Future of block chain technology & big data Integration

Mrs.Sukeshini Satish Gawai[1]
[1]*Lecturer,Y B Patil Polytechnic Akurdi, Pune*

Abstract: **Block chain is a type of database, but it differs entirely from traditional databases in how it stores and manages information. Instead of storing data in rows, columns, tables and files as traditional databases do, blockchain stores data in blocks that are digitally chained together. Moreover, a blockchain is a decentralized database managed by computers belonging to a peer-to-peer network instead of a central computer like in traditional databases and it is an immutable and secure digital ledger that records transactions permanently and verifiably. Bitcoin and Ethereum are examples of such blockchains.**

Blockchain and Big Data are two most advanced technologies that have been significantly transforming various industries. Block chain & big data integration can increase security, data integrity, and privacy efficiently than the traditional systems

***Index Terms*—Block chain, Big data, Cryptographic technique, Distributed Denial-of-Service (DDoS)**.

I. FUTURE OF BLOCKCHAIN TECHNOLOGY

Worldwide organizations are using the blockchain technology to create digital currencies, it increases the transparency of data and lower the chances of potential fraud. It even allows customers to gain access to cloud-based solutions to host, build, and use blockchain apps.

Use Cases for Blockchain and Big Data Integration

1. Healthcare:

Blockchain can provide secure, decentralized storage for patient records, ensuring data integrity and privacy. Big Data analytics can then be applied to this data to derive insights into patient care, disease trends, and more.

2. Supply Chain Management:

Blockchain can track products as they move through the supply chain, ensuring the data collected about each product is accurate, transparent, and immutable. The Big Data system can analyze this data in real-time to optimize supply chain operations.

3. Finance and Banking: Blockchain can ensure secure, transparent financial transactions, while Big Data can analyze customer behavior, market trends, and risk profiles. Together, they enable better decision-making, fraud detection, and compliance.

4. Energy Management: Blockchain can securely track and manage energy transactions (such as from solar panels or wind turbines), while Big Data can be used to analyze energy consumption patterns, predict demand, and optimize energy distribution.

Example: Power Ledger uses blockchain to track renewable energy trading, while Big Data analytics optimizes the grid and predicts energy consumption.

5. IoT (Internet of Things):IoT devices generate vast amounts of data, which can be securely stored on the blockchain to ensure data integrity.
Example: In smart cities, blockchain can manage data from IoT devices like traffic sensors or environmental monitors, while Big Data helps improve urban planning, traffic flow, and resource allocation.

Blockchain and Big Data Integration Architecture

II. BLOCKCHAIN AND BIG DATA INTEGRATION ARCHITECTURE

To integrate blockchain with big data effectively, you need an architecture that balances the strengths of both technologies while addressing their limitations.

1. Data Ingestion & Collection Layer

Data from various sources (IoT sensors, social media, web logs, etc.) is ingested into the Big Data platform through tools like Apache Kafka, Apache Flume, or AWS Kinesis.

The data can be processed and analyzed in real-time or batch mode using tools like Apache Spark or Apache Flink.

2. Blockchain Integration Layer

In this layer, blockchain is used to provide tamper-proof records of the data processed in the Big Data

system. Instead of storing entire datasets, the blockchain stores metadata, transaction hashes, or pointers to the off-chain data.

Smart contracts can be deployed here to enforce business logic for data transactions and automate workflows.

3. Big Data Storage Layer

Big Data is typically stored in distributed storage systems like HDFS, NoSQL databases (e.g., MongoDB, Cassandra), or cloud-based storage like AWS S3 or Google Cloud Storage.

Large datasets that need to be queried and analyzed are kept off-chain to avoid blockchain storage constraints.

4. Data Processing & Analytics Layer

Big Data processing and analytics tools (e.g., Apache Spark, Hadoop, TensorFlow) analyze the ingested data to derive insights, predictions, and actionable business intelligence.

These tools can be integrated with blockchain to ensure that any insights derived from the data are linked to tamper-proof records for auditability and trust.

5. Visualization and Reporting Layer

Tools like Tableau, Power BI, or Apache Superset allow users to visualize insights from the Big Data processing layer.

Blockchain metadata can be used to validate and verify the insights shown on the dashboards, ensuring that the data's provenance and integrity are auditable.

III. OVERVIEW OF HOW BLOCKCHAIN CAN ENHANCE THE SECURITY OF BIG DATA.

- Blockchain offer decentralized storage solutions for Big Data, where multiple copies of data exist across different nodes in the blockchain network. This reduces the risks associated with centralized storage, such as single points of failure and data breaches.
- BlockChain ensures that once data is written, it cannot be tampered with. Each block in the blockchain contains a cryptographic hash of the previous block, forming an immutable chain. This makes it perfect for securing Big Data, where integrity and traceability are critical. For instance, in financial transactions or healthcare records, Big Data systems benefit from blockchain's ability to ensure that records are unalterable, providing transparent data access and audit
- Blockchain leverages advanced cryptographic techniques, such as public-private key pairs and digital signatures, to ensure the privacy and confidentiality of data stored and shared in Big Data systems.
- Users can retain control over their own data by giving permissions for specific data access without exposing everything to third parties. This is particularly valuable in sectors such as healthcare or finance, where data privacy regulations are stringent
- Distributed Denial-of-Service (DDoS) attacks and other cyber threats pose a significant risk to centralized Big Data systems. By decentralizing data management and storage, blockchain makes it much harder for attackers to target and disrupt the system.
- Fraud Prevention: The current big data solutions mainly depend on the analysis of patterns in the stored historical data to detect fraudulent transactions.so its difficult for big data to solve the problem of fraudulent transactions in the financial sector. When storage of the big data is in block chain ,blockchain enables the financial institutions to work & monitor each transaction in real time, so its allowing them to assess the potentially fraudulent transactions quickly. As a result, the integration of blockchain in big data can help the financial institutions to prevent the frauds to protect their customers.
- Real-Time Data Analytics: Blockchain stores every transaction, so the real-time analytics of big data is possible. The banking sector and financial sector can settle the cross-border transactions including large amounts in near real-time as the blockchain integrated big data analytics enables the financial institutes to settle the transactions quickly. Also, banks can track the changes in the data in real time, thus enabling them to make decisions like blocking of the transactions in real time
- Enhancement of the Quality of Big Data: data integration as different sources follow different formats in data collection to maintain the quality of data is difficult, By using blockchain for data storage, the quality of the data can be improved as it is structured and complete. Hence, data scientists can work on the quality data to come up more accurate predictions in real time.
- Streamlining the Data Data Access: The use of blockchain simplify big data analytics by online streamlining the data access. Indeed, by involving

multiple departments in an organization in a common blockchain, authorized users can get access to the secure, trusted data without having to go through several checks.

V. TRADITIONAL BIG DATA, CHALLENGES AND BLOCKCHAIN BASED SOLUTIONS

- File storage system
  Challenges faced by Big data:
  Unauthorized access to the electronic file system. Privacy, security and redundancy problems.
  Solutions provided by Blockchain: Blockchain integrated with IPFS provides the solution by implementing decentralized platforms to solve file redundancy problems and provides security to the file storage system. Hash value of data is stored in blockchain to provide authenticity to the users and an attribute based encryption method is applied before data storage in cloud.
- Data collection
  Challenges faced by Big data:
  Data collection is exposed to various malicious attacks and threats
  Solutions provided by Blockchain:
  Blockchain provides energy efficient data collection and secure data sharing environment using Ethereum.

- Data privacy preservation.
  Challenges faced by Big data:
  User privacy is an issue in digital scenarios in big data era. Services provided by third parties are exposed to security breaches and data misuse.

Solutions provided by Blockchain:

Blockchain provides immutable, verifiable and decentralized ledger to record the transactions in digital scenarios.It provides facilities to the user to control their personal data. Crypto-privacy methods are applied to solve privacy preserving problems.

V. CONCLUSION

The concept of blockchain aid in the management of enormous amounts of data and efficiently secure very large quantities of data, which can be highly beneficial for government organization.

Centralized servers having a high risk of data hacking, loss, or human errors. Cloud storage can be made more secure and robust against hacking with the implementation of Blockchain technology, just like its application in cybersecurity.

The use of blockchain can reduce time delays and human errors and monitor employment, costs, and releases at each step of the supply chain. Through traceability, Blockchain can also ensure the fair-trade status and legitimacy of products.

Digital advertising faced challenges including bot traffic, lack of transparency, domain fraud, inefficient payment models, etc, Blockchain has been found to resolve such issues in the supply chain. Advertisement-related transactions can be better handle with by employing this technology

Blockchain technology was firstly developed for cryptocurrencies like Bitcoin, but its applications have extended to various industries, including finance, supply chain, healthcare, and more. Its decentralized and secure nature makes it attractive for scenarios where trust, transparency, and immutability of records are very important. Blockchain technology will soon attract all business sectors and worldwide organizations to invest more in it.

REFERENCES

[1] Fadi Muheidat et al. / Procedia Computer Science 198 (2022) 15–22.

[2] K. Abbas et al., “Convergence of Blockchain and IoT for Secure Transportation Systems in Smart Cities,” Security and Communication Networks, vol. 2021, 2021

[3] Sukeshini S Gawai, Dr. L . K. Vishwamitra , An in-depth evaluation of Hybrid Access Mechanisms for Real Time Big Data Environments from an empirical perspective,‖ Frontiers in Health Informatics Vol. 13 No. 3 (2024) (Published: 10-09-2024)

[4] S. M. Tadaka, “Applications of Blockchain in Healthcare, Industry 4, and Cyber-Physical Systems,” in 2020 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS), 2020, pp. 1–8.

[5] G. Saldamli and A. Razavi, “Surveillance Missions Deployment on the Edge by Combining Swarm Robotics and Blockchain,” in 2020 Fourth International Conference on Multimedia Computing, Networking and Applications (MCNA), 2020, pp. 106–112. [

[6] S. K. Sharma, B. Bhushan, A. Khamparia, P. N. Astya, and N. C. Debnath, Blockchain Technology for Data Privacy Management. CRC Press, 2021

[7] M. Sun and J. Zhang, Research on the application of block chain big data platform in the construction of new smart city for low carbon emission and green environment," Computer Communications, vol. 149, pp. 332–342, 2020

[8] H. Yu, Z. Yang, and R. O. Sinnott, "Decentralized big data auditing for smart city environments leveraging blockchain technology," IEEE Access, vol. 7, pp. 6288–6296, 2018

[9] K. Rabah, "Convergence of AI, IoT, big data and blockchain: a review," The lake institute Journal, vol. 1, no. 1, pp. 1–18, 2018 .

[10] A. A. Siyal, A. Z. Junejo, M. Zawish, K. Ahmed, A. Khalil, and G. Soursou, "Applications of blockchain technology in medicine and healthcare: Challenges and future perspectives," Cryptography, vol. 3, no. 1, p. 3, 2019.

[11] Q. Liu and X. Zou, "Research on trust mechanism of cooperation innovation with big data processing based on blockchain," EURASIP Journal on Wireless Communications and Networking, vol. 2019, no. 1, pp. 1–11, 2019.

[12] A. Alammary, S. Alhazmi, M. Almasri, and S. Gillani, "Blockchain-based applications in education: A systematic review," Applied Sciences, vol. 9, no. 12, p. 2400, 2019.

[13] G. Chen, B. Xu, M. Lu, and N.-S. Chen, "Exploring blockchain technology and its potential applications for education," Smart Learning Environments, vol. 5, no. 1, pp. 1–10, 2018.

[14] A. Mikroyannidis, J. Domingue, M. Bachler, and K. Quick, "Smart blockchain badges for data science education," in 2018 IEEE Frontiers in Education Conference (FIE), 2018, pp. 1–5.

[15] N. Deepa et al., "A survey on blockchain for big data: Approaches, opportunities, and future directions," arXiv preprint arXiv:2009.00858, 2020.