# Big Data and Cyber Security: Challenges and Solutions

DR. RISHI KUMAR SRIVASTVA
*Assistant Professor & Trainer Placement, Department of Arts and Commerce, Amity University, Jharkhand, Ranchi, India*

***Abstract-*** ***The Big Data is boosting up in every field of research, and it has almost no untouched area. Thus, the Big Data has taken further strides in data-intensive computing field to boost up the performance of the organizations. The Big Data Analytics (BDA) is logical analysis of very large scale dataset to gain insight and facts of the analyzed data sets. In addition, the BDA is used in the Big Data Security field for several reasons. Moreover, the Big Data analytics is merged with Big Data Security which results in another research direction, called Big Data Security Analytics (BDSA). The BDSA poses several research issues and challenges. In this paper, we present a study report on Big Data Security Analytics.***

***Indexed Terms-*** ***Big Data, Big Data Security, challenges, Big Data Security***

## I. INTRODUCTION

The dataset known as "Big Data" is enormous, and what's more, its size is continually expanding as more and more information is added to it. The unchecked expansion of data may become a problem for certain companies and organisations. Besides, firms are acquiring more data in spite of the fast expansion of their data warehouse. The corporation views these data as assets, and they are used to drive revenue for the business. However, Big Data is concerned with giving value to data that has been discarded as being worthless via the use of Big Data analytics (BDA) [1]. Big Data helps to generate more money by extracting value from the data that is collected and analysing it. As a result, big data analytics is quickly becoming a force to be reckoned with in the research arena across a wide range of disciplines [2]. Logical analysis is performed on a very high volume of datasets as part of the BDA process. The BDA is used in a wide variety of domains. The BDA's scope extends beyond the confines of the study of computing itself and includes other fields as well. Because of this, the BDA in inter-disciplinary research is an excellent opportunity for academics, professionals in industry, and other practitioners. Moreover, the Big Data analytics is gaining acceptance unbelievably large region. Therefore, Big Data Security Analytics also arises. Many prospects for study have opened up to us as a result of the convergence of big data analytics and big data security.

Big data and cyber security are two areas that have taken on a growing amount of significance over the course of the last few years. The term "big data" refers to the enormous quantity of data that is produced on a daily basis, while "cyber security" refers to the practise of guarding this data against unauthorised access, theft, and other forms of criminal activity. The convergence of these two industries gives rise to a number of difficulties, but there are also a variety of approaches that may be taken to overcome these difficulties.

The Big Data Security (BDS) [3], [4], [5], [6] and Big Data Analytics (BDA) [7], are merged to form Big Data Security Analytics (BDSA). As a direct result of this, ensuring the safety of a very large-scale system is now much simpler. Moreover, the data are gathered in a daily basis in IT organization to dump the data in the data warehouse. After some time, the data were useful. The data warehouse stores very huge collections of data, each of which includes a variety of security-related details that are quite helpful to the security system. For example, the activities of users are recorded in the log files. As a result of this, the Big Data Security Analytics presents the research community with a once-in-a-lifetime opportunity to mine the breach from the extensive data set.

The Big Data Security Analytics (BDSA) initiative calls for the capacity to analyse and manage

enormous datasets. The BDSA places an emphasis on the following fundamental aspects:-

Anomaly detection- The anomaly detection includes fraud detection, vulnerability and intrusion detection [8].

- Misuse detection- For instance, DDOS.
- Real-time monitoring- For Instance, security alarms
- Prediction- The prediction includes possibility analysis, and prediction of an event.
- Prevention- The security prevention is a prior security preventive measure to be taken care of. However, the prevention takes place after security attacks.

The huge amount and diversity of data, together with their exponential rate of expansion, are the primary challenges posed by big data [9]. Certain organisations gather an enormous quantity of information that is pertinent to security [10]. The organisations go through the discarded data in search of incidents connected to data security. Alongside the proliferation of Big Data for a variety of applications comes a rise in data-driven information security. Forensics, intrusion detection, fraud detection, and anomaly detection are some examples of detecting technologies. Any system that may connect to the Internet is susceptible to attack [11], and it is absolutely necessary to take precautions in order to protect the system. Mondek et al. place a strong emphasis on the importance of defence security systems. The defence security system must be operated with the greatest caution in order to safeguard their expansive system. Additionally, via the use of Big Data Security Analytics, the defence security system investigates the potential weaknesses of any system. In this article, we will discuss the most recent developments in Big Data Security Analytics[12].

Challenges:

- Big data contains vast amounts of information, which makes it difficult to handle, analyse, and keep safe. This presents a number of challenges.
- The fact that data may originate from a broad variety of sources and in a variety of forms makes it difficult to guarantee that it is accurate and that it has not been tampered with.
- The speed at which data is created and processed makes it difficult to keep up with possible dangers in real time. This is due to the fact that the speed at which data is generated and processed[13].
- The complexity of the systems: The dangers posed to computer networks are growing more advanced, necessitating the use of sophisticated computer systems and algorithms to identify and stop them[14].
- The following are possible solutions:
- Make use of sophisticated analytics: The analysis of big data may assist uncover trends and spot abnormalities, both of which might be signs of a potential breach in cyber security. The use of machine learning algorithms may also assist in identifying possible risks and automatically responding to them.
- Strengthen the protection of personal data It is crucial to data's confidentiality and integrity that it be managed and stored in a secure manner. The protection of data may be assisted by the use of encryption, access limits, and monitoring.
- Make use of security solutions that are hosted in the cloud. Cloud-based solutions may provide scalability and flexibility, making it simpler to manage and safeguard massive amounts of data[15].
- Raise the level of awareness as well as the level of training: Cyber security risks are not confined to the IT department alone; rather, each employee has to be made aware of the possible dangers and taught to recognise and avoid them.
- Adopt a risk management strategy. Organisations should carry out frequent risk assessments to detect possible threats and vulnerabilities, and then prioritise security measures in accordance with the results of those assessments[16].
- In conclusion, the convergence of big data and cyber security presents a number of obstacles; nevertheless, there are solutions available to solve these problems. Some of the methods in which organisations may maintain the security and integrity of their big data include the implementation of advanced analytics, the

enhancement of data privacy, the utilisation of cloud-based security solutions, the increase of awareness and training, and the adoption of a risk management strategy[17].

## II. OBJECTIVE

1. The Study The Big Data Is Boosting Up In Every Field Of Research.
2. The Study The BDSA Poses Several Research Issues And Challenges.

## III. BIG DATA

The term "Big Data" refers to a large amount of data that includes many different datasets and grows at an exponential rate. The traditional database system is unable to manage very large datasets; hence, the Big Data paradigm has come into being. The enormous dataset presents a number of challenges when it comes to storing, processing, and managing it [18]. Big Data is comprised of several different data kinds, which are pieced together to create the massive amount of data [19]. As a direct result of this, the data warehouses are more susceptible to having their security compromised. [20] Structured data, semi-structured data, and unstructured data are the three primary categories of data. [21] Unstructured data may be in any form. It's interesting to note that 90% of Big Data are sorts of unstructured data. The petabyte is no longer a unit of measurement used by Big Data[22].

## IV. BIG DATA SECURITY

Concerns about users' right to privacy and data protection pose the greatest threat posed by big data. The problem is that analytics engineers are changing the personal data or the unauthorised data in order to provide erroneous findings[23]. This is a hurdle. Due to the fact that Big Data is a compilation of a huge number of interconnected data sets, it requires its own security to prevent unauthorised access.

## V. BIG DATA SECURITY ANALYTICS

The purpose of Big Data security analytics (BDSA) is to provide a complete and current picture of IT operations; as a result, choices based on data and made in a timely manner are produced by security analytics . The big data analytics both contribute to the reshaping of the security intelligence and provide possibilities for the agencies that deal with security and intelligence. The surroundings of information security are comprised of massive dispersed systems that exist in both physical and virtual realms. Information security is of the highest significance in business settings since it has a direct impact on the availability and continuity of services, and as a result, customer satisfaction. In order for the security teams to be effective, they need to stay current with the most recent developments in the technology, practises, and controls that are used in their industry. Any gadget that can connect to the internet is at risk of being attacked by malicious actors. As a result, storing data associated with security in preparation for pre-analytical, analytical, and post-analytical assessments (real-time or post hoc forensic analysis) is an absolute need. Big Data Analytics is used to manage the enormous amount of data.The statistics are transformed into revenue. As a result, data are quickly becoming the new oil in the current information technology sectors. As a result, the Cloud Security Alliance places a strong emphasis on the following security data: [24]-

Acquiring a vast quantity of data from a variety of sources, including external sources like vulnerability databases

- Conducting further in-depth analyses on the available data.
- Offering a unified perspective of the information that pertains to security.
- The ability to do real-time analysis on streaming data

The data are produced in a variety of text encodings at varying speeds, also dependent on circumstances external to the system such as day and night[25]. The analytics around security concerns in the infrastructure of modern businesses are a classic example of a big-data challenge that calls for big-data solutions. The users essentially have two alternatives available to them in order to get insights and handle the vast amounts of data that have been recorded:

1. Admit that the rules of the game are fair and begin developing solutions that are very complicated.
2. Make an effort to alter the guidelines and reduce the complexity of the issue as much as you can.

The findings of the first study lead to a considerable reduction in the complexity of major issues, which, as a consequence, leads to an improvement in the effectiveness, reliability, and capacities of security defence solutions[26]. Database management systems that are scalable and highly accessible, such as MongoDB, ElasticSearch, and Apache Hadoop, among others, are able to store massive volumes of heterogeneous data in a distributed way, allowing for complicated off-line analysis to be performed using dispersed data[27]. In addition to this, they allow for the incorporation of more compact datasets that may be analysed online in real time[28]. To put it simply,

1. The activities of online analysis should pick data that is relevant and important.
2. The collecting of all relevant data that is accessible, particularly from dispersed storage systems, is a prerequisite for forensic examination.
3. The off-line analysis is carried out in the background, in a manner that is fully automated.

There is an urgent demand for security solutions that can deliver continually growing analytic output in the form of AI that is reliable and intelligible for upper corporate management as well as technical professionals in order to assist rapid decision making[29]. The following is how Chen et al. outlines the goal of the BDSA: [30]-

- Mining and grouping of criminal association rules
- Analysis of criminal networks
- Analysis of spatial-temporal relationships
- Visualisation of spatial-temporal relationships
- Multilingual text analytics for security
- An investigation of affect and sentiment for the sake of security
- Investigation and Assignment of Responsibility for Cyber attacks

## VI. SECURITY DIAGNOSTIC ANALYTICS

The application of security diagnostic analytics may either lead to the discovery of something or the determination of why something occurred[31]. By promoting a product or anything else via social media, for instance, a company may evaluate the number of posts, mentions, followers, fans, page views, and reviews, among other metrics, in order to gather information about how they compare to their rivals[32]. The data obtained from the online platforms are then compiled into a single perspective, allowing for an analysis of both the successful and unsuccessful strategies [33]. The phrase "security diagnostic analytics" most often refers to the use of data discovery tools and visualisation, in which data is used for the purpose of determining patterns[34]. It is useful in determining the elements that have a direct or indirect impact on the bottom line[35].

## VII. PREDICTIVE SECURITY ANALYTICS

The use of predictive security analytics enables the prediction and prevention of potential attacks based on the contextual analysis of data gathered from a variety of security devices and systems included inside Big Data[36]. The data are then turned into vital information and insights that can be put into practise by using both this gathering and the use of predictive analytics[37]. The processing of predictive analytics does, however, call for high-end computational resources. Computing that is data-intensive demands a significant amount of processing power and storage space[38]. The integration of an appropriate approach, such as machine learning processes, with predictive analytics is what it takes to get the desired outcomes[39].

## VIII. SECURITY DECISION ANALYTICS

The approach that handles significant choices in a formal way is known as security decision analytics[40]. This methodology assists in locating, representing, and evaluating significant parts of decisions, as well as providing appropriate recommendations for those who are in charge of making such decisions[41].The end objective is to get actionable insights from the analysed data in order to successfully implement the security solution[42]. The

security decision analytic framework provides the decision maker with assistance in methodically thinking about the goals and preferences, as well as the structure and uncertainty in the issue, and modelling numerically these, together with other significant components of the problem and their interrelationships[43].

## IX. USER BEHAVIORAL SECURITY ANALYTICS

User Behavioural Security Analytics refers to the process of monitoring, collecting, and analysing information on user actions[44]. These patterns and insights are uncovered so that evidence of an intruder, insider threat, and unsafe behaviour on the network may be identified. In addition, since it focuses on behaviour, the UBSA is able to detect assaults that circumvent threat intelligence and raise an alarm on harmful behaviour at an early stage in the attack, which enables security measures to swiftly react. The UBSA places a strong emphasis on the system's and network's perimeter[45]. The system does an evaluation of how the users' actions will affect the system. When it comes to resources that are less sensitive, the effect is significantly reduced.This system takes into account the behaviour of the entity in addition to the behaviour of the user. Because of this, the system now has access to a new section known as User Behaviour and Entity security Analytics (UBESA). The UBESA provides assistance to the organisation in addressing anomalies that may be posed by potential risk[46].

## X. BIG DATA SECURITY ANALYTICS AS A SERVICE

In the information technology industry, the security problem has become increasingly urgent to address in tandem with the expansion of network infrastructure and utilisation. And because to the benefits offered by BDSA, the big data security analytics as a service may now consider expanding into other areas of research and development[47]t. As the old adage goes, "with great power comes great responsibility," and businesses who provide BDSA as a service need to be aware of the potential dangers that might occur during implementation. The Big Data Security Architecture as a Service (BDSA as a Service) is an essential component of Big Data security over the paradigm of cloud computing. The pay-as-you-go concept is made possible by the cloud paradigm of BDSA[48]. As a consequence, the price of BDSA drops by a significant amount[49]. As well as the company's end users, the service provider reaped significant advantages from the provision of BDSA as a Service. The provision of BDSA as a Service paves the way not only for the creation of potentially very large income from security analytics but also for its provision. Therefore, the BDSA as a Service will be the next paradigm-shifting innovation [50].

## CONCLUSION

Big Data encompasses a broad area of the study field; yet, there are many sectors that have not yet begun to use Big Data to improve the functioning of a system. When data are gathered from a wide variety of sources, their reliability for research purposes is compromised. For this reason, the lion's share of the money generated by the IT industry comes from big data security. In order to handle and analyse data in the most effective manner, analytics are collaborating with Big Data security. The security of big data is presently a subject of intense interest in the research and development sector. Computing that is data-intensive presents new problems for study in the field of security. The overwhelming amount of data is one of the most significant problems with BDSA. The enormous data made the job of ensuring the security a tough and difficult one. Additionally, BDSA makes it simpler to research weaknesses, which means that breaching a security system is likewise made less difficult by its use. On the other hand, the BDSA is able to find the weaknesses in a system. As a consequence, the BDSA is used often in modern times. The BDSA helps improve the identification of abuse as well as anomalies. As a result of this, there is a need for BDSA as a Service in order to bring down the cost of analysis.

## REFERENCES

[1] Navaneetha Krishnan Rajagopal, Mankeshva Saini, Rosario Huerta-Soto, Rosa Vílchez-Vásquez, J. N. V. R. Swarup Kumar, Shashi Kant Gupta, Sasikumar Perumal, "Human Resource Demand Prediction and Configuration

Model Based on Grey Wolf Optimization and Recurrent Neural Network", Computational Intelligence and Neuroscience, vol. 2022, Article ID 5613407, 11 pages, 2022. https://doi.org/10.1155/2022/5613407

[2] Navaneetha Krishnan Rajagopal, Naila Iqbal Qureshi, S. Durga, Edwin Hernan Ramirez Asis, Rosario Mercedes Huerta Soto, Shashi Kant Gupta, S. Deepak, "Future of Business Culture: An Artificial Intelligence-Driven Digital Framework for Organization Decision-Making Process", Complexity, vol. 2022, Article ID 7796507, 14 pages, 2022. https://doi.org/10.1155/2022/7796507

[3] EshragRefaee, Shabana Parveen, Khan Mohamed Jarina Begum, Fatima Parveen, M. Chithik Raja, Shashi Kant Gupta, Santhosh Krishnan, "Secure and Scalable Healthcare Data Transmission in IoT Based on Optimized Routing Protocols for Mobile Computing Applications", Wireless Communications and Mobile Computing, vol. 2022, Article ID 5665408, 12 pages, 2022. https://doi.org/10.1155/2022/5665408

[4] Rajesh Kumar Kaushal, Rajat Bhardwaj, Naveen Kumar, Abeer A. Aljohani, Shashi Kant Gupta, Prabhdeep Singh, Nitin Purohit, "Using Mobile Computing to Provide a Smart and Secure Internet of Things (IoT) Framework for Medical Applications", Wireless Communications and Mobile Computing, vol. 2022, Article ID 8741357, 13 pages, 2022. https://doi.org/10.1155/2022/8741357

[5] BramahHazela et al 2022 ECS Trans. 107 2651 https://doi.org/10.1149/10701.2651ecst

[6] Ashish Kumar Pandey et al 2022 ECS Trans. 107 2681 https://doi.org/10.1149/10701.2681ecst

[7] G. S. Jayesh et al 2022 ECS Trans. 107 2715 https://doi.org/10.1149/10701.2715ecst

[8] Shashi Kant Gupta et al 2022 ECS Trans. 107 2927 https://doi.org/10.1149/10701.2927ecst

[9] S. Saxena, D. Yagyasen, C. N. Saranya, R. S. K. Boddu, A. K. Sharma and S. K. Gupta, "Hybrid Cloud Computing for Data Security System," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-8, doi: 10.1109/ICAECA52838.2021.9675493.

[10] S. K. Gupta, B. Pattnaik, V. Agrawal, R. S. K. Boddu, A. Srivastava and B. Hazela, "Malware Detection Using Genetic Cascaded Support Vector Machine Classifier in Internet of Things," 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), 2022, pp. 1-6, doi: 10.1109/ICCSEA54677.2022.9936404.

[11] Natarajan, R.; Lokesh, G.H.; Flammini, F.; Premkumar, A.; Venkatesan, V.K.; Gupta, S.K. A Novel Framework on Security and Energy Enhancement Based on Internet of Medical Things for Healthcare 5.0. *Infrastructures***2023**, *8*, 22. https://doi.org/10.3390/infrastructures8020022

[12] V. S. Kumar, A. Alemran, D. A. Karras, S. Kant Gupta, C. Kumar Dixit and B. Haralayya, "Natural Language Processing using Graph Neural Network for Text Classification," 2022 International Conference on Knowledge Engineering and Communication Systems (ICKES), Chickballapur, India, 2022, pp. 1-5, doi: 10.1109/ICKECS56523.2022.10060655.

[13] M. Sakthivel, S. Kant Gupta, D. A. Karras, A. Khang, C. Kumar Dixit and B. Haralayya, "Solving Vehicle Routing Problem for Intelligent Systems using Delaunay Triangulation," 2022 International Conference on Knowledge Engineering and Communication Systems (ICKES), Chickballapur, India, 2022, pp. 1-5, doi: 10.1109/ICKECS56523.2022.10060807.

[14] S. Tahilyani, S. Saxena, D. A. Karras, S. Kant Gupta, C. Kumar Dixit and B. Haralayya, "Deployment of Autonomous Vehicles in Agricultural and using Voronoi Partitioning," 2022 International Conference on Knowledge Engineering and Communication Systems (ICKES), Chickballapur, India, 2022, pp. 1-5, doi: 10.1109/ICKECS56523.2022.10060773.

[15] V. S. Kumar, A. Alemran, S. K. Gupta, B. Hazela, C. K. Dixit and B. Haralayya, "Extraction of SIFT Features for Identifying Disaster Hit areas using Machine Learning

Techniques," 2022 International Conference on Knowledge Engineering and Communication Systems (ICKES), Chickballapur, India, 2022, pp. 1-5, doi: 10.1109/ICKECS56523.2022.10060037.

[16] V. S. Kumar, M. Sakthivel, D. A. Karras, S. Kant Gupta, S. M. Parambil Gangadharan and B. Haralayya, "Drone Surveillance in Flood Affected Areas using Firefly Algorithm," 2022 International Conference on Knowledge Engineering and Communication Systems (ICKES), Chickballapur, India, 2022, pp. 1-5, doi: 10.1109/ICKECS56523.2022.10060857.

[17] Parin Somani, Sunil Kumar Vohra, Subrata Chowdhury, Shashi Kant Gupta. "Implementation of a Blockchain-based Smart Shopping System for Automated Bill Generation Using Smart Carts with Cryptographic Algorithms." CRC Press, 2022. https://doi.org/10.1201/9781003269281-11.

[18] Shivlal Mewada, Dhruva Sreenivasa Chakravarthi, S. J. Sultanuddin, Shashi Kant Gupta. "Design and Implementation of a Smart Healthcare System Using Blockchain Technology with A Dragonfly Optimization-based Blowfish Encryption Algorithm." CRC Press, 2022. https://doi.org/10.1201/9781003269281-10.

[19] Ahmed Muayad Younus, Mohanad S.S. Abumandil, Veer P. Gangwar, Shashi Kant Gupta. " AI-Based Smart Education System for a Smart City Using an Improved Self-Adaptive Leap-Frogging Algorithm." CRC Press, 2022. https://doi.org/10.1201/9781003252542-14.

[20] Rosak-Szyrocka, J., Żywiołek, J., & Shahbaz, M. (Eds.). (2023). Quality Management, Value Creation and the Digital Economy (1st ed.). Routledge. https://doi.org/10.4324/9781003404682

[21] Dr. Shashi Kant Gupta, Hayath T M., Lack of it Infrastructure for ICT Based Education as an Emerging Issue in Online Education, TTAICTE. 2022 July; 1(3): 19-24. Published online 2022 July, doi.org/10.36647/TTAICTE/01.03.A004

[22] Hayath T M., Dr. Shashi Kant Gupta, Pedagogical Principles in Learning and Its Impact on Enhancing Motivation of Students, TTAICTE. 2022 October; 1(2): 19-24. Published online 2022 July, doi.org/10.36647/TTAICTE/01.04.A004

[23] Shaily Malik, Dr. Shashi Kant Gupta, "The Importance of Text Mining for Services Management", TTIDMKD. 2022 November; 2(4): 28-33. Published online 2022 November doi.org/10.36647/TTIDMKD/02.04.A006

[24] Dr. Shashi Kant Gupta, Shaily Malik, "Application of Predictive Analytics in Agriculture", TTIDMKD. 2022 November; 2(4): 1-5. Published online 2022 November doi.org/10.36647/TTIDMKD/02.04.A001

[25] Dr. Shashi Kant Gupta, Budi Artono, "Bioengineering in the Development of Artificial Hips, Knees, and other joints. Ultrasound, MRI, and other Medical Imaging Techniques", TTIRAS. 2022 June; 2(2): 10–15. Published online 2022 June doi.org/10.36647/TTIRAS/02.02.A002

[26] Dr. Shashi Kant Gupta, Dr. A. S. A. Ferdous Alam, "Concept of E Business Standardization and its Overall Process" TJAEE 2022 August; 1(3): 1–8. Published online 2022 August