SAFEGUARDING DATA INTEGRITY: A COMPREHENSIVE EXPLORATION OF DATABASE BACKUP AND RECOVERY USING *TAMPER-RESISTANT PROPERTIES OF BLOCKCHAIN*

BUOYE PETER* & **AKINBOLA SHERIFAT**
Department of Computer Science, The Federal Polytechnic, Ilaro, Nigeria
*Corresponding author:adewuyi.buoye@federalpolyilaro.edu.ng

ABSTRACT:

Data integrity is a critical concern in database management systems, particularly in ensuring the security and reliability of stored information. Traditional methods of database backup and recovery face challenges such as centralized control, vulnerability to tampering, and single points of failure. This paper presents a comprehensive exploration of leveraging blockchain technology to enhance database backup and recovery processes, thereby safeguarding data integrity. Drawing on the decentralized and tamper-resistant properties of blockchain, this study investigates how blockchain can be integrated into database backup and recovery systems to mitigate risks associated with data manipulation, unauthorized access, and data loss. The proposed approach utilizes cryptographic hashing, decentralized storage, consensus mechanisms, and smart contracts to create secure, transparent, and auditable backups of database records. Through a detailed analysis of existing literature, case studies, and technical implementations, this paper evaluates the effectiveness and feasibility of blockchain-based solutions for database backup and recovery. It discusses the benefits of blockchain technology in ensuring data immutability, integrity verification, and fault tolerance while addressing potential challenges and limitations.

Keywords: Blockchain, Consensu-mechanism. Cryptography, Decentralization, Hashing, Interoperability

1.0 INTRODUCTION

In the ever-evolving landscape of information technology, data serves as the lifeblood of organizations, fueling critical decision-making processes and ensuring seamless operations. The integrity and availability of this data, housed within databases, are paramount to an organization's success. However, the vulnerability of digital assets to various threats necessitates a robust strategy for database backup and recovery.

Ketan & Gurpreet (2022) described a data backup as the practice of copying data from the first to the second location. A database backup is a systematic process of creating and storing duplicate copies of a database or its components. This redundancy serves as a safety measure, providing a means to recover data in the event of accidental deletion, system failures, or catastrophic events. The primary goal of database backup is to safeguard the integrity, availability, and consistency of critical data.

A database backup involves creating a copy of the entire or a subset of a database, capturing its current state, and storing it in a secure location. This process serves as a safeguard against data loss, providing a mechanism to restore the database to a previous state in case of corruption, accidental deletions, or system failures. Database recovery, on the other hand, refers to the process of restoring the database from a backup to its consistent and usable state

The importance of database backup and recovery extends beyond mere data preservation; it is a cornerstone of business continuity and disaster recovery planning. Organizations invest in backup strategies to minimize downtime, maintain data consistency, and ensure compliance with regulatory requirements. As databases grow in complexity and scale, implementing efficient and reliable backup and recovery mechanisms becomes paramount for sustaining operational resilience.

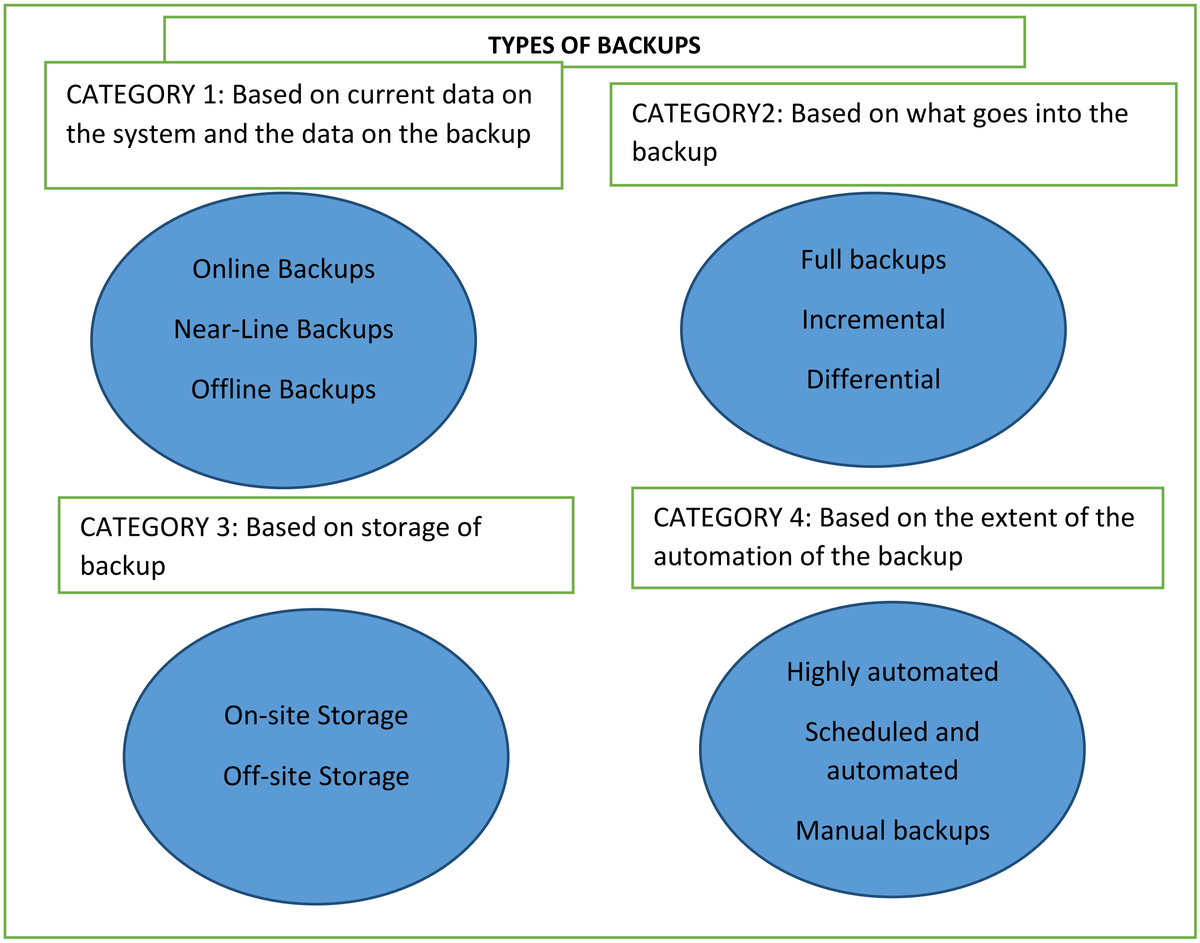
Various backup methods, such as full, incremental, and differential backups, offer flexibility in tailoring strategies to specific organizational needs. According to Lorent (2022), a type of backup defines how data is copied from source to destination and lays the grounds of a data repository model. Additionally, the choice between online and offline backups influences the impact on system performance during backup operations. Implementing a well-defined backup schedule and retention policy further enhances the overall data management strategy.

According to Brooke (2024), Blockchain is an immutable digital ledger that enables secure transactions across a peer-to-peer network. She described blockchain technology as a decentralized and distributed ledger that stores the record of ownership of digital assets.

Recovery, on the other hand, pertains to the restoration of a database to a previous state after a data loss event. Whether due to system failures or accidental deletion, the recovery process is critical for minimizing downtime and ensuring business continuity. Recovery mechanisms typically involve restoring data from the latest backup and applying transaction logs or incremental backups to bring the database up to the desired point in time.

The principles of database backup revolve around creating copies of the data within a database to ensure data integrity, availability, and recovery in the event of data loss or system failures. These principles (such as backup types, backup frequency, transaction log etc.) encompass various considerations, strategies, and practices to effectively implement and manage database backup processes.

There are four major types of backup, grouped into categories, based on certain criteria. These types are diagrammatically demonstrated below.


TYPES OF BACKUPS
CATEGORY 1: Based on current data on the system and the data on the backup
Online Backups
Near-Line Backups
Offline Backups
CATEGORY2: Based on what goes into the backup
Full backups
Incremental
Differential
CATEGORY 3: Based on storage of backup
On-site Storage
Off-site Storage
CATEGORY 4: Based on the extent of the automation of the backup
Highly automated
Scheduled and automated
Manual backups

The modern digital landscape is characterized by an unprecedented reliance on databases, which store and manage vast volumes of critical information. Whether it be financial transactions, customer records, or intellectual property, the value of this data cannot be overstated. However, with the proliferation of cyber threats, system failures, and human errors, the need for a robust database backup and recovery mechanism becomes paramount

A comprehensive understanding of database backup and recovery involves exploring the intricacies of safeguarding data at multiple levels. This encompasses not only the technical aspects of implementing efficient backup strategies but also the strategic considerations in aligning these processes with organizational goals.

2.0 BACKGROUND AND RELATED WORK

According to Yashodha ,S. & Rajashekarappa. (2016) highlighted the importance of efficient backup mechanisms in database systems, emphasizing comprehensive coverage and minimal system performance impact to meet the dynamic demands of modern organizations. Kruti and Kavita (2012) discussed the necessity of cloud computing for database backup and recovery, identifying various contemporary strategies such as HSDRT, PCS, ERGOT, Linux Box, Cold and Hot Backup Technique, SBBR, and REN. They concluded that PCS is somewhat dependable due to its cost-efficiency and privacy maintenance but struggles with implementation complexity. Zhang, Zhou, Li, Liu, Xie, Cheng, & Xing (202) explored incremental and differential backup methods, analyzing the trade-offs between data recovery speed and storage efficiency, and provided insights on tailoring backup approaches. Praveen, Ambika, & Mahantesh, (2017). investigated the implications of cloud-based solutions on database backup and recovery, emphasizing the need for organizations to adapt their strategies to the features and constraints of cloud environments.

Malatesh (2018) offered options for storage repository models, such as native file systems and Hadoop Distributed File System (HDFS), concluding that HDFS is cost-effective and can be built with low-cost commodity hardware. Johnson & Wang (2019) delved into the role of training and user awareness in mitigating data loss risks, emphasizing the need for a holistic approach to data protection that integrates technology and human factors. Wang & Gupta (2020) explored the integration of artificial intelligence (AI) and machine learning (ML) techniques to enhance the efficiency and reliability of backup processes. Karina, Amol, Damini, Rupali, & Pachghare (2021) analyzed Cassandra Database Recovery

Mechanisms, concluding that multiple servers yield better recovery outcomes than a single server and highlighted the benefits of trigger-based backup mechanisms for achieving lower RPO and RTO values.

Li & Patel (2021) discussed securing backup data, focusing on the vulnerabilities during storage and transmission. Khan, Laghari, Gadekallu, Shaikh, Javed, Rashid,& Mikhaylov (2022). provided insights into the role of geographic redundancy and distributed architectures in disaster recovery, emphasizing data availability during natural disasters or large-scale outages. According to Shaanxi & Xi'an (2023) introduced blockchain technology and an electronic data secure storage model using the Ethereum virtual machine and Practical Byzantine Fault Tolerance (PBFT) consensus algorithm. Their model effectively detects tampering and restores data using blockchain backups.

Blockchain is a decentralized and distributed digital ledger technology that records transactions across a network of computers in a secure and tamper-resistant manner. It gained prominence as the underlying technology for cryptocurrencies like Bitcoin, but its applications extend far beyond digital currencies.

Information is kept in uniform-sized blocks on a blockchain. To ensure cryptographic security, each block includes the hashed data from the one before it. The one-way hash function used for hashing is SHA256. The data and digital signature from the most recent block, as well as the hashes of earlier blocks dating all the way back to the "genesis block," the first block ever created in the blockchain, are included in this hashed data. A hash function is applied to the data, and the result is an address for the previous block. One example of a Merkle Tree, which is utilized as an effective means of data verification, is a blockchain data structure.

3.0 BLOCKCHAIN TECHNOLOGY

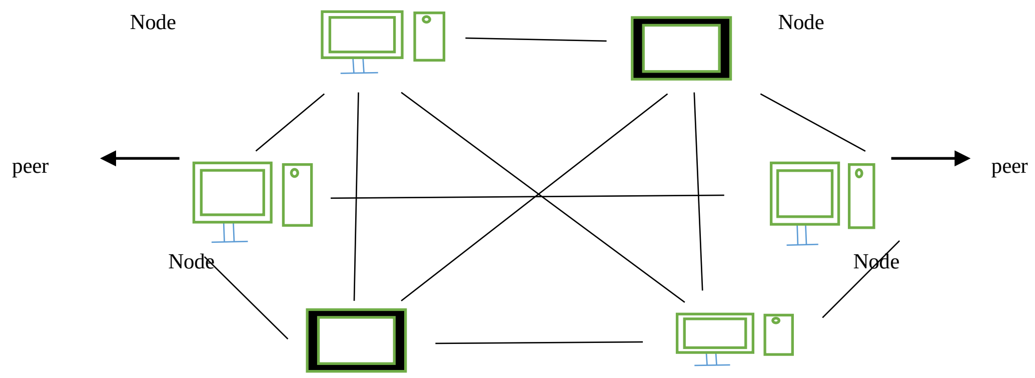


Fig 1: **P2P Network: Author**

Every node, or participant, in a peer-to-peer network using blockchain possesses a copy of the complete ledger. Depending on the blockchain protocol, a consensus process such as Proof of Work, Proof of Stake, or another technique is used to validate a newly created transaction when it is broadcast to the network. In the case of Proof of Work, adding a new block to the chain requires nodes to solve difficult mathematical puzzles, and in the case of Proof of Stake, staking a particular quantity of cryptocurrency. The new block is appended to the chain and all nodes update their copy of the ledger upon reaching a consensus. Immutability is one of blockchain's primary characteristics. The integrity of the whole transaction history is ensured by the fact that once a block is added to the chain, changing its contents would necessitate modifying all subsequent blocks, rendering the change computationally impractical.
Blockchain technology is not just for currency. For instance, smart contracts are self-executing contracts containing coded terms that, when certain requirements are satisfied, automatically carry out and enforce agreements. This feature affects several industries, including supply chain management, healthcare, and finance.

Enhancing data backup and recovery through the tamper-resistant nature of blockchain.

The capacity of blockchain technology to securely preserve data in a way that makes it very difficult to change or manipulate historical transactions or records is known as its "tamper-resistant nature."

Dispersion

In the context of blockchain, decentralization is the division of power and jurisdiction across several computers, or nodes, as opposed to depending on a single central authority. It is the cornerstone of blockchain technology and the secret to its dependability, security, and reliability.

Unlike traditional centralized systems where a central authority controls and manages the entire system, blockchain operates on a decentralized network. In a blockchain network, there isn't a single point of failure, ownership, or control. To provide transparency and prevent any one party from dominating the entire system, each participant, or node, keeps a copy of the full ledger comprising all transaction records.
Blockchain functions as a peer-to-peer network in which nodes exchange information directly to verify and spread transactions. By doing away with the need for middlemen and enabling direct communication between participants, this boosts productivity and lowers expenses.
Consensus techniques are employed by decentralized blockchain networks to reach a consensus among nodes regarding the legitimacy of transactions and the ledger's current state. Consensus techniques improve the security and dependability of the network by preventing manipulation by a single actor.
Decentralization enhances the resilience and fault tolerance of

blockchain networks. Since there is no central point of failure, the network remains operational even if some nodes fail or are compromised. This makes blockchain resistant to censorship, hacking, and other forms of attacks.

Decentralization promotes trust and transparency by allowing participants to verify transactions independently without relying on a central authority. Since every transaction is recorded on the blockchain and visible to all participants, the integrity of the data can be easily verified.

Cryptographic hashing

Cryptographic hashing is a fundamental concept in cryptography, transforming input data into fixed-size hash values using mathematical algorithms. Key principles include deterministic output, fixed size, pre-image resistance, and collision resistance. Common hash functions include MD5, SHA-1, SHA-256, SHA-384, and SHA-512. Applications range from data integrity verification and password storage to digital signatures and blockchain technology. Security considerations involve algorithm selection, salt and pepper techniques, and key stretching. Despite its benefits, cryptographic hashing has limitations and vulnerabilities, especially in weaker hash functions like MD5 and SHA-1. Overall, cryptographic hashing is crucial for ensuring security and integrity in various cryptographic operations and information systems.

Cryptographic hash functions are mathematical procedures that take an input (or message) and output a fixed-length string of characters. These algorithms are what create hash values. These routines compress the input data into an output of a constant length by processing it using particular algorithms. They guarantee predictable output, which means that the same input consistently yields the same result. Since cryptographic hash functions are one-way, deriving the original input from the hash value cannot be done computationally.
Good hash functions produce uniformly distributed hash values, preventing collisions and ensuring security. Examples include SHA-256, SHA-1, and MD5, used in various cryptographic applications like data integrity verification and password storage.

Let us consider the strings 'my data' and 'my data' using the SHA-256 cryptographic hash function.

	'my data'	'mydata'
	f5d2e22521a862dc579a8555db3b1c3e29b83d77726c6edc0d98cf59cd05cf30	59f34e7b19dc295ed39e0f94a1cfab60ff7aa3abcc826e0bfa5459af1086ae16

Table 1:SHA-256 hash value

From the hash function table above, “my data” and “mydata” shows different hash functions. Although “mydata” and “my data” looks very the same, the space in between “my” and “data” makes the algorithm to generate a different hash value. This is the bases on which data integrity is achieved using hash function.

Consensus mechanisms

Blockchain technology relies heavily on consensus mechanisms, which allow decentralized networks to reach consensus over the legitimacy of transactions and preserve the distributed ledger's integrity. These procedures ensure that no single entity controls the network by distributing decision-making authority among network nodes. Decentralization, security, scalability, and fault tolerance are important features.
Proof of Work (PoW), Proof of Stake (PoS), Delegated Proof of Stake (DPoS), Proof of Authority (PoA), and Proof of Burn (PoB) are examples of common consensus procedures. Each mechanism has its own approach to transaction validation and block creation, balancing factors like energy consumption, decentralization, and security.

Emerging consensus mechanisms like Proof of Space (PoSpace), Proof of History (PoH), and Proof of Replication (PoRep) offer innovative solutions to scalability and performance challenges. However, considerations such as energy consumption, centralization risks, and scalability remain important factors in selecting the appropriate consensus mechanism for a blockchain network.

For safeguarding data integrity in a public blockchain context where decentralization and security are paramount, Proof of Work (PoW) or Proof of Stake (PoS) might be preferred due to their established security models. However, for private or enterprise-focused applications where trust and identity are established, Proof of Authority (PoA) could be a suitable choice. However, the selection of the best consensus mechanism depends on the specific requirements, trade-offs, and priorities of the blockchain application. It's essential to carefully evaluate each option and consider factors such as security, decentralization, energy efficiency, scalability, and governance considerations before making a decision.

Immutability

In the context of blockchain technology, immutability refers to the feature wherein data entered into the ledger is virtually impossible to change or remove once verified. Immutability, which is attained through decentralization, consensus processes, and cryptographic hashing, guarantees the security and integrity of blockchain data. Because every block in the chain includes a hash of the data from the previous block, manipulation is very difficult. Immutability reduces the danger of fraud and promotes transparent record-keeping, which finds widespread uses across industries. On the other hand, inaccurate data correction presents difficulties. Overall, immutability is a foundational feature of blockchain technology, offering robustness and trustworthiness in data management and transactions.

Interoperability

Interoperability between blockchain and databases represents a pivotal advancement in data management, enabling organizations to leverage the strengths of both technologies in tandem. By seamlessly integrating blockchain with traditional databases, interoperability bridges the gap between decentralized, tamper-resistant ledger systems and established centralized data storage solutions. One of the key facets of interoperability lies in its ability to facilitate data exchange and synchronization. Through standardized protocols and interfaces, blockchain platforms can communicate with existing databases, enabling the seamless transfer of data between disparate systems. This integration allows organizations to capitalize on the security and transparency of blockchain while maintaining compatibility with their existing data infrastructure.

Moreover, interoperability opens the door to a wide array of use cases across industries. In supply chain management, for instance, blockchain can be utilized to track and trace products throughout the supply chain, while traditional databases store detailed product information and transaction records. Interoperability enables real-time data sharing between blockchain-based supply chain networks and legacy database systems, enhancing transparency, traceability, and efficiency throughout the supply chain ecosystem. Additionally, interoperability fosters innovation by enabling the integration of blockchain capabilities, such as smart contracts and decentralized applications (DApps), with traditional database-driven applications. This convergence enables the development of hybrid solutions that harness the security and immutability of blockchain alongside the scalability and flexibility of traditional databases, paving the way for transformative applications across industries such as finance, healthcare, logistics, and beyond.

Furthermore, interoperability holds the promise of simplifying cross-chain communication, allowing assets and data to seamlessly traverse multiple blockchain networks. This interoperability not only enhances liquidity and accessibility within blockchain ecosystems but also fosters collaboration and synergy between diverse blockchain platforms, driving greater innovation and value creation in the decentralized landscape.

4.0 CONCLUSION

In conclusion, the tamper-resistant nature of blockchain serves as a formidable safeguard for data integrity in various applications. By leveraging cryptographic hashing, decentralization, consensus mechanisms, immutable data structures, and digital signatures,

The cryptographic hashing algorithms used in blockchain create unique digital fingerprints for each block of data, ensuring that any attempt to tamper with the data would result in detectable change to the hash values. Decentralization ensures that no single entity controls the network, making it challenging for attackers to compromise the integrity of the data without consensus from the majority of the network.

Consensus mechanisms play a crucial role in maintaining data integrity by ensuring that all participants in the network agree on the validity of transactions and the state of the ledger. Immutable data structures ensure that once data is recorded and confirmed on the blockchain, it becomes practically impossible to alter or delete. Furthermore, digital signatures provide an additional layer of security by verifying the authenticity and integrity of transactions, ensuring that they cannot be tampered with during transmission.

In essence, the tamper-resistant nature of blockchain technology establishes trust and transparency in data management by providing a secure, decentralized, and immutable ledger. This makes blockchain an invaluable tool for safeguarding data integrity in a wide range of applications, including financial transactions, supply chain management, healthcare records, and more. As blockchain continues to evolve, its role in ensuring data integrity and trust in digital ecosystems is expected to grow even further.

References

Brooke, B. (2024). Understanding blockchain technology. Builtin. Retrieved March 25, 2023, from https://builtin.com/blockchain

Cheng, B., Zhang, J., Zhou, K., Li, G., Liu, Y., Xie, M., & Xing, J. (2021). CDBTune+: An efficient deep reinforcement learning-based automatic cloud database tuning system. The VLDB Journal, 30(6), 959-987.

Johnson, K., & Wang, Y. (2019). Human factors in database backup and recovery: The role of training and awareness. Human-Computer Interaction Journal, 25(3), 167-184.

Karina, B., Amol, B., Damini, S., Rupali, C., & Pachghare, V. K. (2021). Backup and recovery mechanisms of Cassandra database: A review. Journal of Digital Forensics, Security and Law, 15, Article 5.

Ketan, S., & Gurpree, K. (2022). Data backup and recovery. International Journal For Technological Research In Engineering, 9(10). ISSN (Online): 2347 – 4718.

Khan, A. A., Laghari, A. A., Gadekallu, T. R., Shaikh, Z. A., Javed, A. R., Rashid, M., ... & Mikhaylov, A. (2022). A drone-based data management and optimization using metaheuristic algorithms and blockchain smart contracts in a secure fog environment. Computers and Electrical Engineering, 102, 108234.

Kruti, S., & Kavita, R. S. (2012). Online data back-up and disaster recovery techniques in cloud computing: A review. International Journal of Engineering and Innovative Technology (IJEIT), 2(5).

Li, H., & Patel, R. (2021). Security considerations in database backup: A comprehensive review. Cybersecurity Insights, 18(4), 201-218.

Lorent. (2022). Types of backup. Backup4all. Retrieved March 25, 2023, from https://www.backup4all.com/backup-types-kb.html

Malatesh, K., & Meenakshi, S. (2018). Backup and recovery for file system and databases. Asian Journal of Engineering and Technology Innovation (AJETI), ISSN: 2347-7385.

Praveen, S. C., Ambika, S. D., & Mahantesh, N. B. (2017). Efficient and reliable data recovery technique in cloud computing. Internet of Things and Cloud Computing, 5(5-1), 13-18.

Shaanxi, P. C., & Xi'an, S. (2023). Research on secure storage of electronic data through blockchain technology. International Journal of Network Security, 25(3), 449-455.

Wang, X., & Gupta, S. (2020). Integrating AI and ML techniques in database backup processes. Journal of Advanced Technology Applications, 15(2), 88-105.

Yashodha, S., & Rajashekarappa. (2016). Efficient data backup mechanism for cloud computing. International Journal of Advanced Research in Computer and Communication Engineering, 5(7), 92-96