A number of nodes are contemporary in a cluster which does processing of data computations. This data processing can be done by any of the node in a cluster. It will then become tough to discover which node is active in processing. This arises to a problem that to which node security should be delivered. Nodes can interconnect to each other via network. RPC (Remote Procedure Call) tools can also be used for communication but is uncertain as it is not encrypted.
4) General level issues: In Big data various technologies are used for processing. For security purposes, traditional tools are also used, but these tools may not accomplish well with new distributed environment of big data. Big data uses many new technologies for data storage, processing and retrieval. There may be complexities in using these traditional technologies.

V. APPROACHES TO TACKLE ISSUES SECURITY

The approaches are tackling the security issues of Big data. These are as follow[1]:

1) Data encryption: It is the transformation of simple message into an encrypted form that is conversion of plain text into cipher text and is for data level issues. Several encryption algorithms such as AES, RSA, DES, and ECC algorithm are used which uses private keys to secure data. Encryption is done at sender's side and decryption is done at receiver's side. If the data is encrypted then hacker catches it tough to take or modify data. For encryption/decryption process, two types of algorithms are basically used: Symmetric key cryptography and Asymmetric key cryptography.
    a) Symmetric key cryptography: This algorithms use the same key for both encryption and decryption. Examples are Data Encryption Standard (DES) and Advanced Encryption Standard (AES).
    b) Asymmetric key cryptography: This algorithms use different keys for encryption and decryption. Examples are Rivest-Shamir-Adleman (RSA) and Elliptic curve cryptography (ECC).
2) Network encryption: This approach is for network level issues. In the meantime network is used for communication then network must be encrypted. With appropriate algorithm if network is encrypted then hacker is incapable to crack the network and steal the data. Then data over the network will be secure enough from unauthorised access. RPC (Remote Procedure Call) tools can also be used for communication but is insecure as it is not encrypted. RPC is secure only if it is used with some other techniques like SSL. SSL (Secure Socket Layer) is a standard security technology for founding an encrypted connection between a server and a client. When data is sent between browsers and web servers is in the form of plain text which is helpless to snooping. If an attacker is rather able to intercept all data being sent the information can be seen and used by them. In order to generate an SSL connection, a web

server requires an SSL Certificate. With the aim of activate SSL on web server, number of questions about the identity of website and company will be asked. The Web server then creates two cryptographic keys- a private key and a public key. The Public Key does not necessity to be secret and is located into a Certificate Signing Request (CSR) that a data file also containing details. Then CSR will be submitted. During the SSL Certificate application process, the Certification Authority will validate details and issue an SSL Certificate having details and permits using SSL. Then web server will match issued SSL Certificate to the Private Key. The Web server will at that time be able to establish an encrypted link between the website and customer's web browser.

3) Logging**:** In order to prevent unauthorised access from malicious node, logging is used. Logging plays an important role in Big Data and used is for authentication level issues. It is used to keep an eye on all the activities and record them. If there is absence of logging then no activities will be recorded and data may be modified or deleted. If single log is continued by all nodes then each activity is recorded and malicious node can be detected easily.

VI. Big data Analytics for Security

Big data analytics aims to obtain beneficial information from large scale and complicated data [7]. The increase of stored or streamed data and development of analysis systems has led to using these activities in information security. The anomaly detection, intrusion detection, fraud detection, advanced persistent threats (APT) detection, and forensics from big data has been accomplished by examining the logs, system events, network traffic, website traffic, security information and event management (SIEM) alerts, cyber-attack patterns, business processes and other information sources. To detect these attacks, large volume and variety of data is accumulating and associate with network history. The advantageous uses of big data, such as performing without deletion of logs after a certain period, running complex queries on large and unstructured datasets, and facilitating human computer interactions via visual interfaces, for security is becoming quicker and cheaper than traditional methods. There is no need to delete the cancelled accounts or old logs as they can be used for the purpose of forensics later. In addition, real time and agile decision support applications, automatic defense and risk reduction systems, prediction of attack, determining of zero-day attack duration and tracking of attackers can be developed by analyzing suspicious and malicious patterns from information security data [7].

A method to detect malware using big data has been proposed. For this purpose, Large Iterative Multitier Ensemble (LIME) classifiers have been used to handle big data. The performance of LIME classifier with other base classifier was evaluated. The results showed that the presented method performed better than base classifiers. In another study, a framework with big data environment has been suggested such as big data analytic tool and NoSQL database for android application security assessment. The white-box, black-box and mobile environment forensic approaches have been used to determine security assessment level. Then authors inserted assessment results into Couch DB, which is one of the NoSQL database. Using this database, they tried to discover security issues or visualization with big data analytic tools like Sckit, Matplotlib. Finally, they used SOA controller to publish their results via web service [7].

VII. Techniques for Big Data

Big data has great probable to produce useful information for businesses which can benefit the way they manage their problems. Big data analysis is becoming indispensable for automatic find out of intelligence which elaborates in the frequently arising patterns and hidden rules. These huge datasets are too large and complex for humans to well extract useful data without the help of computational tools. Developing technologies like the Hadoop framework and MapReduce offer new and exciting techniques to process and transform big data, well-defined as complex, unstructured, or large amounts of data, into important knowledge [16].

A. Hadoop

Hadoop stands for Highly Archived Distributed Object Oriented Programming which is a scalable, open source, fault tolerant virtual grid operating system architecture for data storage and processing [16]. It runs on goods hardware, uses HDFS (Hadoop Distributed File System) which is fault-tolerant high bandwidth clustered storage architecture. It turns MapReduce for distributed data processing and is works with structured and unstructured data [16]. For handling the velocity and heterogeneity of data used of tools such as Hive, Pig and Mahout which are parts of Hadoop and HDFS framework. Hadoop and HDFS by Apache are broadly used for storing and handling big data.

Hadoop has distributed file system, data storage and analytics platforms and a layer that handles parallel computation, rate of work flow and conformation administration. An HDFS inning through the nodes in a Hadoop cluster and composed links the file systems on lots of input and output data nodes to create them into one big file system. The present Hadoop ecosystem, as shown in Figure 3, consists of the Hadoop kernel, MapReduce, the HDFS and a number of connected components like Apache Hive, HBase, Oozie, Pig and Zookeeper which are explained as below [16]:

- HDFS: A highly faults tolerant distributed file system which is responsible for storing data on the clusters.
- MapReduce: A great parallel programming technique for distributed processing of huge amount of data on clusters.
- HBase: A column oriented distributed NoSQL database for random read/write access.
- Pig: A high level data programming language for analyzing data of Hadoop computation.
- Hive: A data warehousing application that provides a SQL like access and relational model.
- Sqoop: A project for transferring/importing data between relational databases and Hadoop.
- Oozie: An orchestration and workflow management for dependent Hadoop jobs.

Figure 4 gives an overview of the Big Data analysis tools which are used for efficient and detailed data analysis and management works. The Analysis and management setup of Big Data can be understood through the layered structured defined in the figure. The data storage portion is controlled by the HDFS distributed file system architecture and other architectures available are Amazon Web Service, Hbase and CloudStore etc. The data processing tasks for all the tools is Map Reduce which is the Data processing tool which effectively used in the Big Data Analysis [16].

As above mentioned for handling the velocity and heterogeneity of data, tools such as Hive, Pig and Mahout are used which are parts of Hadoop and HDFS framework. Note that for all the tools used, Hadoop over HDFS is the basic architecture. Oozie and EMR with Flume and Zookeeper are used to management the volume and veracity of data, which are standard Big Data management tools [16].

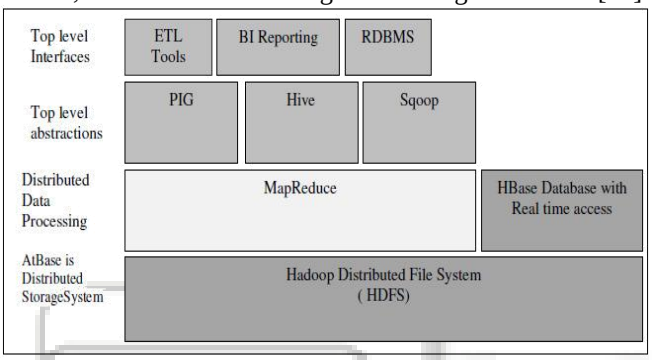

Fig. 3: Hadoop Architecture Tools [16]

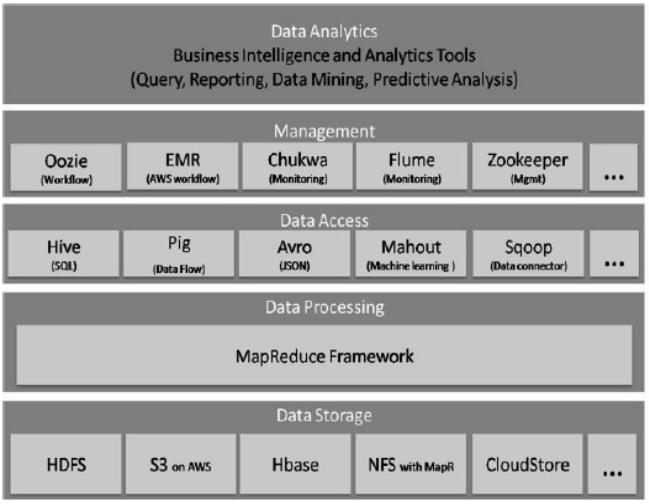

Fig. 4: Big data analysis tools [16]

B. Map Reduce

Map Reduce is a programming model for processing large data sets with a parallel distributed algorithm on a cluster. Map Reduce of Hadoop is a programming model and software structure for writing applications that rapidly process huge quantities of data in parallel on large clusters of add nodes.

The Map Reduce consists of two functions: map() and reduce(). Mapper achieves the responsibilities of cleaning and sorting and reducer achieves the responsibilities of brief the result. There may be many reducers to parallelize the combinations. Users can implement their own processing logic by specifying a customized map() and reduce() function. The map() function is taking an input value pair and creates a list of middle value pairs. The MapReduce runtime system collections collected all middle pairs based on the middle keys and permits them to reduce() function for creating the final results. Map Reduce is extensively used for the Analysis of big data [16].

Large-scale data processing is a tough mission. Handling hundreds or thousands of processors and handling parallelization and distributed environments creates it harder. Map Reduce offers solution to the mentioned issues since it supports distributed and parallel I/O scheduling which fault tolerant and supports scalability and it has inherent procedures for status and observing of heterogeneous and large datasets as in Big Data [16].

VIII. LITERATURE SURVEY

Sr. No	Public-ation / Year	Title	Review / Techniques	Advantages	Disadvantages/ future work
1.	IEEE 2013	Big Data: A Review [6]	This paper present an overview of big data's content, scope, samples, methods, advantages and challenges and discuss privacy concern on it. Techniques : MapReduce Hadoop	It considered all the connected information about Big Data, so we know the current status of Big Data.	This paper clearly has not resolved the entire subject about this substantial topic.
2.	IEEE 2015	Big Data Analysis: Issues and Challenges [7]	This paper presents the comparison between traditional database and big data. Mentioned theissues and	It provides the comparison of data mining and big data which is not same but	In future work, the researcher will have to deal with many of challenges of big data.

			challenges of big data and the tools currently used to implement and analyze the big data are detailed. Techniques : MapReduce MyHADOOP and MongoDB tools	co-related with each other.	
3.	IEEE 2013	Big Data Analytics for Security [8]	This paper is focus on big data analytics and challenges for security. Techniques : Hadoop ecosystem (including Pig, Hive, Mahout, and RHadoop) Stream mining, complex-event processing, NoSQL databases	It includes all security challenges of big data analytics.	Big data is changing the landscape of the security technologies foe everywhere i.e. it is used. So in future search and work on the privacy issues and its approaches with to develop the tools for big data systems.
4.	ELSEVIER 2015	Big Data and Hadoop-A Study in Security Perspective [9]	This paper shows the Big data issues and focused more on security issue arises in Hadoop Architecture base layer. Techniques: HADOOP	This paper shows the big data information and characteristics used in world wide. The issues are also mentioned to give idea about the big data issues in real time. The security issue is pointed more in order to increase the security in big data	Given approaches are used only in Hadoop Distributed File System which is the base layer in Hadoop. In Future these approaches are also implemented in other layers of Hadoop Technology.
5.	IEEE 2012	Big Data: Issues and Challenges Moving Forward [10]	This paper indicates Big data is a moving target. Put another way, it is that amount of data that is just beyond our immediate grasp, e.g., we have to work hard to store it, access it, manage it, and process it. The current growth rate in the amount of data collected is staggering.	Identified some of the major challenges – going forward – that believe must be addressed within the next decade and which will establish a framework for Big Data.	Future research will concentrate on developing a more complete understanding of the issues associated with big data, and those factors that may contribute to a need for a big data analysis and design methodology.
6.	IEEE 2014	Big-Data Security Management Issues [11]	This paper proposed the study will be a general framework defining security management on Big data. Techniques: NIST SP800-30.	The adapted framework that NIST Risk Assessment framework described in NIST SP800-30 will accelerate the progress of security issues to be able to cope up with the growth and dynamics of Big Data.	The future works will be implementing the adapted framework of security for Big Data.
7.	IEEE 2013	Attribute Relationship Evaluation Methodology for Big Data Security [12]	This paper considers big data as a single object which has its own attributes. Here assume that an attribute which have a higher relevance is more important than other attributes and includes the data mining and its techniques (classification, prediction, clustering, and association rule	It is applicable to big data with multiple attributes.	In future, we will complement the attribute generation and then try to evaluate the proposed method through experiments by actual big data and tools.

			discovery) and Big Data and its analysis. Also proposed the big data security using attributes. Techniques: Attribute Relationship Evaluation Methodology		
8.	IEEE 2014	A Survey of Cryptographic Approaches to Securing Big-Data Analytics in the Cloud [13]	The main aim of this paper is to describe the cryptographic techniques in the context of the cloud model and highlight the differences in performance cost associated with each. Techniques: Cryptographic techniques Homomorphic encryption Verifiable computation Multi-party computation	Present a computation model for big data analytics in the cloud and surveyed several cryptographic techniques that can be used to secure these analytics in a variety of settings.	In future the research direction is to design and develop secure multi-party computation techniques personalized specifically for a private semi-trusted cloud setting.
9.	IEEE 2014	Enhance Enterprise Android Application Security with Cloud Computing and Big Data Analytics [14]	This paper objective is to develop a framework that promotes the use of cloud computing and Big Data Analytics on building the assessment system. Techniques: Android Application Assessment Framework Big Data Analytic Model Experimental Assessment System	This research explores the potential value for utilizing the concepts of cloud computing and big data analytics in Android application assessment for enterprises.	As future work to allow the cloud base system to be extended to other systems of the enterprise, SOA is encouraged to be implemented to provide the analytic results as web services.
10.	IEEE 2013	Addressing Big Data Issues in Scientific Data Infrastructure [15]	The paper proposes the SDI generic architecture model that provides a basis for building interoperable data or project centric SDI using modern technologies and best practices.	The paper explains how the proposed models SDLM and SDI can be naturally implemented using modern cloud based infrastructure services provisioning model and suggests the major infrastructure components for Big Data.	The future research will require more formal approach and taxonomy of the general Big Data use cases both in science and industry.

Table 2: Literature Survey of Big data Security

IX. CONCLUSION

Big data needs extra requirements for security and privacy in data gathering, storing, analyzing, and transferring. In this paper, we examined studies on big data security and privacy, comparatively. Security and privacy are among the most important requirements in Big Data. We have seen that the best solution of implementing Big Data security and privacy is the law rather than the security technology but the laws can't keep pace with the development of technology and is different between countries. Thus, security technology and other methods are always necessary. Some possible methods and techniques to ensure security and privacy in Big Data have been discussed above. Moreover, we noted that correlation of massive data is the key which is the basis of use of Big Data as well as the reason of Big Data security and privacy issues. We recommend that the study of "no correlation" may be realization of security and privacy in Big Data. Security and privacy are two main and most important concepts to be considered whenever big data is used. The characteristic of data poses various challenges and chances in accomplishing security goals for example privacy, authentication, integrity, availability, and non-repudiation.

Big data privacy, safety and security are the biggest issues to be discussed more in the future, so new techniques, technologies and solutions need to be developed in terms of human-computer interactions or existing technologies should be improved for accurate results. It is hoped that this study would help understand the big data and its ecosystem better and develop better systems, tools, structures and solutions not only for today but also for the future.

REFERENCES

[1] Matturdi Bardi, Zhou Xianwei, Li Shuai, Lin Fuhong, "Big Data Security And Privacy: A Review", Big Data, Cloud & Mobile Computing, China Communications, Supplement No.2 , vol.11, issue: 14, Pp. 135-145, 2014.
[2] Duygu Sinanc Terzi,Ramazan Terzi,Seref Sagiroglu,"A Survey on Security and Privacy Issues in Big Data", IEEE International Conference, Pp. 202-207, 2015.
[3] Bernard Marr, "Big Data: The 5 V's Everyone Must Know", March 19, 2015. Available on internet Link: https://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters on September 16 2016.
[4] Risha Tabassum, Dr. Nidhi Tyagi, "Issues and Approaches for Big Data Security", International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS) ,Volume V, Issue VII, Pp. 72-74, 2016.
[5] Lohr Steve, "The age of big data [J]". New York Times, 2012, February 11.
[6] Seref Sagiroglu and Duygu Sinanc, "Big Data: A Review", IEEE, Pp. 42-47, 2013.
[7] Vibha Bhardwaj, Rahul Johari,"Big Data Analysis: Issues and Challenges",IEEE 2015.
[8] Alvaro A. Cárdenas ,Pratyusa K. Manadhata, Sreeranga P. Rajan,"Big Data Analytics for Security", IEEE ,Pp. 74-76, 2013.
[9] B. Saraladevi, N. Pazhanirajaa, P. Victer Paul, M.S. Saleem Basha, P. Dhavachelvan ,"Big Data and Hadoop-A Study in Security Perspective", Elsevier Procedia Computer Science 50 , Pp. 596 – 601, 2015.
[10] Stephen Kaisler, Frank Armour, J. Alberto Espinosa, William Money,"Big Data: Issues and Challenges Moving Forward", IEEE, Pp. 995-1004, 2013.
[11] Marisa Paryasto, Andry Alamsyah, Budi Rahardjo, Kuspriyanto,"Big-Data Security Management Issues", IEEE Internatinational Conference, Pp. 59-63,2014.
[12] Sung-Hwan Kim, Nam-Uk Kim, Tai-Myoung Chung, "Attribute Relationship Evaluation Methodology for Big Data Security", IEEE 2013.
[13] Sophia Yakoubov, Vijay Gadepally, Nabil Schear, Emily Shen, Arkady Yerukhimovich," A Survey of Cryptographic Approaches to Securing Big-Data Analytics in the Cloud", IEEE 2014.
[14] Hongye Zhong, Jitian Xiao, Xiangxin Zheng," Enhance Enterprise Android Application Security with Cloud Computing and Big Data Analytics", IEEE ,Pp. 239-243, 2014.
[15] Yuri Demchenko, Paola Grosso, Cees de Laat, Peter Membrey, "Addressing Big Data Issues in Scientific Data Infrastructure", IEEE, Pp. 48-55, 2013.
[16] Jaseena K.U. and Julie M. David," Issues, Challenges, and Solutions: Big Data Mining", NeTCoM, CSIT,Pp. 131–140, 2014.

