An improvement in the safety of big data using blockchain technology

Archana[1*] **and Gaurav Aggarwal**[2*]

[1]Department of Computer Science, Govt. PG College for Women, Rohtak, India
[2]Department of Computer Science, Jagannath University, Jhajjar, India

*Correspondence:
Archana,
archana300030@gmail.com
Gaurav Aggarwal,
gauravaggarw@gmail.com

The development of big data in the information technology sector has made data management and analysis far more challenging. It is necessary to take into account everything, including volume, variety, speed, value, and complexity. Clustering makes the processing of vast volumes of data simpler. This is especially helpful when working with unstructured data. By the usage of cloud computing, which makes use of the internet as its delivery channel, it is possible to provide a variety of computer services, including servers, storage, databases, and networking, in addition to analytics and intelligence, at a lower cost. The main problem is the security of such big amounts of data. One way to enforce strong security is blockchain technology which is also the backbone of cryptocurrency. The distributed, incontrovertible, and publicly verifiable record of every transaction activity that may be provided by blockchain technology has the potential to revolutionize security in a big way for different industries.

Keywords: big data, security, block chain, clustering, inconsistency

Introduction

Big data is a reference to the ever-increasing amount of data that cannot be processed by traditional database techniques. Due to its huge size, unstructured as well as structured data, and complexity, standard data management technologies are unable to effectively store or process the information. Some examples of big data include the following: New York Stock Exchange, which generates around one terabyte worth of fresh trading data every single day. The estimates provided by Facebook indicate that more than 500 gigabytes of fresh data are being uploaded to the site's databases on a daily basis. The key ways in which these data are collected include users uploading photographs and videos, engaging in discussions with one another, posting comments, and so on.

Characteristics of big data

Big data can be described by the following characteristics:

a) Volume: The phrase "big data" refers to an extremely large quantity of data. When you have access to a vast amount of information, it is essential to make the most of its potential applications. The quantity of information that is required to categorize anything is one major aspect that determines whether or not it is considered "big data." The volume of data is particularly high in banking and other financial institutions, social media, and entertainment industries.

b) Variety: The second facet of big data that has to be taken into consideration is its level of diversity. The concepts of structured and unstructured data are both included under the umbrella word "variety." In

the past, the majority of applications did little more than save data in spreadsheets and databases and then retrieve it when necessary. At the moment, analytical applications make use of a broad variety of data sources with structured, semi-structured, as well as unstructured data; some examples of which include emails, photographs, videos, documents in PDF format, monitoring devices, and audio recordings. Data storage, data mining, and data analysis are all made more challenging by the presence of unstructured data.

c) Velocity: In the context of big data, the word "velocity" refers to the pace at which new data are generated. The speed at which new data are generated and processed will influence how much of the data's potential can be realized. What is known as "Big Data Velocity" is a concept that was developed as a result of the contributions made by business processes, log files, networks, social media, sensors, and mobile devices. There is a rapid flow of information that involves a significant volume of data. The velocity of data is particularly high for the banking, healthcare, and manufacturing sectors.

d) Inconsistency: It may be difficult to efficiently manage and analyze the data due to the variability that may arise in the data. This makes it a potential challenge.

Advantages of big data processing

There are several reasons why big data processing is beneficial:

1. Businesses have the option of incorporating outside information into decision-making. Due to the availability of social data gleaned from search engines and sites such as Facebook and Twitter, organizations are able to fine-tune their business plans according to user sentiment and changes in choices.

2. Improved customer service: Conventional means of gathering input from customers are being phased out in favor of big data technology. With these new platforms, the responses received from customers are studied and assessed with the use of big data and natural language processing technologies.

3. An early evaluation of any possible risks that may be associated with the product or service will lead to preventive actions for the risks.

4. The big data environment allows for the pre-processing of larger data sets before they are transferred to the warehouse, which improves the operational efficiency of the business. The combination of big data and data warehouses also assists businesses in offloading data that are needed less often.

Cluster

In a computer cluster, two or more individual computers, also known as nodes, collaborate with one another to carry out a specific task. Because of this, it is feasible to distribute huge tasks that can be done in parallel throughout the nodes that make up the cluster. Performance is boosted as a result of the fact that the combined memory and processing capabilities of each machine may benefit a variety of operations. Since each individual node in a computer cluster has to be able to interact with the other nodes in the cluster, the building of a computer cluster requires the use of an internodes network. In order to construct a cluster, the nodes must first be connected via the use of cluster software. A storage device that is shared across the nodes and/or a storage device that is local to each node are both potential solutions. It is common practice to designate at least one of the nodes as a leader node, which then operates as the entry point for the cluster. This node could be in charge of delegating responsibilities to the subordinates and, if that is required, gathering the results before relaying them to an external party. Also, the communication between nodes in a cluster has to be optimized so that latency may be reduced and bottlenecks can be avoided.

Blockchain technology

Blockchain technology is the term that describes the technical basis upon which Bitcoin currency is built. Because of this technology, it is possible to carry out any transaction in a way that is not governed by any central authority. The participation of a middleman or broker is not required in order for it to take place. As it creates records of transactions that are decentralized, unchangeable, and publicly verifiable, Blockchain technology has the potential to revolutionize a wide variety of industries.

Blockchains store the data differently from standard databases, which store the data in rows and columns that are hash-linked together, while blockchains store the data in blocks that are cryptographically connected together. Each new piece of information that is received results in the creation of a new block in the database. After a block has been filled with information, the data included inside it are connected to one another in the order that chronological events occurred.

While a blockchain may be used to store a wide range of data types, the use that has proven to be the most successful to this point is that of a ledger for the recording of financial transactions. There is not a single person or group in control of Bitcoin's use of the blockchain; rather, all users collectively hold the reins. Blockchain is used to record transactions related to Bitcoin transactions. With the use of blockchains, digital information may be saved and distributed, but it cannot be changed once it has been recorded. Because of this,

a blockchain may be used as the foundation for immutable ledgers, which are essentially recordings of transactions that are incorruptible in any way. Blockchains are often referred to as "distributed ledger technology" as a direct consequence of this fact.

Features of blockchain technology

The key points depicting features of blockchain technology are cited ahead:

a) Immutability is the state of being incapable of undergoing any kind of change or transformation. This is an essential component of blockchain that assures the long-term survival of the technology as it is a network that cannot be altered and is permanent.

b) Decentralized: As the network is decentralized, there is not a single person who is in charge of supervising the infrastructure. This means that security and privacy are not compromised. Decentralization of the network is ensured through the use of a dispersed collection of nodes that are responsible for its management. It is able to store anything of value, including Bitcoins, important documents, contracts, and other valuable digital assets. After then, with the assistance of the blockchain, you will be able to have direct control over them by making use of your private key. As a consequence of this, the general public regains both control and ownership of the assets as a result of the decentralized structure.

c) Improved security: In the blockchain, every piece of data is encrypted and hashed using a new method each time. This provides an additional layer of protection against unauthorized access. To put it another way, the information provided by the network hides the fundamental characteristics of the data. Mathematical techniques may be used to any data as an input to generate various values, but the length of the output always stays the same. A one-of-a-kind identification is assigned to each individual piece of data. In the ledger, each new block is given its own one-of-a-kind hash and also includes the hash of the block that came before it. A fresh set of hash IDs will be generated in the event that the data are altered in any way. And even that is a bit of a stretch, to say the least. The user is going to need both a private key and a public key in order to get access to the data.

d) Hashing cannot be undone in any case. Hashing is a very complicated process, and it is not feasible to modify or undo it in any way. Nobody will ever be able to take a public key and turn it into a private key.

e) Distributed ledger: With a public ledger, this information is often accessible to anybody who is a member of the general public. There is nowhere to run since everything is out in the open. There is, however, an exception to this general rule for blockchains that are either private or federated. In spite of this, a significant number of users are able to see the ledger at any one time under these circumstances. As a consequence of this, the ledger that represents the network is continually updated by everyone else who is using the system. This distributed the processing power among the several machines in order to get a more favorable outcome.

f) Consensus: The techniques used to reach consensus are essential to the operation of any blockchain. The intelligent design of this architecture is based on consensus algorithms, which serve as its foundation. Every blockchain has to have a consensus in order to function properly. The network's ability to reach a consensus is directly responsible for the credibility of the network. The nodes that make up a network may not have much confidence in one another, but they could have faith in the algorithms that drive it. As a consequence of this, the blockchain is improved by every decision that is taken on the network. Using the blockchain provides this as one of its many benefits.

g) Quicker settlement: The processing of transactions via traditional banking systems may often take a very lengthy period. But, with the assistance of blockchains and other contemporary systems, monetary transactions may be completed in a shorter amount of time, saving the user important time.

Role of blockchain in big data

The fact that everything that takes place on a blockchain is encrypted makes it capable of providing the greatest possible degree of security. In a similar fashion, the data that are saved on the blockchain cannot be changed. To be on the safe side, the file signatures of all of the nodes in the network may be checked across all of the ledgers in the network to make sure that they have not been altered. If the record is modified in any way, the signature will no longer be valid.

The applications of big data and blockchain are mutually beneficial. Big data technologies can handle any data, regardless of its variety, velocity, or volume. Blockchain applications simplify operations in any industry. The traditional information processing architecture and business transaction processing have been rendered obsolete as big data and blockchain technologies have grown over the last several years. This has made it possible to abandon these processes. It is necessary for big data to have processing power that is capable of handling big data's processing capacity; conversely, big data requires processing power that is capable of handling the complexity of blockchain and its fast growth.
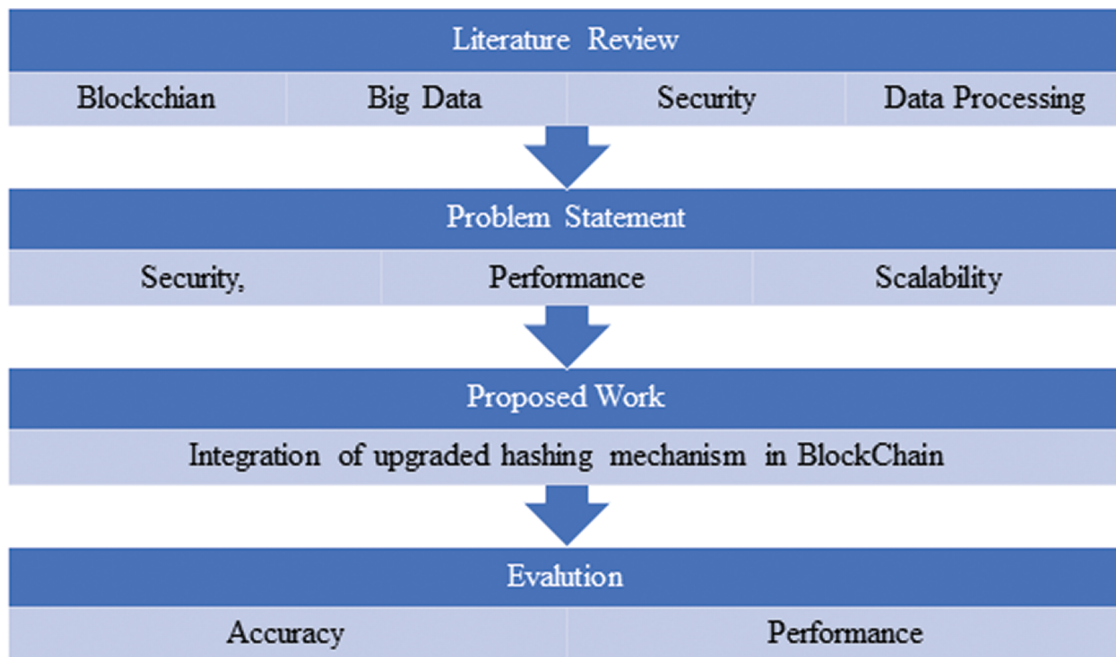
FIGURE 1 | Research methodology.

This sheds light on several implementations of blockchain technology in big data among various industries:

1. Accelerating the financial services sector, the combination of blockchain technology and large amounts of data held by financial institutions will make it possible to estimate risk and spot suspicious trends in real time. Making use of blockchain technology as a way of conducting transactions would assist to safeguard banks and their clients from fraud, speed up the process of transactions, and minimize the cost of transferring money between accounts. For instance, in the most recent few years, in order to streamline the process of moving money from one bank account to another using technology known as blockchain, an association consisting of 47 Japanese banks joined a blockchain business known as Ripple. The goal here is to carry out real-time transfers while minimizing associated costs. The cost of conventional real-time transactions is high because of the possibility of risk factors such as double spending, which may be avoided with blockchain technology.

2. Security in businesses other than banking: Businesses in healthcare, public administration, and other areas have begun adopting blockchain technology to manage data and thwart hacking attempts and avoid data breaches.

3. Monitoring of the supply chain: A blockchain is employed to maintain tabs on the commodities that make up the supply chain, and a mobile app is utilized to monitor the locations of these commodities as they move. Walmart is a wonderful example of this as it uses the technology of blockchains to enhance food safety by allowing for more accurate monitoring of items from the farm to the shop shelf. Users have the ability to get credible information on the provenance of their food while using this strategy (**Figure 1**).

Conclusion

The proposed research is supposed to be capable to resolve the security issue in big data using blockchain. This research would provide a solution to confirm the impact of blockchain on big data security. The goal of the proposed work is to make use of big data with blockchain applicable in real-life scenario. The proposed work offers a broad range of options and flexibility. Improved precision and efficiency are expected as a result of this investigation. The proposed work provides a wide range of possibilities and adaptability.

References

1. Mechkaroska D, Popovska-mitrovikj A, Dimitrova V. Secure big data and IoT with implementation of blockchain technology. *Int Sci J Secur Fut.* (2018) 185:183–5.

2. Tariq N, Asim M, Al-Obeidat F, Zubair Farooqi M, Baker T, Hammoudeh M, et al. The security of big data in fog-enabled IoT applications including blockchain: a survey. *Sensors* (2019) 19:1788. doi: 10.3390/s19081788

3. Ferrag MA, Maglaras L, Janicke H. Blockchain and its role in the internet of things. In: Kavoura A, Kefallonitis E, Giovanis A editors. *Strategic innovative marketing and tourism.* Berlin: Springer (2019). p. 1029–38. doi: 10.1007/978-3-030-12453-3_119

4. Dujak D, Sajter D. *Blockchain applications in supply*. Berlin: Springer International Publishing (2019). doi: 10.1007/978-3-319-91668-2_2

5. Valdeolmillos D, Mezquita Y, González-Briones A, Prieto J, Corchado JM. Blockchain technology: a review of the current challenges of cryptocurrency. *Proceedings of the blockchain and applications international congress*. Berlin: Springer (2020). p. 153–60. doi: 10.1007/978-3-030-23813-1_19

6. Bertino E, Ferrari E. Big data security and privacy. In: Flesca S, Greco S, Masciari E, Saccà D editors. *A comprehensive guide through the Italian database research over the last 25 years*. Cham: Springer (2018). p. 425–39. doi: 10.1007/978-3-319-61893-7_25

7. Babar M, Arif F. Real-time data processing scheme using big data analytics in internet of things based smart transportation environment. *J Ambient Intell Human Comput.* (2018) 10:4167–77. doi: 10.1007/s12652-018-0820-5

8. Dasgupta D, Shrein JM, Datta K. A survey of blockchain from security perspective. *J Bank Financ Technol.* (2018) 3:1–17. doi: 10.1007/s42786-018-00002-6

9. Kumar A, Abhishek K, Nerurkar P, Khosravi MR, Rukunuddin M, Shankar A. Big data analytics to identify illegal activities on bitcoin blockchain for IoMT. *Pers Ubiquit Comput.* (2021) 1–12. doi: 10.1007/s00779-021-01562-z

10. Meng Y, Nazir S. A decision support system for the uses of lightweight blockchain designs for P2P computing. *Peer Peer Netw Appl.* (2021) 14:2708–18. doi: 10.1007/s12083-021-01083-9

11. Niranjanamurthy M, Nithya BN, Jagannatha S. Analysis of blockchain technology: pros, cons and SWOT. *Cluster Comput.* (2018) 22. doi: 10.1007/s10586-018-2387-5

12. Singh SP, Nayyar A, Kumar R, Sharma A. Fog computing: from architecture to edge computing and big data processing. *J Supercomput.* (2019) 75:2070–105. doi: 10.1007/s11227-018-2701-2

13. Zhang P, Lu Q, Hu X, Gu S, Yang L, Min M, et al. Latest progress of the Chinese meteorological satellite program and core data processing technologies. *Adv Atmos Sci.* (2019) 36:1027–45. doi: 10.1007/s00376-019-8215-x

14. Azmoodeh A, Dehghantanha A, Choo KR. Big data and internet of things security and forensics: challenges and opportunities. In: Dehghantanha A, Choo KK editors. *Handbook of big data and IoT security*. Cham: Springer (2019). p. 10–3. doi: 10.1007/978-3-030-10543-3

15. Ahmadvand H, Goudarzi M, Foroutan F. Gapprox: using Gallup approach for approximation in big data processing. *J. Big Data* (2019) 6:20. doi: 10.1186/s40537-019-0185-4

16. Silva BN, Diyan M, Han K. *Big data analytics*. Berlin: Springer (2019). doi: 10.1007/978-981-13-3459-7_2

17. Rathee G, Sharma A. A hybrid framework for multimedia data processing in IoT-healthcare using blockchain technology. *Multimed Tools Appl.* (2019) 79:9711–33. doi: 10.1007/s11042-019-07835-3

18. Calvaresi D, Leis M, Dubovitskaya A, Schegg R, Schumacher M. *Trust in tourism via blockchain technology: results from a systematic review*. Berlin: Springer International Publishing (2019).

19. Czachorowski K, Solesvik M, Kondratenko Y. *The application of blockchain technology in the maritime industry*. Berlin: Springer International Publishing (2019). doi: 10.1007/978-3-030-00253-4_24

20. Khan MM, Ikram M. Use of blockchain in education: a systematic literature review. In: Nguyen NT, Gaol FL, Hong T-P, Trawński B editors. *Intelligent information and database systems*. Cham: Springer (2019). p. 191–202. doi: 10.1007/978-3-030-14802-7_17

21. Xu M, Chen X, Kou G. A systematic review of blockchain. *Financ Innov.* (2019) 5:27. doi: 10.1186/s40854-019-0147-z

22. Firdaus A, Faizal M, Razak A, Feizollah A. The rise of "blockchain": bibliometric analysis of blockchain study. *Scientometrics* (2019) 120:1289–331. doi: 10.1007/s11192-019-03170-4

23. Chen J, Lv Z, Song H. Design of personnel big data management system based on blockchain. *Fut .Gener Comput Syst.* (2019) 101:1122–9. doi: 10.1016/j.future.2019.07.037

24. Nandan S, Ramya KC, Sheeba Rani S, Gupta D, Shankar K, Lakshmanaprabu SK, et al. An efficient lightweight integrated blockchain (ELIB) model for IoT security and privacy. *Fut Gener Comput Syst.* (2020) 102:1027–37. doi: 10.1016/j.future.2019.09.050

25. Deepa N, Pham VQ, Nguyen DC, Bhattacharya S, Prabadevi B, Gadekallu TR, et al. A survey on blockchain for big data: approaches, opportunities, and future directions. *Fut Gener Comput Syst.* (2022) 131. doi: 10.1016/j.future.2022.01.017