WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Big Data Security Challenges

Hariom R. Mishra

Department of Information Technology, B.K. Birla College of Arts, Science & Commerce (Autonomous), Kalyan, 421301, Maharashtra, India.

DOI: https://doi.org/10.55248/gengpi.2022.3.10.31

ABSTRACT:

Big data is usually defined as a huge volume of data that is constantly growing in real time and is difficult to store, retrieve and manage using common database techniques. Traditional technology areas are being transformed by big data technologies, and their effective use will require new security models and security design methodologies to address new security challenges. Big data is a collection of huge volumes of data that can be divided into several categories of structured and unstructured data. The article begins with a definition of big data and then looks at characteristics such as veracity, volume, variety, and dynamics that have the greatest impact on big data security. This huge volume of data cannot be captured, stored or analyzed by conventional database systems. Big data continues to grow along with the growth of the Internet. Big data analytics provide organizations such as government and businesses with new tools to analyze unstructured data. Privacy and security of big data are attracting more and more attention. In this phase of computing, where we are moving to zeta bytes from Giga/Tera/Peta/Exabytes, the risks have simultaneously increased. Next, we discuss several potential strategies and tactics for protecting the privacy and security of big data.

Keywords: big data; the value of big data; security and challenges.

1. Introduction

Today, the phrase "big data" is practically ubiquitous in our daily lives. Around 2005, the phrase "Big Data" was used to describe various huge data sets that are so complicated and large that they are almost difficult to manage and process using conventional data management methods. Such infrastructure is sometimes referred to as Big Data Infrastructure (BDI) or Scientific Data Infrastructure (SDI). Big Data is a field where the vocabulary is not well established. The NIST Big Data Working Group, which has just been established, is likely to find a solution to this problem. Many software companies are developing big data applications and solutions to make analytics accessible to everyone. Big Data is the term for huge, diverse sets of data that are collected through many channels, including call logs, social media platforms, websites, e-checkouts, sensors and product purchases. Three distinct characteristics of big data are volume, velocity, and variety.

Volume: Big data contains unspecified and unfiltered amounts of information. The data collected varies from company to company. So the effort made is unique. Still, it is important to keep valuable data in a huge heap. Organizations must process this vast amount of information to solve their business challenges.

Velocity: It is the speed at which data is generated and collected. Mobile devices, SaaS solutions, e-commerce transactions, and IoT devices are some of the primary sources of real-time data collection. The speed at which data is generated on a large scale requires real-time manipulation and processing to improve data analysis.

Diversity: Traditional data types consist of structured data suitable for relational databases. However, semi-structured and unstructured data in the landscape require additional pre-processing to transform the received information into a digestible format. Structured data can be manipulated quickly, but semi-structured and unstructured data must be transformed into predefined models or formats before they can be transformed into actionable information.

Big Data:

The definition of big data is data that contains more variety and comes in ever-increasing quantity and speed. Big data refers to data that is so large, fast, or complex that it is difficult or impossible to process using traditional methods. Accessing large amounts of information and storing it for analysis has been around for a long time. However, the concept of big data gained momentum in the early 2000s when industry analyst Doug Laney formulated the current definition of big data as his three his vs. Big data is a term that describes the large volumes of data that are difficult to manage, including both structured and unstructured data that inundate organizations on a daily basis. Big data is a term used in traditional information processing application programming to describe any overwhelmingly large or complex amount of data that can be meaningfully managed. While the importance of big data is generally recognized, there are still differing views on its definition. Research scientists, data analysts and technologists define big data

differently as different interests. On Wikipedia, big data is an umbrella term that refers to collections of data sets so large and complex that they are difficult to process using traditional data processing applications. The importance of big data is not just about the amount of data. The value is in how you use it. The current buzz around the term "big data" refers to the use of predictive analytics, the study of customer behavior, or certain advanced data analysis techniques that separate incentives from data and, in some cases, data sets of a certain size. they tend to bend. Big data is the idea of storing, processing and decomposing huge amounts (such as exabytes) of data that are hardly comprehensible to traditional RDBMS. Data Sources Cloud Data Storage Web Data Visualization API Analysis/Reporting Query Engine B. Mahout Distributed for large unstructured data sets such as Hive, NoSQL Hadoop Distributed File System, fault tolerant database. Distributed Configuration and Synchronization Services A programming model for processing large data sets using parallel distributed algorithms on clusters such as the MapReduce framework. It is used to manage semi-structured or unstructured data in addition to organized data. Big data can be divided into several classifications such as data sources, content locations, data stores, data organization, and data preparation. Big data in industry refers to the management of complex technical processes and objects or systems. Modern computer-aided manufacturing generates vast amounts of data that typically need to be stored or stored for effective quality control and diagnostics in the event of a malfunction or accident. Similar to e-science, many industrial applications/scenarios require the collaboration or interaction of many workers and engineers.

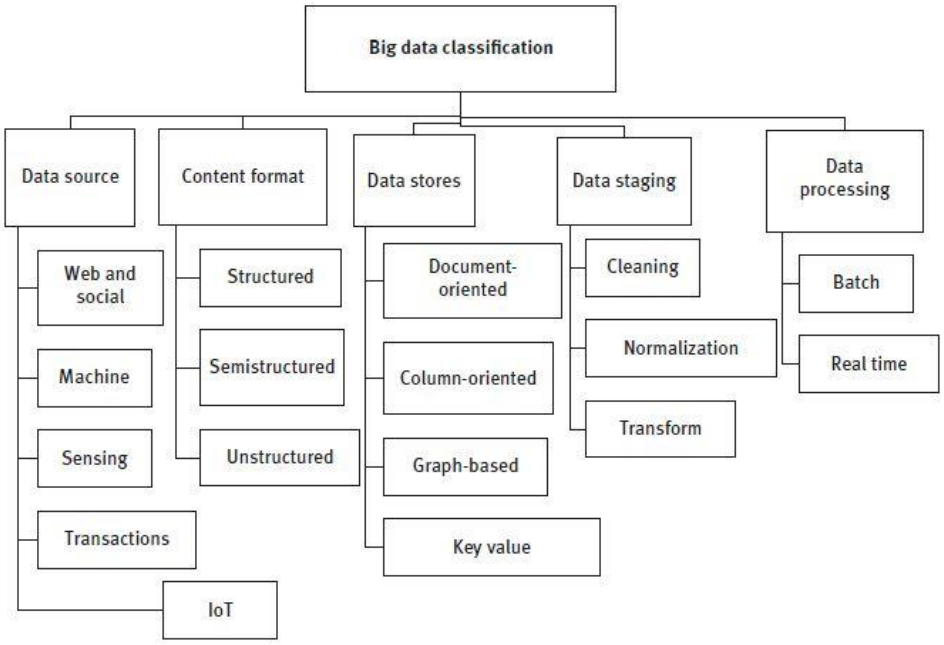

Figure: Big data classification

Figure: Life cycle of big data.

SAFETY CHALLENGES:

Perhaps the most difficult and vexing issue in big data is security. Security mechanisms in big data technologies are generally weak. Big data security frameworks can be divided into different classifications. Some of these can be called infrastructure security, privacy, data management, integrity and responsive security. The emergence of big data has brought new challenges in the field of data security. There is a growing need for research into technologies that can process huge amounts of data and back them up efficiently. Current data backup technologies are slow when applied to large volumes of data. Government agencies, the medical industry, biomedical researchers, and private companies are investing enormous resources in collecting, aggregating, and sharing large amounts of personal data to reap the enormous benefits of big data. Because data is stored on thousands of nodes, "authenticating, authorizing and encrypting data on these nodes becomes a daunting task." In the business world, private companies typically do not share data or expertise. When handling data, companies always try to maintain control over information assets. Third-party sharing facilities such as the cloud or dedicated tools can be used, but additional steps must be taken to ensure workplace security and privacy, such as sanitizing incoming and outgoing data.

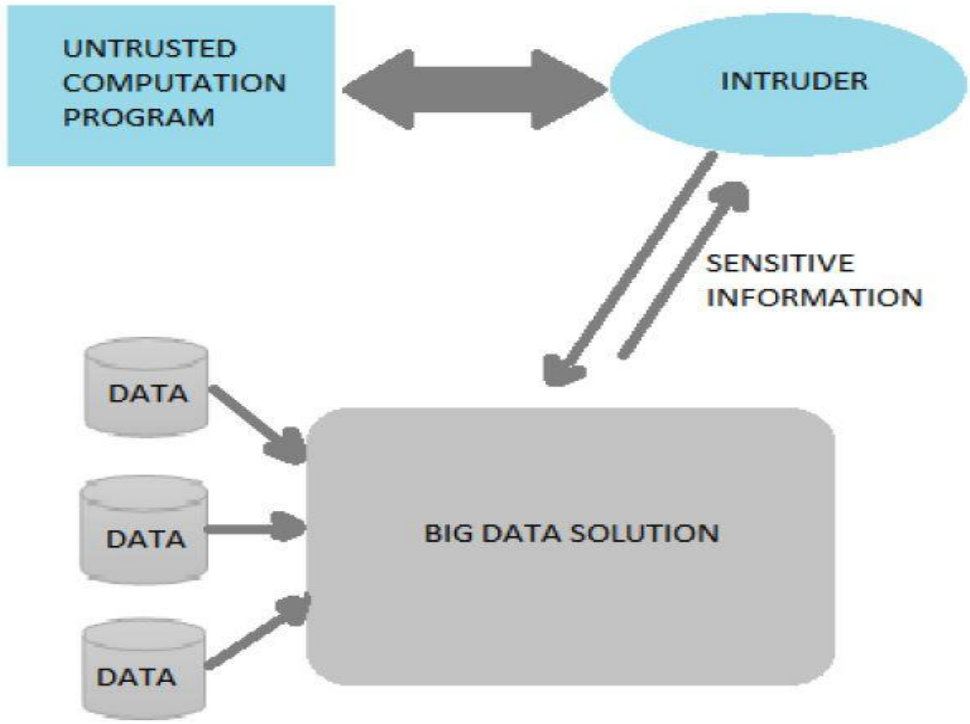

Figure: INSECURE COMPUTATION

1.1. Problem description and overview:

One of the most pressing problems with big data is how to properly store all these massive data sets. The amount of data stored in corporate data centers and databases is growing rapidly. These datasets grow exponentially over time, making them very difficult to deal with. Most of the data is unstructured and comes from documents, videos, audio, text files and other sources, i.e. I can't find it in the database. This can pose huge big data analytics challenges that need to be addressed as soon as possible. Otherwise, it can slow down the growth of your business.

1.2. Objective:

I. Businesses can use big data in their systems to improve operations, provide better customer service, create personalized marketing campaigns and other actions that can ultimately increase sales and profits. Big data is often stored in data lakes. Data warehouses are typically built on top of relational databases and contain only structured data, but data lakes can support a wide variety of data types. The processing of big data places high demands on the basic computing infrastructure.

II. The main goal of big data is to tell stories with numbers. This technology allows organizations or individuals to receive, store, transform and analyze large amounts of data to solve specific problems. A data-driven approach to understanding your business. You can create predictive data models and look for future trends. Future devices will be completely based on big data.

1.3 Solution:

Traditional security mechanisms that focus on protecting small amounts of data that are static in nature (non-streaming) fall short in practice. Big data security challenges are compounded by the velocity, volume and variety of big data. Organizations are hiring more cybersecurity professionals to protect their data. Other security measures include: Data encryption Data isolation Identity and access control Implementation of endpoint security Real-time security monitoring Use security tools such as IBM Guardian. With converged and hyper-converged infrastructure and software-defined storage, you can easily handle rapid data growth. In addition, compression, layering, and deduplication help reduce disk space consumption and lower storage costs. Organizations are also using tools such as big data analytics software, Hadoop and NoSQL databases, Spark applications, AI, machine learning and BI to address this issue. Here are ways organizations can face big data security challenges:

• • Hire more cyber security professionals

• • Data encryption and isolation

• • Identity and authorization check

• • Endpoint security

• • Real-time monitoring

• • Using big data security tools such as IBM Guardium

Conclusions

Security is the most important requirement in big data. As data storage and computing environments become cheaper, cloud environments are better able to store and share systems and analytics applications, software applications become more connected, data security, access control, compression – encryption and compliance. it brings certain challenges that need to be addressed. We deal with it very systematically. Big data security is not a security technology, but a law, but the law cannot keep up with technological developments and varies from country to country. Therefore, safety engineering and other methods are always necessary. There is a large ecosystem for specific big data problems. The topics in this white paper help clarify specific aspects of vulnerabilities across big data infrastructures. Some possible methods and techniques for securing big data have been discussed above. I hope the given problem and solution will serve the researchers and big data fraternity in the best way. Big data has a big space and R&D has a big space.

References

1. https://www.aimspress.com/fileOther/PDF/Math/math-04-03-860.pdf
2. https://www.researchgate.net/profile/Vaibhav-Hans/publication/295907307_Big_Data_Security_-_Challenges_and_Recommendations/links/56d00b3408ae059e375a513a/Big-Data-Security-Challenges-and-Recommendations.pdf
3. https://www.iaeme.com/MasterAdmin/Journal_uploads/IJCET/VOLUME_9_ISSUE_4/IJCET_09_04_025.pdf
4. Kantarcioglu, M. and Ferrari, E., 2019. Research Challenges at the Intersection of Big Data, Security and Privacy. *Frontiers in Big Data*, 2.
5. Bentotahewa, V., Hewage, C. and Williams, J., 2021. Solutions to Big Data Privacy and Security Challenges Associated With COVID-19 Surveillance Systems. *Frontiers in Big Data*, 4.
6. https://www.dataversity.net/big-data-security-challenges-and-solutions/
7. https://www.datamation.com/big-data/big-data-security
8. https://cprimestudios.com/blog/big-data-security-biggest-challenges-and-best-practices
9. https://www.analyticsinsight.net/big-data-security-challenges-how-to-overcome-them/
10. https://addepto.com/blog/big-data-security-issues-and-challenges/
11. https://www.scnsoft.com/blog/big-data-challenges-and-their-solutions