Name: Kartik Vedi Student ID:300374518  
CSIS 4260 Section:001 Seminar: 5

# Topic: Big Data Security and Privacy: Challenges, Solutions, and Impacts.

# Executive Summary

Big data security is the term used to describe the procedures and controls implemented to guard against theft, tampering, and unauthorised access to enormous amounts of data. It entails protecting data at every stage of its lifespan, which includes gathering, storing, processing, and analysing it. On the other hand, big data privacy is concerned with making sure that sensitive data and personal information about persons is handled in compliance with ethical standards and privacy laws. This entails getting permission before collecting data, anonymizing it when appropriate, and putting safeguards in place to stop data breaches and unauthorised publication.

# **Identify and describe the big security challenges and concerns in big data.**

In the digital age, data is growing exponentially, posing previously unheard-of difficulties for data security and privacy. The enormous amount of data generated every day is one of the biggest obstacles to big data security. The amount of data generated by the widespread use of social media, digital transactions, and internet-connected gadgets is astounding, making it more and more challenging for organisations to efficiently monitor and safeguard all data points. Because of the volume of data, conventional security measures are insufficient, necessitating creative solutions to properly protect information.

Furthermore, the variety of data formats and sources adds to the difficulty of protecting large amounts of data. Structured databases, unstructured text, multimedia files, and streaming data from Internet of Things devices are just a few of the sources from which data might come. various data types may have various security needs and risks, therefore a diversified strategy to data protection is required. For example, access controls and encryption may be used to secure organised data in databases, but sophisticated methods such as natural language processing may be needed to identify and categorise sensitive information in unstructured data, such as text documents or audio recordings.

In distributed and decentralised big data systems, protecting data integrity and avoiding unwanted access are critical issues. Data in distributed systems is kept on several nodes or servers, frequently spread across different regions or cloud computing infrastructures. Because of this decentralisation, it becomes more difficult to enforce access rules in a variety of contexts and preserve data consistency. Furthermore, the dynamic character of big data ecosystems—where information is always moving between many parts and platforms—raises the possibility of unwanted access or manipulation. Robust authentication methods, encryption strategies, and audit trails to trace the provenance and modifications of data during its lifecycle are necessary for preserving data integrity.

Insider attacks, gaps in data processing and storage systems, and potential exploitation aimed at data transmission protocol flaws further exacerbate the security picture in big data environments. Insider threats, whether deliberate or accidental, present serious concerns to data security because they allow authorised users with rightful access credentials to misuse those privileges or unintentionally reveal confidential information. Attackers can exploit weaknesses in data processing and storage systems, such as obsolete software or improperly designed databases, to obtain unauthorised access to data. Insecure data transmission protocols further raise the risk of data tampering or interception while in transit, such as poor authentication procedures or unencrypted network interactions.

To tackle these obstacles, a comprehensive strategy that incorporates advanced technology, strong regulations, and preemptive risk mitigation techniques is needed. Cutting-edge technologies with the potential to improve data security in big data contexts include blockchain, artificial intelligence, and machine learning. Proactive threat detection and response are made possible by machine learning algorithms, which can analyse enormous volumes of data to find abnormalities and patterns suggestive of possible security concerns. Similar to this, blockchain technology provides decentralised, unchangeable data storage options that use cryptography to guarantee data integrity and resistance to tampering.

To mitigate risks and protect data privacy, extensive rules and compliance frameworks must be implemented in addition to technology solutions. Organisations must implement measures like data encryption, access controls, and data minimization practices in order to comply with strict requirements for data protection and privacy imposed by regulatory frameworks like the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). Additionally, encouraging employee training programmes and cultivating a culture of security awareness can aid in reducing insider threats and improving overall cybersecurity posture.

In summary, tackling the security issues posed by big data necessitates a coordinated effort to adjust to the changing threat landscape and put preventative measures in place to protect privacy and data integrity. Organisations can guarantee the availability, confidentiality, and integrity of sensitive data in the digital age by utilising cutting-edge technologies, strong regulations, and proactive risk management techniques.

**Identify and describe technologies AND policy solutions that will help secure big data.**

**Innovations in technology are essential to improving large data environments' security. One essential tool for safeguarding data whether it's in transit or at rest is encryption. Organisations can make data unreadable for unauthorised users even in the event that it is intercepted or accessed without the required authority by encrypting it. Robust protection against unauthorised access and data breaches is provided by advanced encryption algorithms like Rivest-Shamir-Adleman (RSA) and Advanced Encryption Standard (AES).**

**For large data systems to prevent unwanted access to sensitive data, access control measures are crucial. Granular access regulations based on the roles, responsibilities, and attributes of users are often enforced through the use of role-based access control (RBAC) and attribute-based access control (ABAC). Organisations can lower the risk of insider threats and unauthorised data breaches by limiting access to data to only authorised workers through the use of access controls.**

**Systems for detecting and preventing intrusions (IDPS) are essential for quickly recognising and addressing security risks. IDPS keeps a close eye on user activity, system logs, and network traffic in order to spot anomalies and unusual activity that could point to security breaches. In order to lessen the effects of an attack and stop additional harm, IDPS can automatically respond to security threats by banning questionable IP addresses or quarantining infected systems.**

**Big data settings are further secured by implementing strong authentication techniques like multi-factor authentication (MFA). Before getting access to sensitive information or systems, MFA requires users to give several kinds of authentication, such as a password plus a biometric scan or a one-time passcode texted to a registered device. Multi-factor authentication, or MFA, lowers the risk of brute-force attacks or compromised credentials leading to unauthorised access.**

**Regulations like the California Consumer Privacy Act (CCPA) and the General Data Protection Regulation (GDPR) establish strict guidelines for data security and privacy procedures. Organisations are required by these regulations to put in place measures like purpose limitation, which limits the use of data to predetermined purposes specified at the time of collection, and data minimization, which entails collecting and retaining only the minimum amount of data necessary for a specific purpose. Furthermore, the CCPA and GDPR place a strong emphasis on transparency in data processing operations, requiring businesses to tell people about the reasons for data processing and, if necessary, get their express consent. Following these rules promotes accountability and trust in big data ecosystems while also improving data security.**

**Identify and discuss the impacts of ANY recent security breach or privacy violations in big data industry (last 5 years).**

**Significant security lapses and privacy violations have shook the big data sector in recent years, putting millions of people at risk of financial crime and identity theft. One well-known instance is the 2019 Capital One data breach, in which over 100 million customers' personal information was compromised by hackers taking advantage of flaws in the company's cloud-based storage systems. This hack demonstrated how crucial it is to have strong encryption and access controls in place to protect sensitive data kept in cloud environments. It also brought attention to the necessity of proactive security measures and ongoing monitoring in order to quickly identify and address security concerns.**

**Comparably, the 2017 Equifax data breach, which impacted around 147 million people, stands out as one of the biggest and most significant breaches in recent memory. The company's web application had a known vulnerability that was not patched, which led to the breach and gave hackers access to private data without authorization. This incidence brought to light how important it is to apply vulnerability management techniques and timely security patches to stop data breaches. It also underlined how crucial it is to put in place thorough security policies and carry out frequent security audits in order to find and fix any flaws before bad actors may take advantage of them.**

**These well-publicized hacks are harsh reminders of the constant dangers and weaknesses found in large data environments. Organisations need to prioritise cybersecurity and take preventative measures to shield sensitive data from unauthorised access and exploitation as they continue to gather and analyse massive amounts of data. Organisations may reduce the risks associated with big data security breaches and protect consumer privacy and confidence by investing in strong security technology, enforcing strict access rules, and being on the lookout for new threats.**

# **References**

Author(s): CNN Business Title: Capital One data breach affects 100 million; bank says hacker accessed credit card application data URL: https://edition.cnn.com/2019/07/29/business/capital-one-data-breach/index.html

Author(s): CNN Business Title: Capital One data breach affects 100 million; bank says hacker accessed credit card application data URL: https://edition.cnn.com/2019/07/29/business/capital-one-data-breach/index.html