iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 10 **Issue:** VI **Month of publication:** June 2022

**DOI:** https://doi.org/10.22214/ijraset.2022.44544

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 10 Issue VI June 2022- Available at www.ijraset.com*

# Challenge and Security Issues UsingBig Data: A Survey

Bhavesh Thale[1], Dr. Suhasini Vijayakumar[2]
[1]*Student MCA Dept., Bharati Vidyapeeth's Institute of Management and Information Technology, Navi Mumbai*
[2]*Principal and Prof. MCA Dept., Bharati Vidyapeeth's Institute of Management and Information Technology, Navi Mumbai*

***Abstract:*** ***Data is always one of the most valuable assets. Data volume in the world is expanding day by day because of internet access, smartphones, social networking sites, ubiquitous computing, and many other technological advances. The problem is that they do not understand how to structure and convertthis large and complex set of data big data, these days popular term that refers to a huge collection of very large and complex sets, is dealing with severe security and privacy challenges. An important characteristic of big data is that data from various sources have life cycles from collection to destruction, and new information can be derived through analysis, combination, and utilization. Data analysis is now used in almost every aspect of our society: communication, marketing, banking, and research. There are huge opportunities in the big data phase for science, health care, economic decision-making, education, and novel forms of public interaction and entertainment. But these opportunities also present challenges related to security and privacy. In this paper we are focusing on the big data and its related security issues.***

***Keywords: Big Data, Security, Utilize, Privacy issues***

## I. INTRODUCTION

The Big Data field applies to manage data sets whose size istoo large for commonly used software tools to capture, man- age, and analyse that amount of data effectively. Data volume is expected to double every two years. Data from all these sources are very often unstructured, and come from a wide range of sources, including social media, sensors, scientific applications, surveillance, video and image archives, Internet search indexing, medical records, business transactions, and system logs. Currently, big data is gaining more and more traction as devices connected to the Internet of Things (IoT) are increasing rapidly, producing large amounts of data that must be transformed into valuable information. [1]

## II. CHARACTERISTICS OF BIG DATA

1) *Volume:* Big Data contains an undecided and unfilteredvolume of information. The data collected is different fordifferent businesses. Therefore, the efforts paid are unique.Nonetheless, filtering valuable data from the voluminous pileis essential. Companies need to process this high-volumeinformation to address their business challenges. [2] **Velocity:** It is the speed at which data is created and collected.

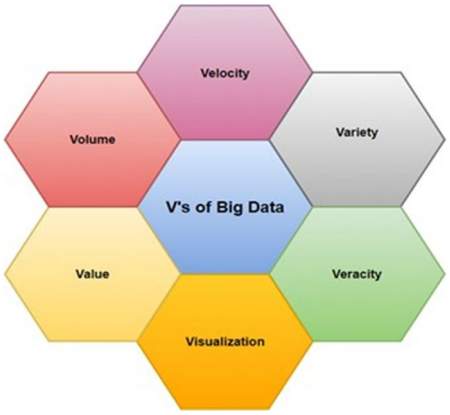



Fig. 1. Characteristics of Big Data

Mobile, SaaS solutions, e-commerce transactions, and IoT devices are a few of the primary sources of acquiring real-time data. The velocity at which data is generated at scale requires real-time handling and processing for augmentingData Analytics. [2]

©IJRASET: All Rights are Reserved 

2) *Variety:* Conventional data types consist of structured data that fit well with relational databases. However, with Semi-structured and Unstructured data in the landscape, the infor- mation received requires additional pre-processing to convert it into digestible formats. While Structured data can be quickly dealt with, Semi-structured and Unstructured data need to be converted into predetermined models or formats before turning them into actionable information. [2]
3) *Veracity:* The veracity of your data can be defined as its accuracy. Big Data veracity is one of the most important characteristics since low veracity can greatly affect the quality of your results.
4) *Visualization:* Big data visualization refers to presenting your insights via visual representations such as charts and graphs. It has become popular in recent years as big data professionals often share their insights with non-technical audiences.
5) *Value:* Velocity refers to the speed of data processing. High velocity is crucial for the performance of any big data process. It consists of the rate of change, activity bursts, and the linking of incoming data sets. [3]

## III. UNDERSTANDING THE CHALLENGES OF BIG DATA SECURITY




Fig. 2. Big Data Concern

## IV. LITERATURE SURVEY

Vishal Joshi 2020 [4]present some important concept of big data-specific security and privacy challenges so that it will bring renewed focus on fortifying big data infrastructures.Security/Privacy Issues and Challenges in Big Data paper provide explanation of how various users of big data faces issues in dealing with day-to-day operations in big data at various stages of big data ecosystem. This paper explains the research took place in order to address the main problemsand challenges related to security in Bigdata and thrown light on points of consideration while working with big data.

R. Sumithra 2018 [5] presents a comprehensive surveyon Security, Privacy Issues and Challenges in Big Data and Cloud Security. Big data is giving good decision making, it provides benefits of a data warehouse and an added functionality of analyzing data from distributed file systems. Big data is cost effective. This paper also resolved legal issues around intellectual property rights, data privacy andintegrity, cyber security and big data code of conduct.

Jose Mura 2014 [1]Security and Privacy Issues of BigData. This paper discusses current security and privacy issues.There are many sources of unstructured data including social media, sensors, scientific applications, surveillance, video andimage archives, Internet search indexing, medical records, business transactions, and system logs.

P. Kamakshi Dec 2014 [6]Survey on big data and related privacy issues. In this paper the strength and applications of big data as well as various privacy issues are discussed

Lo'ai A. Tawalbeh and Gokay Saldamli In this [7]is paper discussed existing layered cloud architectures andpresent a solution addressing the big data storage, use of P2P Cloud System (P2PCS) for big data and hybrid mobile cloud computing model based on cloudlets concept and apply this model to health care systems as a case study processing and analytics.

Minit Arora and Dr Himanshu Bahuguna (2016) [8] presents a survey that organizations used various methods of de-identification to ensure security and privacy. The most common method of ensuring security and privacy is through verbal and written pledges. However, history has shown that this method is flawed. Passwords, controlled access, and two factor authentication is low-level, but routinely used, technical solution to enforce security and privacy when sharingand aggregating data across dynamic, distributed data systems.

K.P.Maheswari, P.Ramya and S.Nirmala Devi(2017) [9] provides a study and analyses of security levels in big dataand cloud computing. The big data issues are most acutely felt in certain industries and in certain government activities. The security issues of big data systems and technologies are also applicable to cloud computing because it is veryimportant for the network which interconnects the systems.

J.L. Joneston Dhas, S. Maria Celestin Vigila and C.Ezhil Star(2017)[10] designed a framework on Security and Privacy-Preserving for Storage of Health Information Using Big Data. The storage of health records as big data presents many real-time problems. Among them is how to protect the data in the cloud. The next one is how to identify the record and how to protect the health information from the unauthorized user.

## V. BIG DATA: SECURITY ISSUES

### *A. Fake Data*

Creating fake data poses a serious threat to businesses since it consumes time that could otherwise be used for identi- fying or resolving other problems. In large-scale analytics, companies are more likely to be able to leverage inaccurate information, since evaluating individual data points is difficult. As a result of false flags for fake Data, unnecessarily costly actions may also be undertaken that negatively impact pro- duction or other critical functions required to run a business. One way to avoid this is to ensure that companies shouldbe critical of the data they are working on for enhancing business processes. The ideal approach would be to identify anomalies by validating the data sources by periodic evaluation and evaluating Machine Learning models with diverse test datasets.

### *B. Data Privacy*

Data Privacy is a big challenge in this digital world. Personal and sensitive information is protected from cyberattacks, breaches, and improper or unintentional dataloss. Data Privacy needs to be strengthened by businessesfollowing stricter privacy compliance standards with the help of cloud-based access management tools. It is best to follow a few rules Implementing one or more Data Security technologies should be accompanied by following a few rule There are three general rules for securing your data: know what data you have, control your data stores and backups, secure your network from unauthorized access, and conduct regular risk assessments. [11] The two primary techniques for data privacy are (1) anonymization and (2) pseudonymization.

Anonymization includes randomization and generalization Randomization strategies modify the integrity of the statistics to keep away from the energetic hyperlink between the statistics and the individual. [12] On the different hand, the generalization method generalizes or dilute the attributes of facts topics by using altering the respective scale or order of magnitude. Generalization strategies consist of aggregation, K-anonymity, and L-diversity/T-closeness. Aggregation and K-anonymity protect in opposition to singling out by using grouping them with, at least, K different individuals. On the other side, L-diversity is the extension of K-anonymity such that, in every equivalence class, each and every attribute hasat least L unique values to keep away from inference attacks. And, T-closeness is the expanded L-diversity such that equal training similar to the preliminary distribution of attributes in the table are created to maintain the facts as close to the unique one. Despite a number of methods in anonymization,it is proven now not enough for the privacy guarantee in a latest work. [12]
Pseudonymisation replaces one attribute in the dataset through every other to limit the likability between the unique identification of a records problem and the dataset.

The methods for pseudonymization include [12]

*1)* Encryption with a secret key,
*2)* Hash feature (a characteristic that returns a fixed-size output from a enter of any size and can't be reversed),
*3)* Keyed-hash feature with saved key
*4)* Deterministic encryption and
*5)* Tokenization and covering

### *C. Data Access Controls*

It is critically important for an organization to have a system which is fully secure. Access to data should only be granted to authenticated users. A system's access control needs to be designed in such a way that attackers, hackers, or malicious users will not be able to circumvent it. The problem is that securing a fully reliable and strong access control system involves a major investment and ongoing maintenance. [11]

## VI. BIG DATA APPLICATION IN DIFFERENT DOMAINS

### *A. Big Data Application in Education*

*1) Enhance the Grading System:* Big Data allows educational institutions to monitor students' performance across various disciplines on an individual and collective basis. They can develop solu- tions to assist students in reaching their goals. Analysis of student grades in different subjects helps institutions create study plans that are tailored to the learning abilities of their students. Through data analytics, they can discover the factors affecting student performance and provide effective solutions. [11]

*2) Career Guidance to Students:* Students can use big data to determine which careerpath is right for them. According to the student's area ofinterest and performance in respective subjects, teacherscan advise him or her about the right career path to choose. Career options are expanded for students byusing big data. Schools can offer career guidance by using big data.

*3) Proposing New Learning Plans:* Students have different learning capabilities. Some can learn just by reading, while others can learn by writing. Students may also learn by watching videos or through other methods. Many times, students are forced to adapt to a defined learning structure, which can hinder their growth. This negatively affects students' academic performance. Learning plans can be built based on the capabilities of students using Big Data Analytics. Data analytics can help students identify their strengths and weaknesses and provide them with relevant study plans and learning material.

### *B. Big Data Uses in Health Sector*

*1) Predictive* Analytics in Healthcare: In healthcare, Big Data plays a crucial role in Pre- dictive Analysis. With the help of predictive analysis, doctors can provide excellent treatment to their patients. It ensures the patient's safety. By identifying which patients are at risk for which diseases, doctors can makeinformed decisions that will improve patients' health.

*2) Electronic Health Records (EHRs):* In the medical sector, Big Data is primarily used tocreate electronic health records. Healthcare sectors werefacing challenges coping with the rapidly growing num-ber of patient records earlier. Currently, each patient hastheir own medical records, such as a list of medicines, medical reports, laboratory test results, etc. Healthcare sectors can maintain the data about patients and have access to it easily with the use of electronic health records. There is a separate file for each patient. It can beeasily modified by the doctor and can be shared safely.

3) *Real-Time Monitoring:* Patients are now receiving excellent treatment from healthcare systems that monitor their health in real time. There are many tools available that analyze patient data and suggest actions to be taken by doctors. There are many wearable sensors that track patients' health like blood pressure, pulse rate, heartbeat, etc. which can be monitored by the doctors. It reduces patients' unnecessary visits to hospitals

*C. Big Data Application in Banking and Finance Sector*

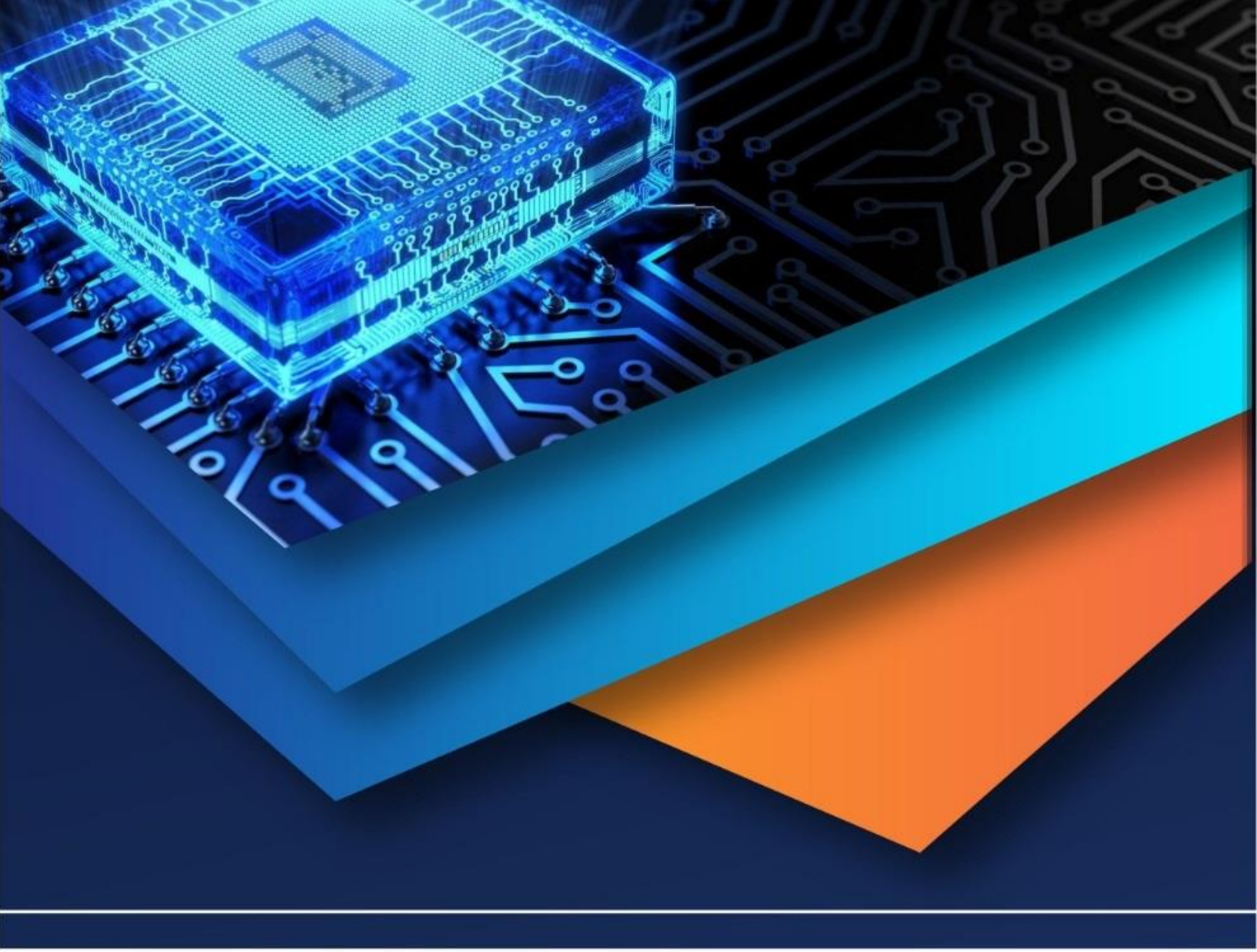
1) *Real-time Stock Market Insights:* Machine getting to know with Big Data permits finance sectors to analyze inventory fees via thinking about social and political tendencies that would possibly have an effect on the inventory market. It permits monitoring of traits in real-time for this reason presenting the functionality to analysts to consider fabulous statistics and make clever decisions.
2) *Fraud Detection and Prevention:* ML fuelled [11] [13] [14]by big data is significantly responsible for fraud detection and prevention. The security risk possessed by the credit card gets mitigated with analytics. When information about any secure and valuable credit card gets stolen, banks instantly freeze the card and transactions. It then informs the customer about the security threats.
3) *Trade Analytics:* Banks, Hedge funds, and others in the economic marketsare the use of Big Data for change analytics. Trade analytics helps finance and banking sectors in pre-trade selection support, high-frequency trading, Predictive An- alytics, and sentiment measurement.

## VII. CONCLUSION

In this paper, we discuss the fundamentals of Big Data.We also explain how various organizations deal with big data issues. There is also some research presented in thispaper regarding these challenges, but it does not provide a solution. There is some information and technologies that mayadd to the most relevant and challenging Big Data securityand privacy issues. Security and privacy concerns are largely related to the fact that huge quantities of personal information are freely available in digital form. By using the personalinformation of customers, many organizations are utilizing BigData for their own benefit, profit and to accomplish their goals. As part of the Big Data code of conduct, we also need to resolve legal issues surrounding intellectual property rights, data integrity, and cyber security. In this paper we also discusschallenges faced by various professional workers in different domains. There have been many surveys produced over the last few years based on the fact that Big Data technology is gaining traction across various industries. [4]

## REFERENCES

[1] J. Moura, "Security and Privacy Issues of Big Data," Handbook ofresearch on trends and future directions in big data and web intelligence., no. 20-52, 2015.
[2] https://hevodata.com/learn/big-data-security/.
[3] "https://www.analytixlabs.co.in/blog/characteristics-of-big-data/".
[4] M. V. Joshi, "Security/Privacy Issues and Challenges in Big Data,"International Research Journal of Engineering and Technology (IRJET),vol. 07, no. 06, 2020.
[5] R. Sumithra, "Security, Privacy Issues and Challenges in Big Data and Cloud," Special Issue based on Proceedings of 4th International Conference on Cyber Security (ICCS), 2018.
[6] P.Kamakshi, "SURVEY ON BIG DATA AND RELATED PRIVACY ISSUES," International Journal of Research in Engineering and Tech- nology, vol. 03, no. 12, Dec 2014.
[7] L. A. T. a. G. Saldamli, "Reconsidering big data security and privacyin cloud and mobile cloud systems," Journal of King Saud University – Computer and Information Science, 2019.
[8] M. A. a. D. H. Bahuguna, "Big Data Security – The Big Challenge," International Journal of Scientific Engineering Research, vol. 7, no. 12, Dec 2016.
[9] P. a. S. D. K.P.Maheswari, "STUDY AND ANALYSES OF SECURITY LEVELS IN BIG DATA AND CLOUD COMPUTING," International Journal of Innvative Research in Science and Engineering, vol. 3, no.02, 2017.
[10] S. M. C. V. a. C. E. S. J.L. Joneston Dhas, "A Framework on Security and PrivacyPreserving for Storage of Health Information Using Big Data," International Science Press, 2017.
[11] "https://techvidvan.com/tutorials/big-data-applications/".
[12] EU. Opinion 05/2014 on anonymisation techniques. ARTICLE 29 DATAPROTECTION WORKING PARTY, 2014.
[13] P. V. J. S. Batcha, "THE FIELD OF BIG DATA FOR SECURITY INTELLIGENCE," IJCRT, vol. 6, no. 2018, 2 April 2018.
[14] M. Parihar, "Big Data Security and Privacy," International Journal of Engineering Research Technology, 07 July 2021.
[15] R. V. Sitalakshmi Venkatraman, "Big data security challenges and strategies," vol. 4, no. 3.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)