JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Blockchain & Big Data : Challenges & Applications

Dr. H.R.Vyawahare

Email ID:harsha.vyawahare@gmail.com

Faculty of Computer Science and Engineering

Sipna College of Engineering and Technology,

Amravati

***Abstract:*—The widespread use of blockchain technology and its enormous range of applications has resulted with much ongoing research in different areas. The Blockchain, however new and yet in the experimenting stage, is viewed as a revolutionary solution that addresses current technology challenges. like decentralization, trust, identity, data ownership and data-driven decisions. The amount and variety of digital data produced by people and machines worldwide is increasing at the same time. Blockchain technology contributes significantly by proviving the best way to store, organise, and process Big Data. Blockchain uses different technologies for bigdata management. This article maily focuses on ways the blockchain uses to deal with big data. Decentralization and encryption of the Blockchain have enormous potential to improve the management and implementation of large-scale data. This article works on cutting-edge methods, openness and future development. The article presents blockchain schematic diagram, different blockchains and motivation that drives them to reconcile. Finally, paper presents the difficulty and the future direction are other driving research that has been conducted in this promising region.**

Keywords: Blockchain, Bigdata, Decentralization, encryption.

I. INTRODUCTION

Since everything is digitised today, data is expanding rapidly. Data could be structured or unstructured. Many times dealing with and processing big data is difficult without compromising availability, performance, scalability, or security. Various sources (such as mobile devices, social media, computer logs, and various sensors around us) generate large amounts of data at high speed. Huge amounts of data is generated globally, and at an unprecedented rate. Thus, the amount of data generated is rising exponentially. The evolution of new technologies and paradigms such as the Internet of Things (IoT) has further increased the pace of data generation.

There are various challenges and issues associated with big data techniques and applications, for example, data security and privacy, energy management, scalability of computing infrastructure, data management, data interpretation, real-time data processing, big data intelligence. Among these challenges, security and privacy have been considered as important issues.

Recently, blockchain as a ledger technology has emerged as attractive solutions for dealing with big data. It was shown in [7] that blockchain can prove beneficial in providing high-quality data and securing data sharing for industrial IoT applications. Due to it's advantages, blockchain can play vital role in transforming current big data systems by enhanching it's security features and network management capabilities for enabling it to provide newly emerging big data services and applications. This paper, presents a review of blockchain for big data, ranging from it's approaches to opportunities and future directions.

The structure of the survey is shown in Figure 1. Section 2 introduces the concept of Blockchain and discusses the motivation for its integration.

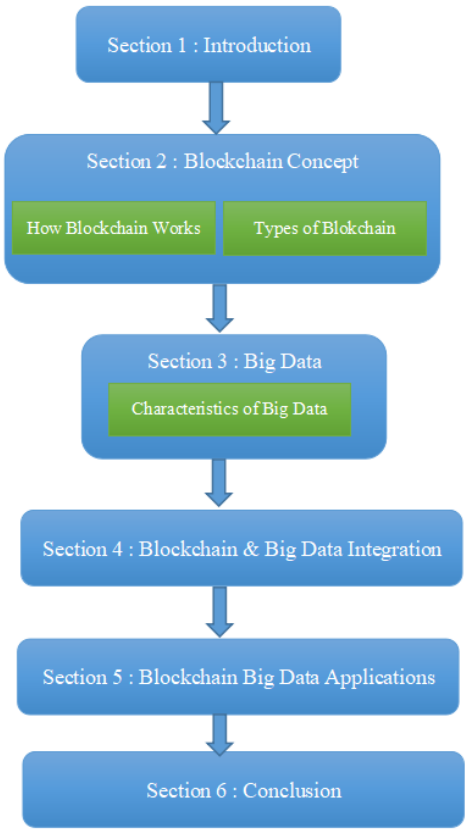

Figure 1: Orgnization of this Article

II. BLOCKCHAIN CONCEPT

Blockchain is a sort of calculation paradigm made up of distributed ledgers, originally developed in 2008 by Satoshi Nakamoto. Blockchain is emphasized by the Bitcoincrypto currency network, comprising all transactions conducted in its

network which are encrypted and executed collectively by peer nodes. Most of the market and exposure of the financial industry to technology is driven by Blockchain. Blockchain is rqecognized as disruptive and effective as a stable and autonomous computing infrastructure.

2.1 HOW BLOCKCAHIN WORKS

Legitimate transaction lots are included in the block that are hashed and encoded as a Markel tree. In each block, the cryptographic hash of the previous block in the Blockchain is maintained, linking the two together. The linked blocks thus create a chain. The trustworthiness of the past square is affirmed by this iterative loop, right to the first unique square. Sometimes, it is possible to create different blocks at the same time, thereby creating temporary forks.

Often, the peers helping the database have various historical versions. They just keep the database's highest-scoring edition they know. Whenever a peer receives a higher-scoring version (usually an old version with a new block added), their database will be extended or overwritten and peer changes will be re-delivered.

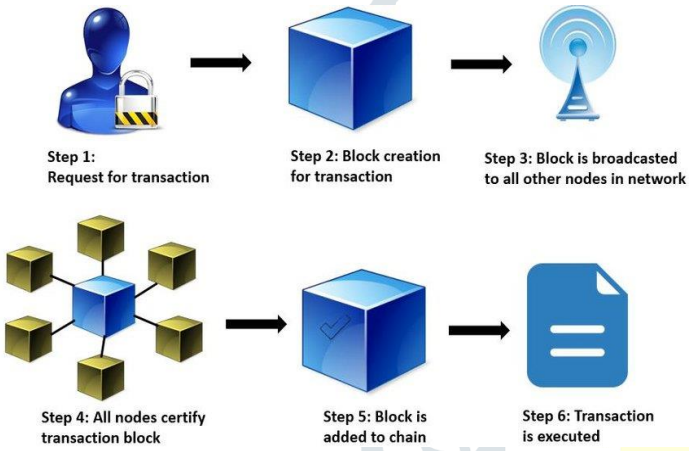

Figure 2: Working of Blockcahin

Blockchain is commonly used to apply new block scores to existing blocks and is encouraged to extend rather than overwrite old blocks using new blocks. Hence, as more blocks are constructed on top, the likelihood of an entry being replaced decreases exponentially and ultimately becoming very low. For instance, Bitcoin uses a proof-of-work scheme, where the network considers the chain with the most accumulated workload as legitimate. There are several techniques that can be used to display a sufficient level.

2.2 .TYPES OF BLOCKCHAIN :

As shown in Figure 3, there are present at least four forms of Blockchain networks: public Blockchain, private Blockchain, consortium Blockchain, and hybrid Blockchain.

1. PUBLIC BLOCKCHAINS

This is the first type of blockchain technology which helped to popularize distributed ledger technology (DLT). It removes the problems that come with centralization and it doesn't store information in any one place, instead distributing it across a peer-to-peer network. Its decentralized nature requires consensus algorithm. Proof of work (PoW) and proof of stake (PoS) are two common consensus methods.Public blockchain is non-restrictive and permissionless, and anyone with internet access can sign on to a blockchain platform to become an authorized node. This user can access current and past records and conduct mining activities. Here anyone can verify the transactions, find bugs or propose changes because the source code is usually open source. Here the examples include Bitcoin, Ethereum and Litecoin

2. Private Blockchains

This Blockchain network works in a restrictive environment like a closed network, or that is under the control of a single entity. A private Blockchain, also referred to as a licensed Blockchain has several important variations from the public Blockchain.

- There is a need of participants to agree to join the network.
- Transactions are confidential and only open to participants in the ecosystem who have been allowed to join the network.
- Private Blockchains are more cohesive than Blockchains that are public.

For businesses who want to cooperate and exchange data but do not want their confidential business data to be available on the public Blockchain in such cases private Blockchains are very useful. These chains are more concentrated, in nature. There is considerable influence over participants and governance mechanisms by the organizations in the organizational chain. A personal Blockchain may or may not contain chain-related tokens.

Examples of private blockchains include Multichain and Hyperledger projects (Fabric, Sawtooth), Corda, etc.

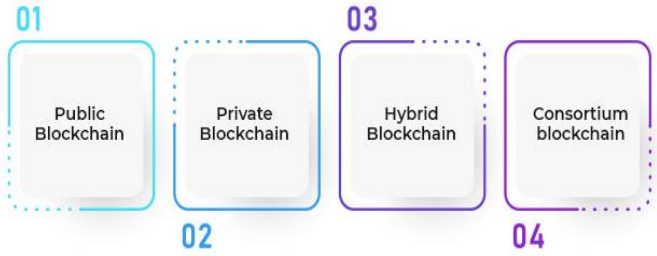

Figure 3: types of Blockchain

3. CONSORTIUM BLOCKCHAINS

A consortium blockchain is a semi-decentralized type where more than one organization manages a blockchain network. This is contrary to a private blockchain, which is managed by only a single organization. More than one organization can act as a node in this type of blockchain and exchange information or do mining. Consortium blockchains are typically used by banks, government organizations, etc.

Examples of consortium blockchain include Energy Web Foundation, R3, etc.

- This model of collaboration offers few of the best use cases for Blockchain's benefit, bringing together a group of "friends," working together but competing with each other.
- By cooperating in certain aspects of the business, they can increase efficiency both individually and collectively.
- The participants of the consortium Blockchain can include anyone from the central bank to the government to the supply chain.

4) HYBRID BLOCKCHAINS

A hybrid blockchain is a combination of the private and public blockchain. Thus one can have a private permission-based system as well as a public permission-less system. With such a

hybrid network, users can control who gets access to which data stored in the blockchain. Only a selected section of data or records from the blockchain can be allowed to go public keeping the rest as confidential in the private network. The hybrid system of blockchain is flexible so that users can easily join a private blockchain with multiple public blockchains. A transaction in a private network of a hybrid blockchain is usually verified within that network. But users can also release it in the public blockchain to get verified. The public blockchains increase the hashing and involve more nodes for verification. This enhances the security and transparency of the blockchain network. Example of a hybrid blockchain is Dragonchain.

Drago chain occupies a unique position in the Blockchain ecosystem in the hybrid ecological chain. This implies that it blends the privacy benefits of authorization and private Blockchains with the protection and accountability benefits of government Blockchains.

III. BIG DATA

Big data is data on an unprecedented scale. Big data is a term used to characterize a large data set. that grows exponentially over time. To put it plainly, such information is so enormous and complex that conventional information, the executive's devices can't store or deal with it proficiently. Big data usually has 5-V characteristics, including capacity, speed, diversity and accuracy. briefly describe these functions of big data.

3.1. CHARACTERISTICS OF BIG DATA:

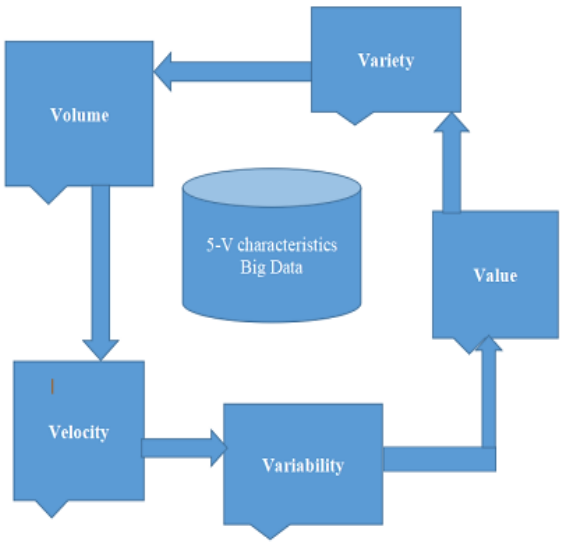

Figure 4: 5-V characteristics

(i) Volume**:-**The name Big Data itself is synonymous with an immense size. In evaluating the importance of data, data size plays a crucial role. Moreover, the quantity of data depends on whether particular data should truly be considered as big data. Volume is a portion along these lines that should be viewed as managing large data.

(ii) *Variety:*-Diversity is the next feature of big data. Variety refers to the essence of heterogeneous sources and data that are organized and unstructured. The only sources of data considered by most applications were spreadsheets and databases in the early days. In research applications, data in the form of e-mail, images, videos, surveillance devices, PDF and audio etc. is now often taken into account. In the storage, mining and analysis of data, all kinds of unstructured data pose certain problems.

(iii)*Velocity***:** Time speed refers to the speed at which it produces information. The speed of generation and processing of data should satisfy the demand that determines the actual potential of the data. Big data speed addresses the speed at which data flows from sources such as business processes, programed logs, web and social media accounts, sensors and mobile devices.

(iii)*Variability* **:–** This implies that often the data can display inconsistencies, which hinders the process of being able to process and handle the data effectively.

(V)*Value***:** Creating meaning is the ultimate challenge in big data. Current systems and procedures are often so complex that it is hard to use data and derive real value.

IV. BLOCKCHAIN AND BIGDATA INTEGRATION

Blockchain and Big data are the number one emerging technologies on the agenda of many companies. It is expected that in the next few years, both will completely change the way businesses and organizations operate. For a long time, it has been developed in a separate way, and at first glance, these technologies may be considered as mutually exclusive. But this idea is changing rapidly. More and more people expect that distributed ledgers will help enterprises finally master big data.

When people are increasingly interested in the intersection of Big data and Blockchain, the overall improvement of captured data and the rapid transmission of said data have caused a great sensation. Blockchain will enable companies to confidently identify the integrity of the data generated. As Blockchain begins to become mainstream, consensus-driven timestamps, proper audit trails, and immutable entries will all become better. PricewaterhouseCoopers has already begun testing Blockchain analysis tools to track digital tokens after they are released. This can assist organizations with shielding their tokens from danger or maltreatment because of illegal tax avoidance and other crimes. The viable mix of huge information and Blockchain can likewise make the healthcare field gradually accepted, because it will enable healthcare providers to easily and safely share records with patients, their doctors, insurance providers, judicial departments, employers, etc. Since the healthcare industry actually operates based on how data is generated and the function of transferring it from one place to another.

- The rich partnership between big data and Blockchain is because the limitations in big data can easily be covered up by Blockchain. Motivation for integration includes security, transperancy, decentralization and flexibility The Blockchain's greatest advantage is the protection it provides to the data stored in it. Also the transparent design of the Blockchain will assists to trace data back to its original stage. The information contained in the Blockchain is not the property of one person. Therefore, there is no risk of data being stolen if the company is compromised in any way.

V. BLOCKCHAIN DATA APPLICATIONS.

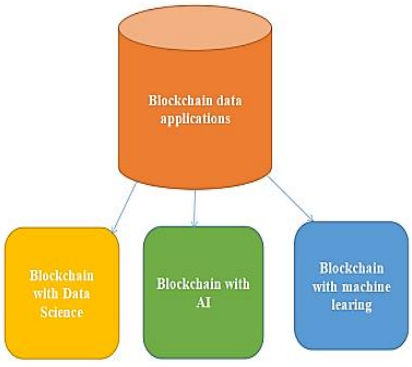

Figure 5:Blockchain data applications

I) *Blockchain with Data Science*: Data science is the science of extracting valuable insights and information from structured and unstructured data to solve practical problems. The growth of data science can be attributed to the rapid development of big data. Big data deals with huge amounts of data, and these data are usually not managed by conventional data processing techniques.Because the Blockchain adopts a peer-to-peer relationship, it makes the channel of the ledger transparent. It also shows users what data can be used reliably, where it comes from, how to change it, etc. Blockchain technology makes possible the entire history of data in a distributed ledger from entry point to exit point. In order to provide powerful data analysis and predictive modeling capabilities, data science needs access to reliable and powerful data sets. The decentralized nature of the Blockchain makes it possible for data scientists to enhance their ability to manage data and establish a reliable data infrastructure to ensure data integrity.

Like any other data, Blockchain data can be used to derive valuable insights about behavior and trends, and can be used to predict future results more accurately. Blockchain can provide data scientists with access to large amounts of structured data.

II) *Blockchain and AI*: Blockchain and synthetic intelligence are currently the 2 freshest technological traits. Despite the fact that the two technologies are very distinct in phrases of developers and applications, researchers have been discussing and exploring their mixture

APPLICATIONS OF AI AND BLOCKCHAIN:

1.*Smart Computing Power*: Lot of computing energy is neded to run a Blockchain with all the encrypted information on a device. For instance, a brute force method is used by the hash set of rules used to mine Bitcoin blocks, including systematically enumerating all feasible applicants for the solution and checking out if every candidate satisfies the hassle announcement before verifying the transaction.

2. *Data Protection*: Blockchain, then again, is largely a technology that permits for secure facts garage on dispensed ledgers. It enables the development by means of events who have accepted this of a totally safe database for evaluation.

3. Data Monetization: Facts monetization is another technological development that may be feasible by means of the mixture of the two technology. Monetizing accumulated information is a main sales hotspot for large groups, together with FB and Google.

III) Blockchain data with machine learning: Machine learning can help in handling many limitations that blockchain-based systems have. The combination of these two technologies (Machine Learning and Blockchain Technology) can provide high-performing and useful results. The learning capabilities of machine learning algorithms can be applied in the blockchain to make the chain smarter than before. This integration can be helpful in the improvement in the security of the distributed ledger of the blockchain. Also, the computation power of ML can be used in the reduction of time taken to find the golden nonce and also the ML can be used for making the data sharing routes better.

VI. CONCLUSION:

This article briefly introduces how beneficial it can be to integrate big data and Blockchain. Paper also highlights the use of Blockchain technology in machine learning. Combined with the main applications of these technologies, the relevance of these technologies and the close connections between them are further discussed. The purpose of this article is to encourage further research on integrating Blockchain technology into Big data and machine learning.

REFERENCES

[1] Cuneyt Gurcan Akcora, Matthew F. Dixon, Yulia R. Gel, Murat Kantarcioglu, “Blockchain Data Analytics”, JOURNAL OF IEEE INTELLIGENT INFORMATICS, VOL. 20, NO. 1, JANUARY 2019

[2] NATARAJAN DEEPA, QUOC-VIET PHAM,DINH C. NGUYEN,DINH C. NGUYEN,PRABADEVI BOOPATHY, “A Survey on Blockchain for Big Data: Approaches, Opportunities, and Future Directions” ACM Comput. Surv., Vol. 1, No. 1, Article . Publication date: September 2020

[3]Mark van Rijmenam Datafloq, “THE CONVERGENCE OF BIG DATA AND BLOCKCHAIN”, December 2017

[4]Daniela Mechkaroska, Aleksandra Popovska-Mitrovikj and Vesna Dimitrova, “SECURE BIG DATA AND IOT WITH IMPLEMENTATION OF BLOCKCHAIN”, INTERNATIONAL SCIENTIFIC JOURNAL,SECURITY & FUTURE, YEAR II, ISSUE 4, WEB ISSN 2535-082X; PRINT ISSN 2535-0668,P.P. 183-185 (2018)

[5] S. Dhanalakshmi, G.CharlesBabu,”An Examination Of Big Data And BlockchainTechnology”,International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-11, September 2019

[6] Hamza ES-SAMAALI ,Aissam OUTCHAKOUCHT and Jean Philippe LEROY”A Blockchain-based Access Control for Big Data”, International Journal of Computer Networks and Communications Security VOL. 5, NO. 7, JULY 2017, 137–147 Available online at: www.ijcncs.org E-ISSN 2308-9830 (Online) / ISSN 2410-0595 (Print)

[7] C. H. Liu, Q. Lin, and S. Wen, “Blockchain-enabled data collection and sharing for industrial IoT with deep reinforcement learning,” IEEE Transactions on Industrial Informatics, vol. 15, no. 6, pp. 3516–3526, Jun. 2019