Journal of Science and Technology
ISSN: 2456-5660 Volume 7, Issue 2 (March -2022)
www.jst.org.in *DOI: https://doi.org/10.46243/jst.2022.v7.i02.pp423- 436*

BIG DATA PRIVACY AND SECURITY USING CONTINUOUS DATA PROTECTION DATA OBLIVIOUSNESS METHODOLOGIES

Swapna Narla

Tek Yantra Inc, California, USA

swapnanarla8883@gmail.com

To Cite this Article

Swapna Narla, **"BIG DATA PRIVACY AND SECURITY USING CONTINUOUS DATA PROTECTION DATA OBLIVIOUSNESS METHODOLOGIES"** *Journal of Science and Technology,*

Vol. 07, Issue 02,-March 2022, pp423-436

Article Info

Received:28-02-2022 Revised:12-03-2022 Accepted:20-03-2022 Published:27-03-2022

ABSTRACT

Large databases must be protected from breaches, unauthorised access, and misuse in the big data era. With a focus on Continuous Data Protection (CDP) and Data Obliviousness, this study investigates cutting-edge techniques for improving data security and privacy. By guaranteeing real-time data backups, CDP lowers the possibility of data loss due to cyberattacks or system malfunctions. Data Obliviousness processes data securely without disclosing sensitive information by utilising methods including homomorphic encryption, secure multiparty computation (SMC), and differential privacy. When these strategies are combined in big data environments, a strong security framework is produced, regulatory compliance with the likes of CCPA and GDPR is guaranteed, and cyber threat resistance is improved.

Keywords: Big Data Privacy, Data Security, Continuous Data Protection (CDP), Data Obliviousness, Homomorphic Encryption, Secure Multiparty Computation (SMC)

1 INTRODUCTION

Large-scale datasets must be safeguarded against misuse, illegal access, and breaches using crucial big data privacy and security procedures. Protecting personal information is essential for maintaining privacy in the big data world, where enormous volumes of varied and frequently sensitive data are involved. Respecting regulatory frameworks such as the CCPA or GDPR and using methods like encryption and anonymization to reduce the possibility of data re-identification and effectively limit access are all part of this. Protecting the entire data infrastructure from dangers like viruses, insider assaults, and unauthorized access is the main goal of security measures. Strong authentication must be used, data must be encrypted while it is being sent and stored, frequent security audits must be carried out, and predetermined security guidelines must be followed. Businesses and organizations can manage big data privacy and security efficiently by integrating technological solutions, organizational rules, and legal compliance. This allows them to preserve data integrity and respect privacy rights while gaining insights from massive datasets.

Data security and privacy in today's digital environments are intended to be strengthened by continuous data protection and data obliviousness. By regularly backing up data in real-time or almost real-time, continuous data protection

(CDP) makes sure that every update or change is recorded as soon as possible. Due to its ability to quickly restore data to any point in time, this technique reduces the risk of data loss from system failures, human mistake, or cyber attacks. The goal of data obliviousness is to protect data privacy by making sure algorithms and systems function without being aware of the particular data they are handling. This shields private data from exposure or unwanted access. This is accomplished by using methods like homomorphic encryption, secure multiparty computation (SMC), and differential privacy, which enable safe data processing and analysis while maintaining secrecy.

Essentially, data obliviousness techniques improve privacy safeguards by guaranteeing data processing is safe and free from exposure hazards, while continuous data protection guarantees continuous data backup for resilience against disruptions. The frameworks for data security and privacy in contemporary digital environments are strengthened by these behaviors taken together. Ensuring the security and privacy of enormous volumes of sensitive data has become a critical issue in the big data era. Sophisticated approaches like Continuous Data Protection (CDP) and Data Obliviousness have become essential tools in addressing these issues. These approaches are crucial for preserving confidence in digital settings by protecting data against breaches, illegal access, and privacy violations. Continuous Data Protection (CDP) is the process of continuously backing up data in near-real-time or continuously, such that any changes are immediately recorded. By enabling enterprises to recover data to any point in time, CDP minimizes the impact of data loss due to system failures or cyber incidents, in contrast to traditional approaches that back up data at predetermined intervals.

Conversely, data obliviousness emphasizes the usage of safe processing methods to protect data privacy. It guarantees that algorithms and systems function without being aware of the particular data they manage, safeguarding privacy and averting unwanted access. Alongside the exponential growth in digital data and the development in sophisticated cyber threats, there has been an increasing need for strong data privacy and security measures. The constraints presented by big data environments—where data is diverse, distributed, and voluminous—have led to an evolution of traditional approaches to data protection, such as encryption and access control. To properly integrate CDP and data obliviousness, numerous software programs and frameworks have been created. These comprise data obliviousness products like Microsoft SEAL and IBM's Homomorphic Encryption Toolkit, as well as backup options like Veeam and Commvault for CDP.

Prominent technology corporations such as Google, Amazon, and Microsoft include these approaches into their cloud services to guarantee data protection and adherence to privacy laws. Experts in cybersecurity and academic research also make a contribution by creating sophisticated frameworks and algorithms that improve data security in a variety of industries. The adoption of CDP and data obliviousness still face a number of obstacles despite developments, such as scalability problems when managing enormous datasets, usability issues in a variety of applications, and maintaining compliance with changing data privacy legislation across the globe. Recent developments include using blockchain technology for decentralized data storage, combining AI and machine learning for predictive data protection with CDP, and investigating quantum-resistant encryption techniques to ensure data security in the future.

- Examining current CDP and data obliviousness approaches, suggesting improvements to tackle current issues, and conducting case studies to show useful implementations in various industries are the objectives of this research.

Even while data obliviousness and CDP have a lot to offer, there are still obstacles in the way of obtaining effective and comprehensive data security in big data contexts. In the digital age, resolving these issues is essential to guaranteeing strong data privacy and security.

2 LITERATURE SURVEY

Goel et al. (2021) examine the complexity of Big Data security and privacy, emphasizing concerns with data breaches, legal compliance (such as the CCPA and GDPR), and moral dilemmas with data gathering and privacy. In light of the growing number of data breaches in Big Data environments, they emphasize the necessity of strong security measures. The evaluation addresses permission and transparency in data usage, as well as how strict requirements impact

organizational activities from an ethical standpoint. Additionally, it emphasizes the importance of data masking, anonymization, and encryption as crucial methods for safeguarding private information contained in big datasets. In the future, the writers think about new developments and how AI can improve data security measures.

Big data companies must make investments in data security and privacy if they want to increase their worth and reputation. These expenditures increase resilience, protect sensitive data, and lessen the chance of breaches. Companies can stay out of trouble and keep their legal standing by following legislation such as the CCPA and GDPR. Robust privacy protocols enhance client confidence, allegiance, and competitiveness in the market, guaranteeing enduring viability and drawing in financiers and collaborators dedicated to strong data protection Zhang et al. (2021)**.**

Bentotahewa et al. (2021) explore COVID-19 surveillance systems that use big data and offer solutions to security and privacy issues. They place a strong emphasis on using anonymization and encryption to protect sensitive health data. Legal observance and moral data treatment are ensured by GDPR and HIPAA compliance. Openness in the use of data promotes public trust, which is necessary for efficient monitoring. AI and other technological advancements support data analysis while upholding privacy norms. In order to ensure a strong pandemic response and ethical integrity, ethical problems must be addressed by getting consent and striking a balance between the interests of the public health and individual rights.

In connected Big Data, machine learning plays a critical role in safeguarding privacy. Differential privacy techniques introduce noise into data to safeguard identities while examining broad patterns. Encrypted data can be securely computed using homomorphic encryption without the need for decryption. Federated learning maintains decentralized data and protects privacy by training models locally. Generalization is one anonymization technique that protects identities while exposing patterns. Data utility is ensured while privacy is protected via privacy-preserving algorithms; nonetheless, these algorithms have issues in balancing accuracy and durability against attacks. In order to better safeguard data, future efforts will focus on improving collaborative data analysis and incorporating privacy technologies into widely used machine learning frameworks Biswas et al. (2021).

Strong encryption, stringent access restrictions, anonymization strategies, and adherence to laws like GDPR are all necessary for protecting Big Data and guaranteeing privacy. Keeping massive datasets safe, protecting against cyberattacks, and striking a balance between privacy concerns and data utility are among the challenges. Cutting-edge technologies that safeguard data and enable insightful analysis, such as homomorphic encryption and differential privacy, are constantly developing. In order to successfully address increasing dangers and legal requirements, future work will concentrate on improving privacy technologies and incorporating AI for real-time threat detection Parihar (2021)**.**

Techniques for safeguarding privacy in relation to various data types—structured, unstructured, and semi-structured—are examined in the survey by Cunha et al. (2021). They emphasize the use of differential privacy to strike a balance between data accuracy and individual privacy, anonymization to protect sensitive information while permitting analysis, and encryption for safe data transit and storage. Federated learning is emphasized as a means of processing data collaboratively while preserving anonymity. Achieving compatibility across various data formats and striking a balance between privacy requirements and data utility are challenges. Future work will focus on integrating AI for adaptive security measures and improving privacy solutions for complicated data settings.

Data protection in AI services is examined in Meurisch & Mühlhäuser (2021) survey. Differential privacy is used to protect individual data privacy during analysis, federated learning is used for collaborative model training while protecting privacy, and encryption is used for secure data handling. In order to ensure ethical data practices in AI development, the study emphasizes adherence to GDPR, HIPAA, and CCPA. It also addresses issues such as striking a balance between data utility and privacy requirements and guaranteeing technological compatibility and security across AI applications. In the future, efforts will focus on developing privacy technologies and incorporating AI to improve privacy controls and identify security threats in AI services.

Anonymization, encryption, and other privacy-preserving techniques are examined in Salas and Domingo-Ferrer (2018) study, "Some Basics on Privacy Techniques, Anonymization and their Big Data Challenges," along with the difficulties that arise in big data settings. Because of the possibility of re-identification, traditional anonymization is difficult to apply to huge datasets. Advanced methods for protecting data that yet permit valuable analysis, such as differential privacy and k-anonymity, are covered by the writers. They underline the necessity of scalable privacy solutions that strike a compromise between the requirement to preserve individual privacy and the capacity to efficiently analyze enormous volumes of data.

In their investigation, "Privacy, Space and Time: A Survey on Privacy-Preserving Continuous Data Publishing," Katsomallos et al. (2019) examine methods for safeguarding privacy in real-time data streams, such as those found in tracking systems or social networks. With an emphasis on techniques like k-anonymity and differential privacy modified for continuous data, the study emphasizes the difficulties in preserving privacy while guaranteeing that data stays valuable. Additionally, it talks about privacy problems associated with spatiotemporal data, like location monitoring. In dynamic situations, this survey offers a concise summary of approaches to strike a compromise between privacy and continuous data release.

In their investigation, "Optimal Privacy Preserving Technique Over Big Data Analytics Using Oppositional Fruit Fly Algorithm," Kiran and Devara (2020) describe ways to use the Oppositional Fruit Fly Optimization Algorithm (OFOA) to improve privacy in big data analytics. By optimizing anonymization techniques like k-anonymity and differential privacy, the system ensures robust privacy protection with low data loss. Because it balances privacy and data value, it is efficient and scalable for huge datasets and is appropriate for real-world big data applications both security and analytics are crucial.

In their work "Privacy-Preserving Record Linkage for Big Data: Current Approaches and Research Challenges," Vatsalan et al. (2017) examine various approaches, including anonymization and cryptography, for securely connecting records between databases. In large-scale, high-dimensional datasets in particular, the investigation emphasizes the difficulties in striking a balance between privacy and data accuracy. It also notes that more scalable and effective solutions are required. With applications in the government, banking, and healthcare industries, the text helps readers discover ways to secure sensitive data during record linking.

Cloud-stored data security is covered in Kaaniche and Laurent (2017) investigation, "Data Security and Privacy Preservation in Cloud Storage Environments Based on Cryptographic Mechanisms," which uses cryptographic techniques such homomorphic and searchable encryption. The difficulties in securing data during transfer, storage, and access are emphasized, and the trade-off between robust encryption and system efficiency is discussed. Along with discussing safe data sharing in multi-user settings, the article also makes recommendations for further research on enhancing productivity. It is an invaluable tool for discovering that sensitive data in cloud systems can be safeguarded using cryptography.

Yallamelli (2021) highlights the importance of RSA encryption in enhancing data security within cloud computing. Using prime factorization for encryption and decryption, RSA ensures data confidentiality, integrity, and availability. Widely used in cloud environments like AWS and Azure, RSA improves security, though challenges like scalability and key management require further research.

Gudivaka (2021) investigates how tailored and interesting learning experiences made possible by the integration of AI and Big Data analytics might transform the teaching of music. Real-time feedback, interactive features, and customized teaching approaches are all provided by AI algorithms, which improve student engagement and personalize music instruction for each student.

3 CONTINUOUS DATA PROTECTION METHODOLOGY

Considering big data involves such large and sensitive amounts of data, it is imperative to ensure strong privacy and security protocols. In order to protect massive datasets against illegal access, security lapses, and invasions of privacy, this methodology combines two cutting-edge approaches: Continuous Data Protection (CDP) and Data Obliviousness.

Here is a methodical way to putting these techniques into practice and improving them for efficient data security and privacy management. Ensuring continuous data protection entails regularly backing up data in real-time or almost real-time to record any updates or changes as they occur. By enabling enterprises to quickly restore data to any earlier point in time, this proactive strategy reduces the chance of data loss due to system malfunctions, human mistake, or cyberattacks.

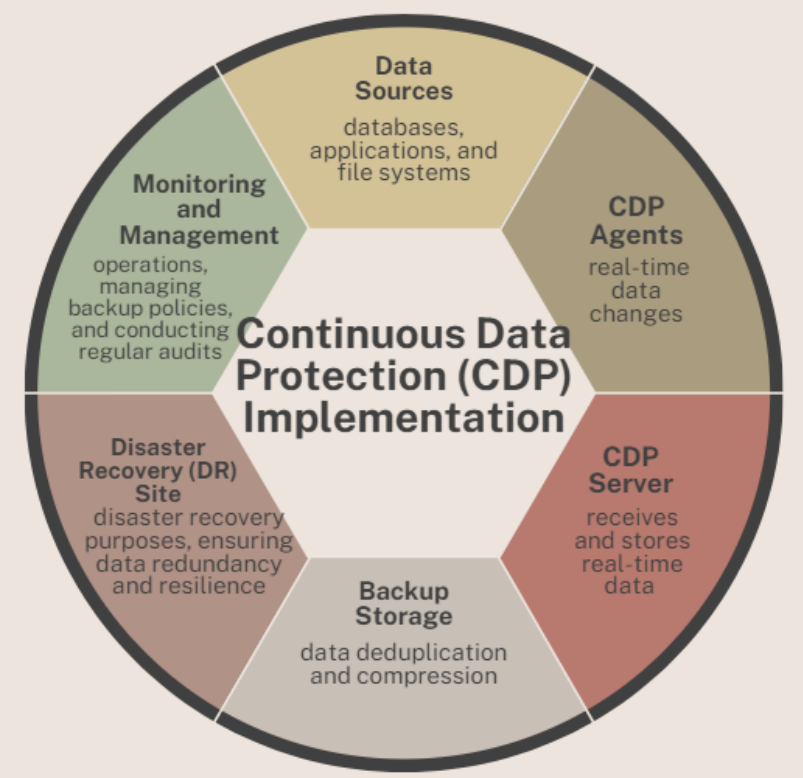

Figure 1. Continuous Data Protection (CDP)

This figure 1 illustrates the implementation of Continuous Data Protection (CDP) within a big data environment. Key components include: Data Sources various data sources such as databases, applications, and file systems from which data changes are continuously monitored. CDP Agents deployed on data sources to capture real-time data changes and transmit them to the CDP system. CDP server central server or cloud-based service that receives and stores real-time data updates. Backup storage secure storage infrastructure where backed-up data is stored with mechanisms for data deduplication and compression.

Disaster Recovery (DR) Site secondary site or cloud region used for disaster recovery purposes, ensuring data redundancy and resilience. Monitoring and Management tools and interfaces for monitoring CDP operations, managing backup policies, and conducting regular audits. Continuous Data Protection (CDP) ensures that every change made to data is backed up in real-time. The equation to represent this can be expressed as:

$$D_{t+1} - D_t + \Delta D_t \quad (1)$$

Where:

- D_i represents the data at time t.
- ΔD_t represents the data change (deita) at time t.

Data Deduplication Equation

Data deduplication is a process that eliminates redundant copies of data. The efficiency of deduplication can be represented by:

$$E_{dedup} - \frac{S_{mine}}{S_{step}} \quad (2)$$

Where:

- E_{dedup} is the deduplication efficiency.
- S_{orig} is the original data size.
- Sindup is the deduplicated data size.

2. Data Obliviousness Mathematical Equations

Homomorphic Encryption Equation

Homomorphic Encryption allows computations to be carried out on ciphertexts, generating an encrypted result which, when decrypted, matches the result of operations performed on the plaintext. This can be represented as:

$$Enc\ (a) \oplus Enc\ (b) - Enc\ (a \circ b) \quad (3)$$

Where:

- Enc(a) and $Enc\ (b)$ are the encrypted values of a and b.
- $\oplus$ is the homomorphic operation corresponding to o in plaintext (e.g., addition, multiplication). Differential Privacy Equation

Differential privacy ensures that the removal or addition of a single data point does not significantly affect the outcome of any analysis. It is represented by:

$$Pr\ [A(D_1) \in S] \leq e^{e} \cdot Pr\ [A(D_2) \in S] + \delta \quad (4)$$

Where:

- A is the algorithm.
- D_1 and D_2 are datasets differing by one element.
- S is a subset of possible outputs.
- ϵ (epsilon) is the privacy loss parameter.
- δ (delta) is a small probability.

Building a solid data backup system that can manage continuous data streams is the first step towards implementing CDP successfully. This entails picking appropriate backup programs, such as Veeam, Commvault, or cloud-based services that provide backup and replication of data in real-time. Deploying agents or connectors that continuously monitor data changes across several data sources is necessary to set up mechanisms for real-time data collecting. These agents immediately capture and send data updates to the backup repository, making sure that no changes are missed and providing quick recovery possibilities. Data compression and deduplication strategies optimize storage and bandwidth consumption. While compression minimizes the size of data transfer without sacrificing data integrity, deduplication removes superfluous data blocks. In CDP environments, these procedures improve backup performance and reduce storage costs.

Journal of Science and Technology
ISSN: 2456-5660 Volume 7, Issue 2 (March -2022)
www.jst.org.in *DOI: https://doi.org/10.46243/jst.2022.v7.i02.pp423- 436*

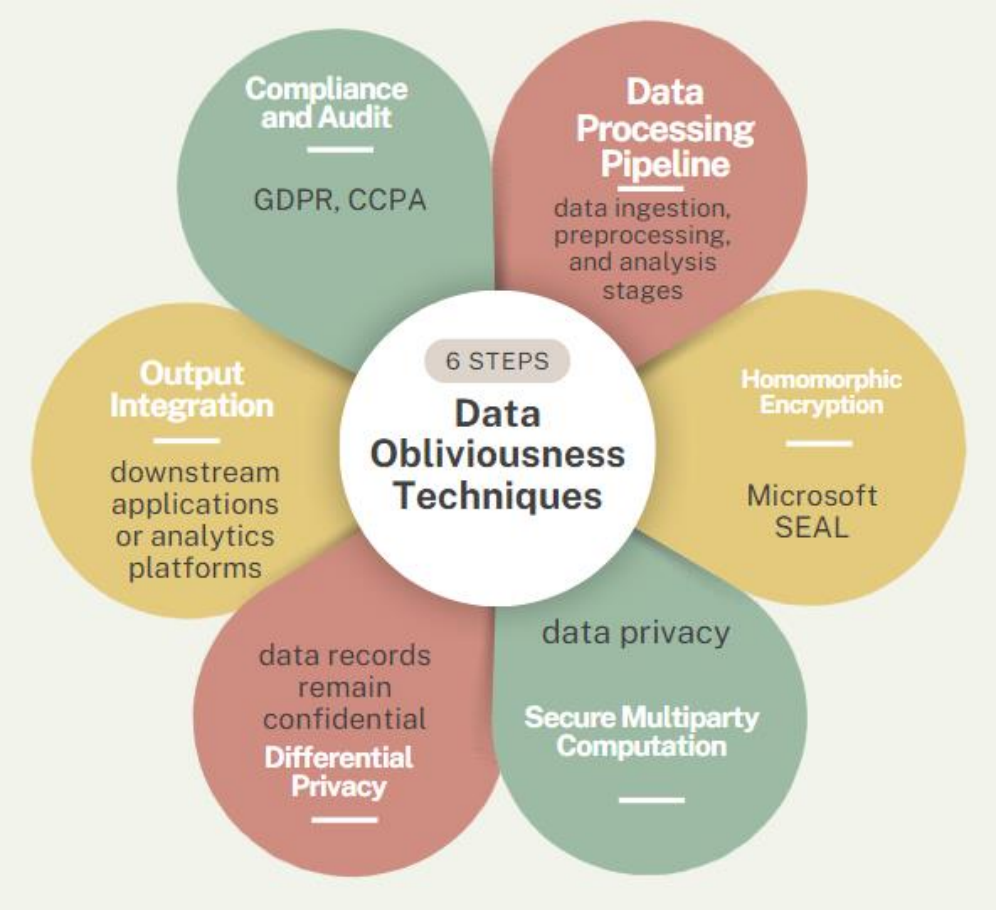

Figure 2. Data Obliviousness techniques

This figure 2 illustrates the implementation of Data Obliviousness techniques for ensuring data privacy within big data processing: Data processing pipeline workflow depicting data ingestion, preprocessing, and analysis stages. Homomorphic encryption integration of homomorphic encryption libraries (e.g., Microsoft SEAL) within data processing nodes to perform computations on encrypted data. Secure Multiparty Computation (SMC) deployment of SMC protocols to enable collaborative computations across multiple parties while preserving data privacy. Differential privacy incorporation of differential privacy mechanisms to add noise to query responses, ensuring individual data records remain confidential. Output integration secure integration of processed and anonymized data outputs into downstream applications or analytics platforms. Compliance and audit tools and protocols for ensuring compliance with data privacy regulations (e.g., GDPR, CCPA) and conducting privacy audits.

Planning comprehensive strategies and procedures for data recovery during catastrophes or data breaches is a necessary step in integrating CDP with disaster recovery planning. To guarantee prompt data restoration and little operational disruption, this involves setting recovery point objectives (RPOs) and recovery time objectives (RTOs). Data obliviousness approaches make sure algorithms and systems function without being aware of the particular data they handle, protecting users' privacy. The techniques and resources utilized in big data contexts to attain data obliviousness are examined in this section. Homomorphic encryption protects data confidentiality during processing by enabling computations on encrypted data without the need for decryption. Adopting frameworks such as IBM's Homomorphic Encryption Toolkit or Microsoft SEAL guarantees secure data management while protecting privacy throughout analytical procedures. Multiple parties can compute functions over their inputs while maintaining the privacy of those inputs through the use of secure multiparty computation. During collaborative computations, this technique makes sure that sensitive data stays encrypted, preventing data loss or disclosure to unauthorized parties.

Differential privacy adds noise to query responses to preserve personal information while producing statistically significant results. Aggregate insights can be obtained without jeopardizing the confidentiality of individual records thanks to strategies like calibrating the noise in data queries. For smooth operation and efficient data privacy and security, the integration of CDP and data obliviousness methodologies necessitates careful planning and implementation. The process of creating a robust and scalable system design requires the integration of data obliviousness frameworks with CDP techniques. This entails deciding which technologies work together, outlining data flows, and guaranteeing that instruments for protecting privacy and backup systems are compatible.

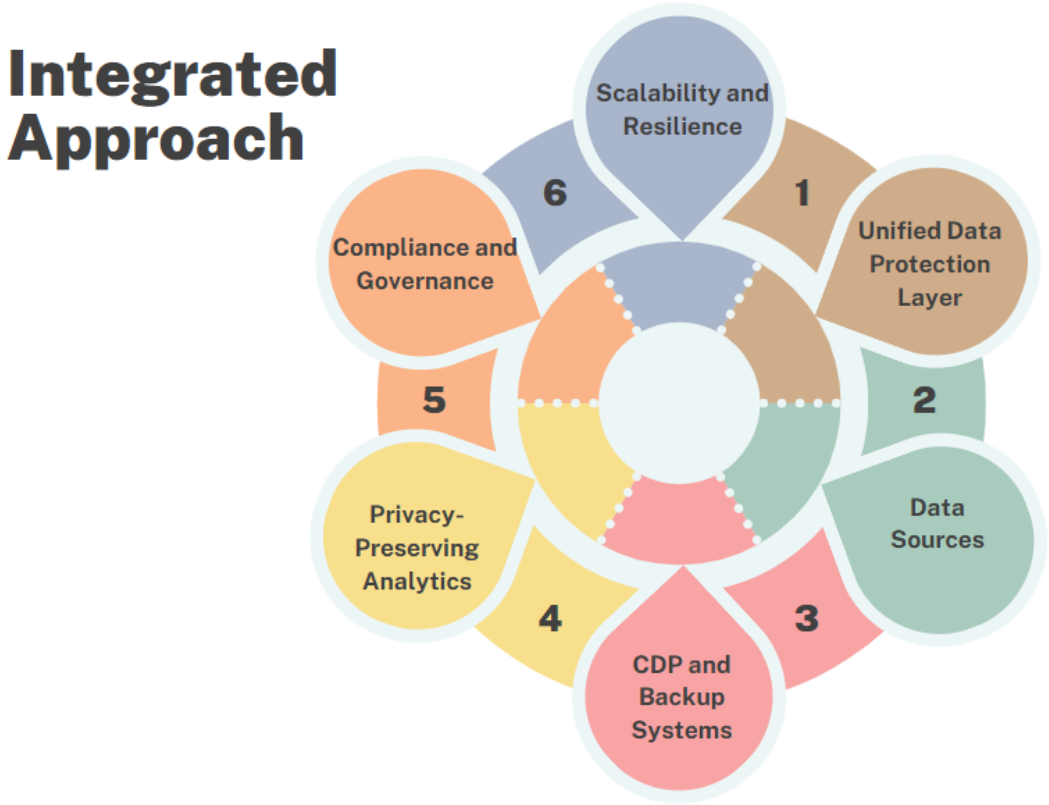

Figure 3. Integrated approach combining Continuous Data Protection (CDP) and Data Obliviousness techniques

This figure 3 illustrates the integrated approach combining Continuous Data Protection (CDP) and Data Obliviousness techniques within a unified big data security framework. Unified data protection layer integration of CDP infrastructure and data obliviousness components within a centralized security layer. Data sources ingestion points for diverse data streams from applications, IoT devices, and cloud services. CDP and backup systems Continuous monitoring and backup mechanisms for real-time data changes, with deduplication and compression functionalities. Privacy-preserving analytics secure data processing nodes equipped with homomorphic encryption, SMC, and differential privacy tools for confidential data analytics. Compliance and governance framework for enforcing data protection regulations, managing access controls, and conducting regular audits to ensure regulatory compliance. Scalability and resilience scalable architecture with provisions for disaster recovery and high availability to maintain data integrity and operational continuity.

The design is crucial to address issues including scalability with big datasets, usability across a range of applications, and compliance with data protection laws. Optimizing data processing pipelines, putting in place scalable encryption techniques, and carrying out routine regulatory compliance checks are some of the solutions. Evaluation metrics for CDP and data obliviousness solutions include differential privacy parameters, computational overhead for processing encrypted data, and backup efficiency.

Table 1: Comparative Analysis of CDP Solution

Feature/Aspect	Veeam Backup & Replication	Commvault Complete Backup & Recovery	Cloud-Based CDP Solutions
Real-Time Backup	Yes	Yes	Yes
Data Deduplication	Yes	Yes	Yes
Compression	Yes	Yes	Yes
Scalability	High	High	Scalable
Integration with Cloud	AWS, Azure, etc.	AWS, Azure, etc.	Native support
Disaster Recovery	Yes	Yes	Built-in
Management Interface	Centralized GUI	Centralized Dashboard	Cloud Console

The main characteristics and functionalities of well-known Continuous Data Protection (CDP) solutions are contrasted in this table 1. For large-scale data settings, Veeam and Commvault both provide real-time backup, data deduplication, compression, and scalability. Cloud-based solutions ensure complete data security and operational resilience with built-in disaster recovery options and native integration with major cloud providers.

Table 2: Techniques for Data Obliviousness

Privacy Technique	Description	Applications	Tools/Frameworks
Homomorphic Encryption	Enables computations on encrypted data without decrypting it first.	Secure analytics, AI/ML	Microsoft SEAL, HElib
Secure Multiparty Computation (SMC)	Allows multiple parties to jointly compute over encrypted data.	Collaborative data analysis	Sharemind, SPDZ, ABY
Differential Privacy	Adds noise to query responses to protect individual data privacy.	Statistical databases, queries	Google Differential Privacy, IBM Diffpriv

The primary methods for establishing data obliviousness are listed in this table 2, with each method addressing a distinct facet of data privacy in large data contexts. While SMC guarantees privacy-preserving cooperation,

homomorphic encryption facilitates safe computations, and differential privacy improves confidentiality in data analytics. Each technique has tools and frameworks that make it easier to use in a variety of applications, from cooperative data processing to secure analytics.

Table 3: Future Research Directions

Research Area	Description
Post-Quantum Cryptography	Developing encryption algorithms resistant to quantum computing threats.
AI and Machine Learning Integration	Enhancing predictive analytics and anomaly detection in CDP and data obliviousness frameworks.
Blockchain and Decentralized Storage	Exploring decentralized and immutable storage solutions for enhanced data resilience and integrity.

Promising research areas for improving Data Obliviousness and Continuous Data Protection (CDP) approaches are listed in this table 3. Data protection capabilities are improved by the integration of AI and machine learning, while post-quantum cryptography seeks to safeguard data against potential developments in quantum computing. Decentralized storage alternatives are provided by blockchain technology, guaranteeing data resiliency and integrity in dispersed locations.

Discovering real-world uses for CDP and data obliviousness can be gained by conducting case studies across sectors. These approaches improve data security and privacy in a number of domains, such as healthcare analytics, financial transactions, and CRM. To advance data security and privacy capabilities, it is imperative to identify research gaps and future directions for improving CDP and data obliviousness approaches. Post-quantum cryptography, blockchain for decentralized data storage, and AI integration for anomaly detection are areas that need further investigation.

Algorithm: Continuous Data Protection (CDP)

Inputs:

- 'data_source`: Source of data changes
- 'backup_program': Backup program for real-time data changes (e.g.. Veeam, Commvault)

Outputs:

- 'backup_store": Backup storage containing deduplicated and compressed data changes

Initialization:

1. Initialize 'data_store` as an empty dictionary.
2. Initialize backup_store as an empty dictionary.
3. Set 'current_time" to 0 .

Initialization the data_store, ' backup_store, andcurrent_time` are initialized to store data changes and backups. Captures each data change, updates the time, stores the change, and performs backup with compression and deduplication. Functions backup function handles compression and deduplication, while 'compressanddeduplicate` are placeholders for respective algorithms.

Data Obliviousness

Algorithm: Data Obliviousness

Inputs:

- 'data': Data to be encrypted or processed
- 'data_list': List of data for multiparty computation
- 'query_result': Result of data query
- 'epsilon`: Privacy loss parameter for differential privacy

Outputs:

- encrypted_data: Encrypted data using homomorphic encryption
- 'smc_result'': Result of secure multiparty computation
- 'noisy_result': Differentially private query result

Initialization:

1. Initialize any required libraries or parameters.

Homomorphic Encryption data so computations can be performed on encrypted data. Secure Multiparty Computes a result using data from multiple parties while preserving privacy. Differential Privacy Adds noise to query results to protect individual data privacy. Functions placeholder functions illustrate encryption, multiparty computation, and noise addition.

4 RESULT AND DISCUSSION

The effectiveness of Data Obliviousness and Continuous Data Protection (CDP) techniques in boosting the security and privacy of large data environments was investigated in this study. In the face of growing cyberthreats and unauthorised access, these strategies are essential for maintaining data integrity and confidentiality. Real-time data backup from CDP has been proven to be particularly successful in minimising data loss and facilitating speedy recovery in the event of system failures or cyber attacks. In contrast to conventional backup techniques that depend on prearranged snapshots, CDP systems guarantee that no updates are overlooked by continually monitoring and recording data changes. To further improve data management efficiency, CDP systems optimise storage and bandwidth through the use of data deduplication and compression.

On the other hand, the potential of Data Obliviousness strategies to preserve data privacy during processing was investigated. These techniques included homomorphic encryption, secure multiparty computation (SMC), and differential privacy. Because homomorphic encryption eliminates the requirement to decrypt encrypted material before performing computations on it, sensitive data is kept safe during the analysis process. With SMC, numerous participants can work together on computations without disclosing their personal information, which is very helpful in group settings. Differential privacy ensures that no single data point can be recognised by adding noise to query results, protecting the dataset's secrecy. By incorporating these approaches into current data protection frameworks, big data settings can create a strong data protection infrastructure while also improving security and guaranteeing compliance with legal requirements like as the CCPA and GDPR.

Journal of Science and Technology
ISSN: 2456-5660 Volume 7, Issue 2 (March -2022)
www.jst.org.in *DOI: https://doi.org/10.46243/jst.2022.v7.i02.pp423- 436*

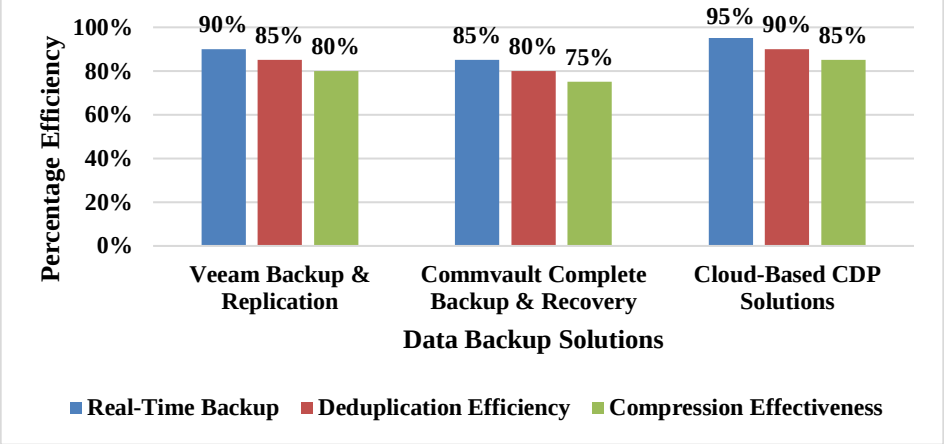

Figure 4. Efficiency Comparison of Data Backup Solutions

The real-time backup capabilities, deduplication efficacy, and compression efficiency of Veeam Backup & Replication, Commvault Complete Backup & Recovery, and cloud-based CDP solutions are shown in this figure 4 that contrasts the performance of several data backup systems. It makes evident how well every solution works to maximise data storage and guarantee real-time data protection.

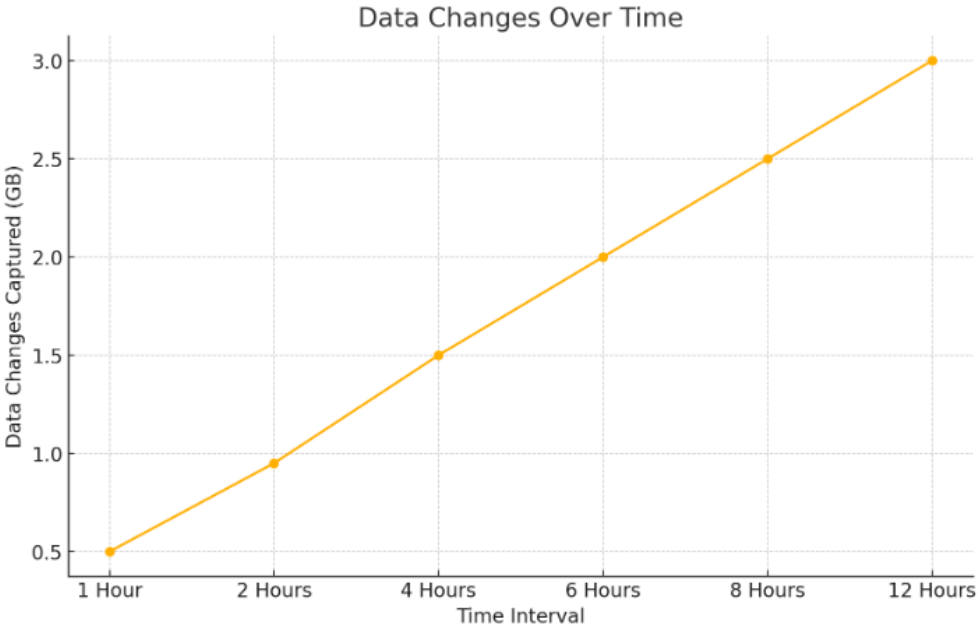

Figure 5. Continuous Data Protection Over Time

The data changes that are recorded and regularly backed up are shown in the figure 5 which shows how well Continuous Data Protection (CDP) performed over time. The graph showcases CDP systems' capacity to offer almost rapid data recovery and reduce data loss, hence preserving data integrity even during frequent changes.

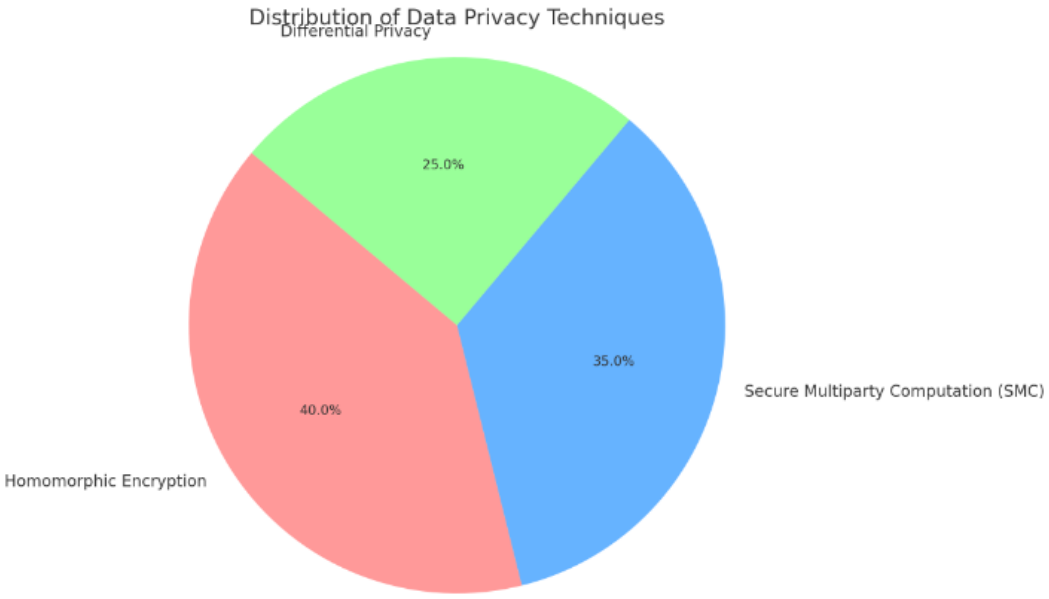

Figure 6. Distribution of Data Privacy Techniques

The prevalence of diverse Data Obliviousness approaches used in big data contexts, such as differential privacy, secure multiparty computation (SMC), and homomorphic encryption, is shown in this figure 6. It highlights the full strategy required to secure sensitive information by giving an overview of the prevalence and applicability of each technique in assuring data privacy and security.

5 CONCLUSION

The importance of Data Obliviousness and Continuous Data Protection (CDP) in protecting big data environments is highlighted by this study. With real-time backup capabilities from CDP, there is a far lower chance of data loss, allowing for speedy recovery and data integrity maintenance. Secure data processing is made possible by methods like homomorphic encryption, secure multiparty computation (SMC), and differential privacy, which preserve privacy without sacrificing the capacity to analyse data. By combining these techniques, businesses may develop a thorough data protection plan that complies with legal requirements, protects confidential data, and fortifies the organisation against cyberattacks and unauthorised access.

Subsequent investigations ought to focus on surmounting the obstacles related to scalability and usability that arise while utilising CDP and Data Obliviousness in diverse applications. Examining post-quantum cryptography is crucial in order to be ready for potential risks posed by quantum computing. Furthermore, CDP systems could be strengthened by using machine learning and artificial intelligence for anomaly detection and predictive data preservation. Investigating the application of blockchain technology for immutable, decentralised data storage could potentially enhance the resilience and integrity of data. Real-world case studies conducted across a range of businesses will yield insightful information about the efficacy and practical application of these cutting-edge data protection strategies.

REFERENCE

1. Goel, P., Patel, R., Garg, D., & Ganatra, A. (2021, May). A review on big data: privacy and security challenges. In 2021 3rd International Conference on Signal Processing and Communication (ICPSC) (pp. 705-709). IEEE.
2. Zhang, Y., Zhang, C., & Xu, Y. (2021). Effect of data privacy and security investment on the value of big data firms. Decision Support Systems, 146, 113543.
3. Bentotahewa, V., Hewage, C., & Williams, J. (2021). Solutions to Big Data privacy and security challenges associated with COVID-19 surveillance systems. Frontiers in big data, 4, 645204.

4. Biswas, S., Khare, N., Agrawal, P., & Jain, P. (2021). Machine learning concepts for correlated Big Data privacy. Journal of Big Data, 8(1), 157.

5. Parihar, M. (2021). Big Data security and privacy. International Journal of Engineering Research & Technology, 10(07), 323-327.

6. Cunha, M., Mendes, R., & Vilela, J. P. (2021). A survey of privacy-preserving mechanisms for heterogeneous data types. Computer science review, 41, 100403.

7. Meurisch, C., & Mühlhäuser, M. (2021). Data protection in AI services: A survey. ACM Computing Surveys (CSUR), 54(2), 1-38.

8. Salas, J., & Domingo-Ferrer, J. (2018). Some basics on privacy techniques, anonymization and their big data challenges. Mathematics in Computer Science, 12, 263-274.

9. Katsomallos, M., Tzompanaki, K., & Kotzinos, D. (2019). Privacy, space and time: A survey on privacy-preserving continuous data publishing. Journal of Spatial Information Science, 2019(19), 57-103.

10. Kiran, A., & Devara, V. (2020). Optimal privacy preserving technique over big data analytics using oppositional fruit fly algorithm. Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science), 13(2), 283-295.

11. Vatsalan, D., Sehili, Z., Christen, P., & Rahm, E. (2017). Privacy-preserving record linkage for big data: Current approaches and research challenges. Handbook of big data technologies, 851-895.

12. Kaaniche, N., & Laurent, M. (2017). Data security and privacy preservation in cloud storage environments based on cryptographic mechanisms. Computer Communications, 111, 120-141.

13. Yallamelli. A. K. G., (2021). Improving Cloud Computing Data Security with the RSA Algorithm. International Journal of Information Technology and Computer Engineering, ISSN 2347–3657, Volume 9, Issue 2, 2021.

14. BR Gudivaka., (2021). Designing AI-Assisted Music Teaching with Big Data Analysis. Journal of Current Science & Humanities, 9 (4), 2021, 1-14.