

## A Survey on Data Security and Challenges

J. Bhattacharya<sup>1</sup>, J. Mistri<sup>2\*</sup>, R. Biswas<sup>3</sup>, D. Dalui<sup>4</sup>, D. Singh<sup>5</sup>, P. Rakshit<sup>6</sup>, S. Bhattacharyya<sup>7</sup>

<sup>1,2,3,4,5,6,7</sup>Dept. of Computer Science & Engineering, JIS College of Engg., Kalyani, India

\*Corresponding Author: [jhumpamistri07@gmail.com](mailto:jhumpamistri07@gmail.com), Tel.: 8981633447

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

**Abstract**—Big Data has gained attention and become a hot topic in scientific and industrial fields. Big data refers to large amount of data which is difficult to store and process by using traditional data processing application. The digitization of our day to day activities generated petabyte and exabyte of data. The data coming from heterogeneous sources and maximum data is residing in unstructured way. So, Dealing with this enormous data many challenges occur and security issue is one of them. Some other big data challenges are – dealing with huge volume of data, integrating heterogeneous data source, validating data etc. In this paper, we review different challenges and security issues in big data environments.

**Keywords**—Big data, Challenges, Security, privacy, volume, velocity, veracity, variety, value, exabyte, petabyte

### I. INTRODUCTION

Big data term describe a large volume of data. The data may be structured, unstructured or semi structured. The volume of data is growing exponentially for various reasons. Our day to day activities generates huge amount of data. The digitization of our day to day life the whole world has gone online. The main sources of big data are cell phone, social media, medical records, e-commerce etc. Researchers have predicted that by 2020, the volume of data will be around 40 zettabytes. The processing, storing and dealing with such data occurs many challenges. Security is also a big issue for any organizations. Some big data stores can be main target for hackers. So, in this paper we review different challenges and security issues in big data environments.

There is available different definition of big data .Among them five characteristics that define big data called ‘5V’. They are –

**Volume:** Volume describe amount of data, which is growing exponentially day by day. This data generated by humans, social media interactions, sensor board etc. Researchers have predicted that by 2020, the volume of data will be around 40 zettabytes.

**Velocity:** Velocity refers to speed of generation of data. By a research, it have been seen that 1.03 billion active users of facebook on mobile. This is increase of 22% year over year. This shows that how fast the data is growing.

**Variety:** Big data coming from different sources and the type of data is different. The data may be structured, unstructured or semi structured. Hence, there is a variety of data which is

getting ingested every day. Unstructured data creates problem in storing, processing and analysing the data.

**Veracity:** Veracity defines the data in doubt or uncertainty of data. This is because of inconsistency and incompleteness.

**Value:** Value refers to profit that can be derived from the use of big data.



Figure 1. 5V's of big data

The main objective of this paper is to survey the literature related to various challenges of big data and security and privacy issues which a big data application chain is facing.

### II. BIG DATA CHALLENGES

#### A. Dealing with huge amount of data:

The most common challenge with big data is storing, processing and analysing all the data. Maximum data is unstructured that means it does not reside in a database. This type of data is photos, videos, PDF, and document, audio. This is difficult to search and analyse.

*B. Integrating heterogeneous data sources*

There is a problem in integration of big data. Because data comes from heterogeneous sources - social media, organizations, sensor board, e-commerce etc. Combining all that data and create a valuable insight from it is incredibly difficult.

*C. Unreliable data and troubles of upscaling*

The main fact is that big data is not 100% reliable. It contains wrong information, duplicates data and contains contradictions. And it's unlikely that data of extremely inferior quality can bring any useful insights or shiny opportunities to your precision demanding business tasks [29].

**III. SECURITY ISSUES**

Big data becomes more popular. The volume of big data is huge. So, the issue of security is becoming increasingly crucial. A security issue of big data is a big challenge. Many cyber-attacks have started to focus on big data. Big data increases the risk of data leakage because of its huge volume and velocity. There are four different sight of Big Data security:

- Infrastructure security
- Data privacy
- Data management
- Integrity and reactive security

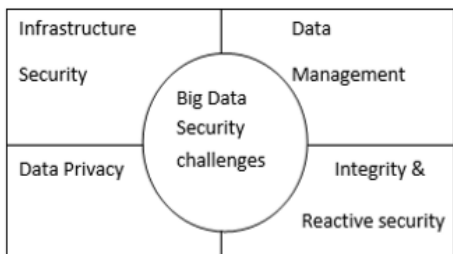


Figure 2. Challenges in Big Data Security

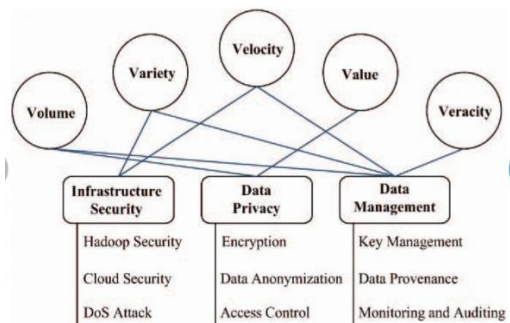


Figure 3. Caegory of Security Challenges in Big Data

*A. Infrastructure security:*

Infrastructure security refers technologies and the architecture of big data system. This is mainly based on

Hadoop technology. The security problems related to this technology have been widely discussed by researchers, who have proposed many methods how to improve the security of the Hadoop system. They have proposed many solutions by the use of different security mechanisms such as authenticity or cryptography.

*B. Data privacy*

Data privacy is main issue about which people are most concerned. Also this is a big concern for organizations while use of big data. A big data contains huge amount of personal data that organizations extract valuable insights from it. Many researchers proposed several techniques and mechanisms with which to protect the privacy of the data. Also allow companies to still make a profit from it.

*C. Data management*

In this part focuses on what to do once the data is contained in the Big Data environment. It shows how to secure the data and how to store the data in the Big Data system and how to share that data. Data management consists of:

- Security at Collection or Storage
- Policies, Laws
- Government
- Sharing Algorithms

*D. Integrity and Reactive Security*

One of the basis on which Big Data is supported is the capacity to receive streams of data from many different sources and with heterogeneous format: either structural data or non-structural data or semi structural.

Integrity and reactive security contains-

- Integrity
- Attack detection
- Recovery

**IV. LITERATURE SURVEY**

Mr. YannamApparao et al. [1] have discussed the security on secure data storage and transaction logs related with data storage management, processing through security. This paper also give the security on transaction logs are move the information from one origin to another origin Unofficially access may interfere and the user may not get data when the data are moving. This paper also demonstrates security matter on secure data and transaction.

Alice Joseph et al. [2] have discussed that big data is explained as large amount of dissimilar of data with unpredictable velocity. This paper is firstly describing definite characteristics of smart industry. Next it rewired huge profit and challenges of big data. At last it describes different securities and privacy threats of big data and refers solution.

Mohammed S. Al-kahtani et al. [3] have described that giving security and privacy in big data equally crucial along with giving quality service (QOS) in big data specially this paper provide network security approach.

Renu Bhandari et al. [4] have discussed the point of convergence on perception of big data. This paper also describes the reasons and solution features that should be taken into deliberation with go through development of secured big data solution and technologies used to hold risks and privacy concern.

Uthayasankar Sivarajah et al. [5] have discussed that big data analytics (BDA) is progressively becoming a move practice that are taken on by various organizations. Different BD challenges are discussed. This paper has recognized applicable BD research studies that have recognize and donate both conceptually and factually to the expansion and accrual of intellectual wealth to the BDA in technology and organizational resource management.

Snehalata Funde et al. [6] have described that today's data are collected from various sources. Enormous data is a traditional term utilized to portray the exponential increment what's more, openness of organized what's more; unstructured data with the huge amount of data security is increased. This paper reviewed various techniques preservation of the big data.

K.P.Maheswari et al. [7] have described that cloud computing is describe as configuration of different technologies such as networks, databases, operating system, virtualization, resource scheduling, and transaction management, loading balancing concurrency control and memory management. This paper is based on security levels, which are regard as issues accepted to cloud computing and big data and conclusion is based on the basis of those over security.

Haina Ye, Xinzhou Cheng et al. [8] have described that big data has become a proceeding factor in both scientific and industrial fields for it's possible value .A essential topic is security and privacy .In this paper the research and development on security and privacy in big data is accessed. Firstly it describes the effects of characteristics of big data on information security then it explained and analyzed the topics and issues on security.

Jamal Raiyn et al. [9] have discussed trade with information security and safety matters in public open spaces. Public open spaces incorporate high streets, street markets, shopping centers, community gardens, parks and playgrounds everyone plays a important role in social, cultural and economic life of a community. This paper describes technologies used in cyber parks to attain information security in big data era.

Ebrahim Saha fizadeh et al. [10] has described survey on security and privacy issued in big data and NOSQL. So, for the conventional security models have problems in trading with huge volume of data. This paper presents few security issues in big data and features the security and privacy challenges in big data infrastructures and NO SQL database. Lixiang Li et al. [11] have described that Big Data draws the attention not only because of its great power but for the severe security and privacy challenges it brings.

Sanchita Gupta et al. [12] have discussed that paper mainly deals with the security issues faced while using data mining technique from an expanded proportion and review different processes that can help to secure the information. We shortly discuss the key points of researched topics and then define states of approaches and also provide some essential ideas on future research.

Minit Arora et al. [13] have described that big data is the huge amount of data with variety of data with different nature .privacy is a big challenge for big data .This paper discussed the issues regarding privacy of this data and some current techniques to assure of the data.

Haval Mohammed Sidqi et al. [14] have described that big data in simple terms is extremely huge sets of data which can reveal trends, interests & patterns and classify data upon computational analysis. In addition, it would be the decision-makers and can get the value of this information is different and rapidly changing tasks, everything from data transferences daily social network clients business.

Alice Joseph et al. [15] have described that provides a literature review on the need of security and privacy issues of various big data applications. The first section gives a brief description of big data. The second section reviews the various big data applications and, hence, explains the importance of privacy and security of Big Data in the third section.

Renu Bhandari et al. [16] have focuses on key which insights of big data architecture somehow lead to top 5 big data security risks and the use of top 5 best practices that should be considered while designing big data solution which can thereby surmount with these risks. Big data architecture, being divisible in nature can undergo partition, replication and distribution among thousands of data and processing nodes for distributed computation thus supporting multiple features associated with big data analytics like real time, streaming and continuous data computation along with massive parallel and powerful programming framework.

Snehalata Funde et al. [17] have focuses on the Enormous data which is a traditional term utilized to portray the exponential increment what's more, openness of organized what's more, unstructured data. In future enormous data will

be essential to business as well as society like web office with later improvement in innovation, organizing furthermore, cost lessening in capacity gadgets, today we are overflowed with immense sum of data.

Sanchita Gupta et al. [18] have focuses on the basic idea here is to identify various types of users who face security issues regarding data mining applications. And for each of them individually debate on their confidentiality concern and the approach that can be adopted to save their information. We shortly discuss the key points of researched topics and then define states of approaches and also provide some essential ideas on future research.

Getaneh Berie Tarekegn et al. [19] have described that paper focuses on important concepts of Big Data. In this paper we discuss various things of big data. We define Big Data and discuss the parameters along which Big Data is defined. There includes the three V's of big data which are velocity, volume and variety. The authors also look at processes involved in data processing and review the security aspects of Big Data and propose a new system for Security of Big Data and finally present the future scope of Big Data.

Anju Abraham et al. [20] have discussed that tries to assist Cloud storage which provides data management and reduces the costs. Various clustering methods like K mean, K nearest neighbouring, DBSCAN clustering methods are implemented for clustering massive data that are related to each other there used map reduce framework in big data analytics. Clustering was done on encrypted partitioned data in order to protect the information from the third party access. Various approaches have been used for securing and maintaining the efficiency and performance of millions of dataset along with variety, velocity, and volume.

Yuri Demchenko et al. [21] have described that provides knowledge for practical implementation of such important security infrastructure components as federated access control and identity management, fine-grained data-centric access control policies, and the Dynamic Infrastructure Trust Bootstrap Protocol (DITBP) which allows deploying trusted remote virtualised data processing environment. The paper refers to the past and ongoing project experience by authors and discusses how this experience can be consolidated to address new Big Data security challenges that identified in this paper.

XiaolongJin et.al [22] have described that first briefly introduce the concept of big data, including its definition, features, and value. Then it was identified from different perspectives the significance and opportunities that big data brings to the world. Next, it was presented representative big data initiatives all over the world. Then described the grand challenges (namely, data complexity, computational

complexity, and system complexity), as well as possible solutions to address these challenges.

R.Kalaivani et al. [23] have described that concentrated on providing security to big data which was stored on cloud. User could not afford all infrastructures to support Big Data technology, Cloud Computing technology was boomed as user dream come true by performing massive-scale and complex computing. All the expensive data were present at various vendor clouds.

Trupti V. Pathrabe et al. [24] have discussed the continuous growth in the importance and volume of data has created a new problem: it cannot be handled by traditional analysis techniques. This problem was, therefore, solved through the creation of a new paradigm: Big Data. There also discussed some security issues as well as privacy issues in big data.

Prof. Asha Patel et al. [25] has described the privacy prevention for Association rule mining in big data. For providing privacy to the data is the major issue so that the third party is not able to access the sensitive information and a key value proposition of big data is access to data from multiple and diverse domains, security and privacy will play a very important role in big data research and technology.

Tilwani Mashook et al. [26] have discussed exhibits the analysis of Employee Tracking and Monitoring System data for fleet management. The new wave of digitizing Fleet Is recorded has seen a paradigm shift in the corporate industry. This paradigm shift from reactive to proactive employee data can result in an overall decrease in risk costs and mismanagement of tasks by employee which would eventually lead to Company's growth.

P.Joseph Charles et al. [27] have described the overview of Big Data Security. Big Data originated new issues related not only to the volume or the variety of the data, but also to data security and privacy. We explain the results obtained after applying a systematic mapping study to security in the Big Data ecosystem.

Julio Moreno et al. [28] have described the results which obtained after applying a systematic mapping study to security in the Big Data ecosystem. It is really impossible to carry out the detailed research into the whole topic of security, and the outcome of this research is, therefore, a big picture of the main problems related to security in a Big Data system, along with the principal solutions to them proposed by the research community.

## V. CONCLUSION AND FUTURE SCOPE

Big data has become one of the most promising and a hot topic in scientific and industrial fields to predict future trends. With enormous amount of data big data faces lots of

challenges and main is security issue. The main objective of this paper is to survey the literature related to various challenges of big data and security and privacy issues which a big data application chain is facing. In survey paper reviewed and explains the main problems and challenges related to security in Big Data, and how researchers are dealing with these problems. We found that the principal problems are related to the different characteristics of a Big Data system. Also, security issues are a vital problem when dealing with big data. Many authors focus their research on protect data. Also focus on how to overcome from the various challenges of big data.

## REFERENCES

- [1] Alice Joseph ,Prof. Mathew Cherian, “*The Quest For Privacy And Security In Various Bigdata Applications: A Study*” International Journal Of Current Engineering And Scientific Research , VOLUME-5, ISSUE-3, 2018, ISSN (PRINT): 2393-8374.
- [2] Renu Bhandari, Vaibhav Hans and NeeluJyothi Ahuja, “*Big Data Security – Challenges and Recommendations*” International Journal of Computer Sciences and Engineering, Volume-4, Issue-1, E-ISSN: 2347-2693.
- [3] SnehalataFunde, SurajKarale, LaxmanBharate, “*Survey Of Big Data Security*”,International Journal Of Current Engineering And Scientific Research, ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-5, ISSUE-2, 2018.
- [4] Sanchita Gupta, AkashKataria, ShubhamRathore, Dharmendra Singh Rajput, “*Information Security Issues In Big Data: Solution Using Ppdm (Privacy Preserving Data Mining)*”, International Journal of Pharmacy & Technology, Dec-2016, Vol. 8, Issue No.4, 25551-25568.
- [5] GetanehBerieTarekegn, YirgaYayehMunaye, “*Big Data: Security Issues, Challenges And Future Scope*”, International Journal of Computer Engineering & Technology (IJCET), Volume 7, Issue 4, July–Aug 2016, pp. 12–24, Article ID: IJCET\_07\_04\_002.
- [6] Anju Abraham, Shyma Kareem, “*Security and Clustering Of Big Data in Map Reduce Framework:A Survey*”,International Journal of Advance Research, Ideas and Innovations in Technology, Volume 4, Issue 1, ISSN: 2454-132X.
- [7] Yuri Demchenko1, Canh Ngo1, Cees de Laat1, Peter Membrey, Daniil Gordijenko, “*Big Security for Big Data: Addressing Security Challenges for the Big Data Infrastructure*”.
- [8] XiaolongJin, Benjamin W. Wah ,Xueqi Cheng , Yuanzhuo Wang , “*Significance and Challenges of Big Data Research*”, <https://www.researchgate.net/publication/272946686>.
- [9] R.Kalaivani, “*Security Perspectives on Deployment of Big Data using Cloud: A Survey*”, International Journal of Advanced Networking & Applications (IJANA), Volume: 08, Issue: 05 Pages: 5-9 (2017) Special Issue.
- [10] Trupti V. Pathrabe, “*Survey on Security Issues of Growing Technology: Big Data*”, IJRST, National Conference on Latest Trends in Networking and Cyber Security, March 2017.
- [11] Prof. Asha Patel, “*A Survey Paper on Security Issue with Big Data on Association Rule Mining*”, IJRST, National Conference on Latest Trends in Networking and Cyber Security, March 2017, IJRST/Conf/NCLTNCS/2017/025.
- [12] TilwaniMashook, Patel Malay, Pooja Mehta, “*Security and Privacy –A Big Concern in Big Data A Case Study on Tracking and Monitoring System*”, IJRST, National Conference on Latest Trends in Networking and Cyber Security, March 2017, IJRST/Conf/NCLTNCS/2017/007.
- [13] P.Joseph Charles, I.Carol, S.Mahalakshmi, “*Big Data Security an Overview*”, International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 02 |Feb-2018,
- [14] Julio Moreno, Manuel A. Serrano and Eduardo Fernández-Medina, “*Main Issues in Big Data Security*”, Future Internet 2016, 8, 44.
- [15] Mr. YannamApparao, Mrs. KadiyalaLaxminarayanamma Security Issue on “*Secure Data Storage and Transaction Logs in Big Data*” International Journal of Innovative Research in Computer Science & Technology (IJRCST) ISSN: 2347-5552, Volume-3, Issue-3, May 2015.
- [16] Alice Joseph, “*Big Data Security And Privacy In Smart Industry*” International Journal of Computer Engineering & Applications, Volume XII Issue III , March 18, ISSN 2321-3469.
- [17] Mohammed S. Al-kahtani “*Security and Privacy in Big Data*” International Journal of Computer Engineering and Information Technology (IJCEIT), Volume 9, Issue 2, February 2017.
- [18] RenuBhandari, Vaibhav Hans and NeeluJyothi Ahuja “*Big Data Security – Challenges and Recommendations*” International Journal of Computer Sciences and Engineering Vol.-4(1), PP (93-98) Jan 2016, E-ISSN: 2347-2693.
- [19] Uthayasankar Sivarajah, Muhammad MustafaKamal, ZahirIrani,Vishanth Weerakkody “*Critical analysis of Big Data challenges and analytical methods*” Journal of Business Research70(2017)263–286.
- [20] Snehalata Funde, Suraj Karale, Laxman Bharate “*Survey Of Big Data Security*” International Journal Of Current Engineering And Scientific Research, ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-5, ISSUE-2, 2018
- [21] K.P.Maheswar, P.Ramya , S.Nirmala Devi “*Study And Analyses Of Security Levels In Big Data And Cloud Computing*” International Conference On Recent Trends In Engineering Science, Humanity and Management, ISBN:978-93-86171-18-4
- [22] Haina Ye, “*A Survey of Security and Privacy in Big Data*” Department of network optimization and management.
- [23] Jamal Raiyn “*Introduction to Big Data Management Based on Agent Oriented Cyber Security*” Journal of telecommunications and information technology (2017).
- [24] EbrahimSahafizadeh, Mohammad Ali Nematbakhsh “*A Survey on Security Issues in Big Data and NoSQL*” ACSIJ Advances in Computer Science: an International Journal, Vol. 4, Issue 4, No.16 , July 2015 ISSN : 2322-5157
- [25] .Lixiang Li, Kaoru Ota, Zonghua Zhang and Yuhong Liu “*Security and Privacy Protection of Social Networks in Big Data Era*” Mathematical problems in engineering, Volume 2018, Article ID 6872587, 2Pages [2108].
- [26] Sanchita Gupta, AkashKataria, ShubhamRathore, Dharmendra Singh Rajput “*Information Security Issues In Big Data: Solution Using Ppdm (Privacy Preserving Data Mining)*” International Journal of Pharmacy & Technology ISSN:0975-766X
- [27] Minit Arora, DrHimanshuBahuguna “*Big Data Security – The Big Challenge*” International Journal of Scientific & Engineering Research, Volume 7, Issue 12, December-2016399ISSN 2229-5518.
- [28] Haval Mohammed Sidqi, “*Big data analytics:survey paper*” International Journal Of Latest Trends In Engineering And Technology, VOLUME 8 ISSUE 4-JULY 2017 e-ISSN:2278-621X
- [29] <https://www.scncsoft.com/blog/big-data-challenges-and-their-solutions>.

### Authors Profile

Jayshree Bhattacharya has received her B.tech degree in Computer Science and Engineering from West Bengal University of Technology in 2018. She is currently pursuing her M.tech degree in Computer Science and Engineering from Maulana Abul Kalam Azad University of Technology.



Mrs Jhumpa Mistri is presently working as an Assistant Professor in the Department of Computer Science and Engineering of JIS College of Engineering, India. She received the Master of Technology (M.Tech.) degree from JIS College of Engineering, India in 2017. She also received the degree of Bachelor of Technology (B.Tech.) in Computer Science and Engineering from Abacus Institute of Engg. & Management.



Dharmpal Singh received his Bachelor of Computer Science and Engineering and Master of Computer Science and Engineering from West Bengal University of Technology. He has done his Ph.D in year 2015. He has about 12 years of experience in teaching and research. At present, he is with JIS College of Engineering, Kalyani, and West Bengal, India as an Associate Professor and Head of the department. He has published 38 papers in referred journal and conferences index by Scopus, DBLP and Google Scholar and editorial team and senior member of many reputed journal index by SCI, Scopus, DBLP and Google Scholar. He has organized seven national levels Seminar/Workshop, published three patents and has applied for the AICTE Project in year of 2019.



*Dr. Pranati Rakshit* is an Assistant Professor in the Department of Computer Science and Engineering of JIS College of Engineering, India. She has more than 19 years of work experience associated with teaching and research. She has completed her Ph.D. degree from Jadavpur University, Kolkata, West Bengal, in the field of Pattern Recognition and Medical image analysis. She has completed her Master degree and B.E. Degree in Computer Science & Engineering. She has worked in the field of Data Mining, image Processing, IOT also. She has supervised more than 20 M.Tech Projects. She has a good number of research publications. She is a life member of Indian Society of Technical Education.



Ms. Sonali Bhattacharyya is an Assistant Professor from Department of Computer Science and Engineering of JIS College of Engineering, West Bengal. She is attached with teaching more than 14 yrs. She completed her M. Tech from University of Calcutta, Raja bazar Science College. Her interest of work is the field of Data Mining, Machine Learning etc. She is a life member of CSI and FOSET.



Debolina Dalui is a 2<sup>nd</sup> year student, pursuing Master of Technology in the department of Computer Science and Engineering, from a renowned engineering college of West Bengal.

