International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

An Examination Of Big Data And Blockchain Technology

S. Dhanalakshmi, G.Charles Babu

Abstract-: ***Blockchain technology is the process of development of bitcoin, the blockchain technology as a distributed ledger of cryptocurrency transactions for digitized, decentralized, trusted and secured manner. The mainstream of blockchain technology is bitcoin, bitcoin concept made with ledger of every single transaction, transactions allows for hashing mechanism to verify the large amounts of data. Big data task requires that large amount of computational space, to generate the terabytes of data for ensuring the successful data processing techniques. The major impact on big data analytics requires more number of data and generated data can be depending upon different sectors from different organizations. This paper presents a state of definition, characteristics, transaction process, and applications, along with discussion of big data analytics are introduced. In blockchain technology covers the flaws of big data in fruitful relationship, with the factors of security, transparency, decentralization and flexibility, so that data to be analyze in different and efficient way for organizations all sizes in data analytics form.***

Keywords: Big Data, Blockchain, Bitcoin, Cryptocurrency, Classification, Components, Use Cases

I. INTRODUCTION

Big data is used for large data volumes, these data volumes are generated from heterogeneous sources. Its generally refers to massive volumes of complex, high velocity, variable data, diverse types of data, to enable the different technique, and analyze the data, then the data to be managed in effective manner. The scales of big datas; memory capacity of the human brain about 2.5 peta bytes, within two minutes 13 peta bytes amount of data that could be downloaded from the internet. Some examples of big data are social networks, web logs, video and audio archives, military surveillance, astronomy, atmospheric science, genomics, commercial, sensor networks, RFID, biogeochemical and etc. Blockchain technology is a ledger of transaction, the ledger of transactions are shared in trusted environment, then the transactions and records are maintained and growing continuously in distributed manner, the transaction can be inspect with everyone but no single user control. The data are stored in blocks, and blocks are chained together, then block is sealed by hash function, these types of systems are centralized systems. Blockchain work like public ledger, a ledger has different entries of transactions, new entries of transactions are added to the end of the block, if you make change in any blocks the hash of all subsequent blocks also change. If the transactions are done in the same order in their ledgers, any reflection is their, also done with all nodes. The ledgers are distributed and synchronized, need for consensus mechanisms, this consensus mechanisms a protocol which makes all the nodes agree on the transactions and their order. Another kind of blockchain is distributed ledger, distributed ledger and blockchain are often used interchangeably. Hyperledger Fabric and R3 Corda is an example of distributed ledger technology.

II. BLOCKCHAIN

Blockchain technology is sequence of blocks, essentially a trustless, peer-to-peer and continuously growing database of records. All the blocks are ordered and recorded in public ledger, no parent in first block of blockchain; this block is named as genesis block, and the next block points to the hash of the previous block with one parent block. Each block of transactions points to Merkle root. The components of blockchain are block header and transactions; block header is called metadata about a block; the metadata consists of a) previous block hash, b) mining statistics used to construct the block and c) merkle tree root. All the blocks are connected in linked list format, each block having some header, the header connects to the transactions in block, if any transaction to be changed the header of block also changed. Figure 1 shows that sequence of blockchain diagram, and the list of records to be connected in sequence.

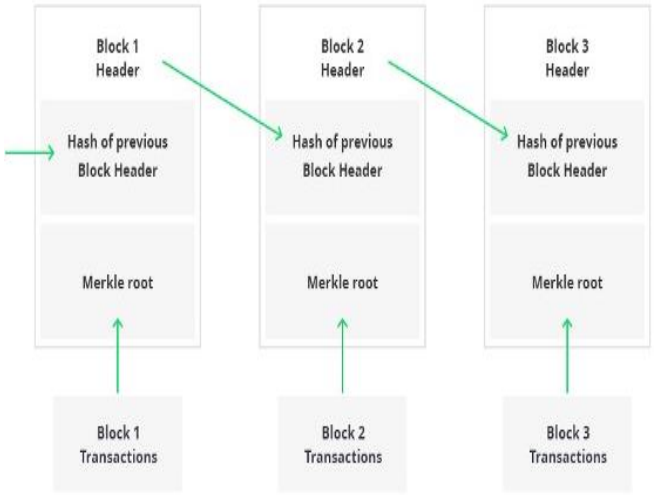


Figure.1. Blockchain Sequence

Manuscript published on 30 September 2019.
*Correspondence Author(s)
S. Dhanalakshmi, Professor, Department of Computer Science and Engineering, Malla Reddy Engineering College (Autonomous), Telangana, India
G.Charles Babu, Professor, Department of Computer Science and Engineering, Malla Reddy Engineering College (Autonomous), Telangana, India

A. Bitcoin

The first application of blockchain is bitcoin systems, the system form that peer to peer network (software that runs on machines of all stakeholders to form the system), permissionless (no identity - no need to sign up anywhere to use; no access control – anyone can participate in any role) and completely decentralized systems (no central party for ordering or recording anything). Bitcoin system releases the token is called BTC, bitcoin is the first cryptocurrency created in 2009, the bitcoin system implements the combination of public transactions & PoW Consensus functions. It's a decentralized system with no central authority for transaction, no one can involved in to the system for direct transactions. Bitcoin systems, its performs two broad operations; transaction management (transfer of bitcoins from one user to another) and Money Issuance (regulate the monetary base). Types of nodes in bitcoin are; Users with wallets - wallets carry user's key pairs, track ownership and authenticate transactions, Miners - compete with each other to add new block of transactions on the blockchain by solving puzzles in PoW system, and Exchanges - portals where users can buy or sell Bitcoin tokens (BTC) in exchange for fiat or other cryptocurrencies.

Table 1. Cryptocurrency Applications on Blockchain

Currency Name	Features	Consensus
Bitcoin(BTC)	First decentralized ledger currency – 2009	PoW
Bytecoin (BCN)	Focused on user privacy through impassive and anonymous transactions - 2012	PoS
Litecoin (LTC)	Uses scrypt as a hashing algorithm in cryptocurrency concept - 2011	PoW (Scrypt)
Peercoin (PPC)	Uses PoW and PoS functions - 2012	PoS with Hybrid
Emercoin (EMC)	Trusted storage for any small data (DNS, PKI, SSL infrastructure etc.) – 2013	PoW
Ripple (XRP)	Designed for P2P debt transfer - 2013	PoS
Waves	An open blockchain platform to develop applications for high volume transactions - 2016	PoS
Omni (MSC)	Both a digital coin and a communication platform built on top of bitcoin blockchain - 2013	PoS
Gridcoin (GRC)	The first cryptocurrency linked to citizen science through Berkley Open Infrastructure for Network Computing - 2013	PoS
Namecoin (NMC)	Used for creating a censor-resistant - 2011	PoW (SHA-256)
Dogecoin (DOGE)	First "Joke Currency" based on internet meme - 2013	PoW (Scrypt)
Myriad	Multiple mining algorithms - 2014	Multi-algorithm PoW
Blackcoin	A type of mining pool - 2014	PoS
Vericoin	Autoexchange to bitcoin for payment - 2014	PoS
Fericoin	Implements a monetary policy - 2012	PoS (SHA2256)

B. Cryptocurrency

Cryptocurrency is the application on blockchain; it's the decentralized currency replacement that acts as to traditional fiat currency (government and central financial institutions). It can have two roles; the currency to represent values of goods and services, other as an investment instrument to the market stock. Cryptocurrency can be classified into two types; altcoins and tokens. Altcoins run their own blockchains, can be divided into, based on bitcoin blockchain (such as Litecoin and Dogecoin) and not based on bitcoin blockchain (ripple or ethereum), and other types are Tokens; its usually represent any asset that is tradable, resides on top of another blockchain; such as non-fungible (used to create verifiable digital scarcity), the examples are cryptokitties, decentralized and fungible tokens (true digital cash), in fungible most tokens are act as a currency, ex., Bitcoin, ether, Zcash. The various cryptocurrency domains are healthcare, mining, eCommerce, Identity, Storage, Gambling, Privacy, Entertainment, Computing, stable, virtual reality, financial, Crowd Funding and etc., The crypto assets can be classified into infrastructure (other systems can be built upon very easily), currency (stores the value in user currency replacement) and services.

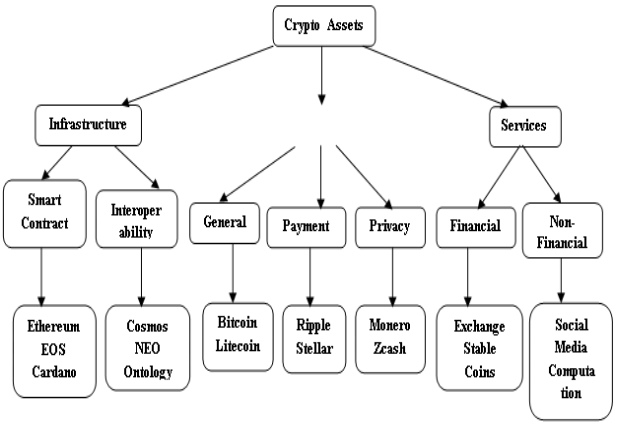


Figure.2. Crypto Assets Classification

C. Building Blocks of Crytpocurrency

Cryptocurrency system is a encryption technique, to generate and verify the units of currency & transfer of funds. It has three main features of cryptocurrency, a secure blockchain, wallets, and mining. Cryptocurrency having three main building blocks; the building blocks are distributed ledger, consensus algorithms, and currency.

a) In distributed ledger – it's a fundamentally asset databases, it has two types unpermissioned (its for public blockchain, everyone has a copy. Ex. Bitcoin, IOTA) and permissioned (private and only accessible to certain parties. Eg. Hyperledger fabric, Quorum)

b) Consensus Algorithms – it's defining that; nodes reflect the same state of ledger, certain set of rules, and remove third party. The variants are Proof-of-Work, Proof-of-Stake, DPoS, PoA, PoT, PoC, PoB, PoR, BFT , PoI, Delayed PoW and etc.,

c) Currency – Currency is define that; representation for values of goods and services. The properties of currency; unforgeable, limited supply, computationally infeasible to duplicate and not be double spent

D. Blockchain Components

The blockchain components can be categorized in to several parts; the components are cryptographic hash function, transactions, asymmetric-key cryptography, addresses, ledgers, blocks, consensus models, forks and smart contracts.

Retrieval Number: K24970981119/19©BEIESP
DOI: 10.35940/ijitee.K2497.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

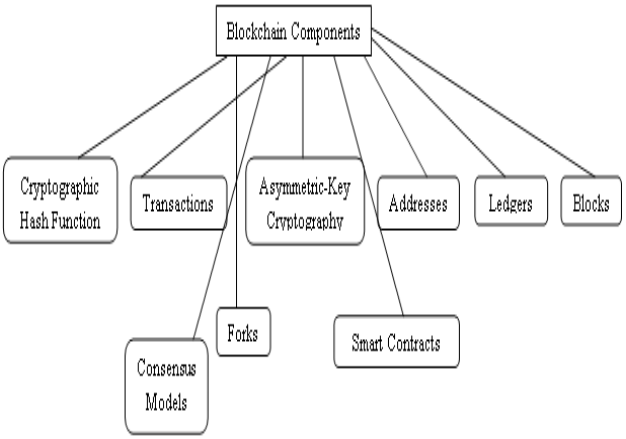
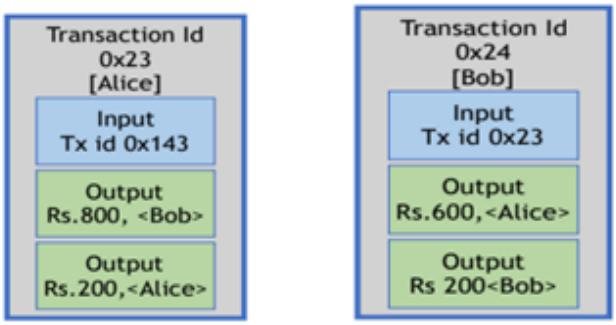


Figure.3. Blockchain Components

The cryptographic hash functions - data can be any size, text, image and file, most common hash function secure hash algorithm (SHA256), the output of SHA256 is 32 bytes (256bits), if every hash computation takes 1 Millisecond.

- Transactions - it represents interaction between parties that is transfer the currency between two users. Each block in a blockchain can contain zero or more transactions; it consists of inputs and outputs, the inputs – assets that are to be 'transferred', and outputs – where the assets 'go'.



a. Input b. Output

Figure.4. Transactions of Input and Output

Asymmetric-key - Asymmetric key consists of public key and private key of two pairs of key systems. In Blockchain network used for asymmetric-key cryptography in secured systems, the systems serves as bitcoin addresses, and used to digitally sign the transactions systems. The encryption key is public and decryption key is private key (anyone can encrypt the data, but decryption only done for intended receiver).

Addresses – In blockchain network, the addresses are used only once, using cryptographic hash function.

Ledgers – It's a collection of transactions, every user maintains their own copy of the ledger

Blocks – A block contains block header (metadata) and block data (list of validated and authentic transactions)

Consensus Models – Determining the next block in the systems, verify and construct a valid block and add to chain, by computing a cryptographic block hash then generating new currency.

Forks – software changes periodically

Smart Contracts – code and data deployed in the blockchain network

E. Types of Blockchain

The blockchain are categorized into public blockchain, private blockchain and consortium blockchain.

1. Public Blockchain – this type of network systems everyone can read and see, & open to all in the platform, and can validate the transactions. Other name of public blockchain is permissionless blockchain; it's a entirely decentralized and peer-to-peer network, very slow compared to private blockchain. Example: Bitcoin
2. Private Blockchain – access permission is limited, and writes access for the private blockchain is controlled by one organization; private blockchain network is shared by encrypted database. Its otherwise called permissioned blockchain, private blockchain also dencentralized and peer-to-peer network. Network members are known, but transactions process are secret for Verifying and validation process, then the transaction systems provide higher level efficiency. Example: Hyperledger Fabric
3. Consortium Blockchain – This system is combination of both public and private blockchain, its partly decentralized network, consortium blockchain network allow for registered participant or everyone to access, and controlled by preset of nodes, then its provides the efficiency and transactional privacy.

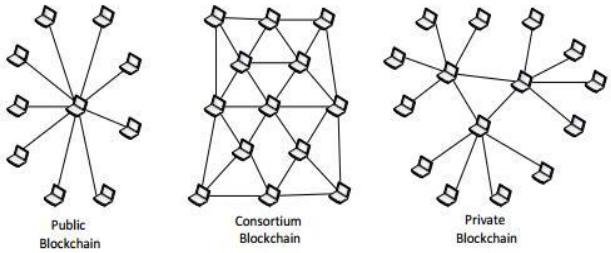


Figure.5. Public vs Consortium vs Private Blockchain

III. BIG DATA TECHNIQUES

Big data techniques, can be classified or identified by 3 parameters (3v's); volume, velocity and variety. The first parameter volume denotes data size, amount of data. The amount of data generated from facebook, the typical data centre like Google, Amazon, the data are generated in the form of trillion of bytes. The second perspective type of parameter in big data technology is velocity, like the speed of change of data in social networking website, and the third important aspect of analytics in big data; it's a kind of geospatial applications, also variety of data applying blockchain for big data applications.

Big Data Enabling Technologies

It's a collection of data and process using different open source tools. The open source tools are; Apache Hadoop, Hadoop Hdfs, Hadoop Yarn, Hadoop Map Reduce, Hive, Cassandra, Apache Zookeeper, Apache Hbase, Apache Spark, No SQL, Kafka, Spark Streaming Ecosystem, Spark MLib, Spark GraphX.

a) Apache Hadoop – Apache hadoop systems are open source framework in big data. It has different basic parts:

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

• Hadoop Distributed File System (HDFS) – this type of systems in hadoop are, performing storage process, and split into different data & distribute the nodes in cluster; Scaling out of H/W resources and Fault Tolerant

• MapReduce - Programming model that simplifies parallel programming. (Map-> apply (), Reduce-> summarize () and Google used MapReduce for Indexing websites, and other parts of Hadoop systems are;

• YARN – A Programming model for processing big data. Flexible scheduling & resource management over HDFS. Yahoo uses YARN to schedule jobs over 40,000 servers

b) MapReduce function, In MapReduce function, hadoop distributed file system move to the map function with the input of (key,value) pairs, the pairs are shuffle & sort ; its produce the intermediate values, the values merged into reduce function, after reduce function final key and value can be produced.

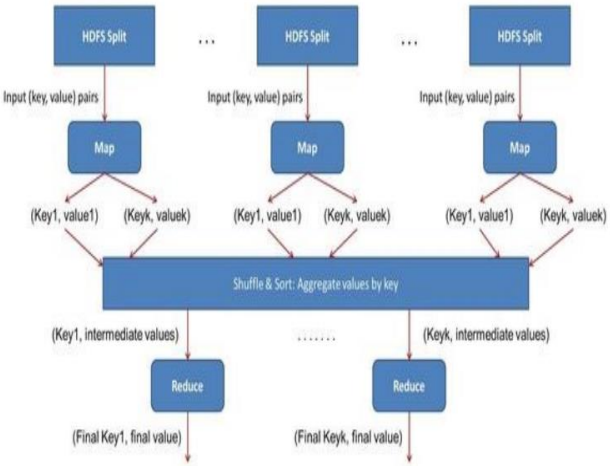


Figure.6. MapReduce Function

c) YARN – YARN is a Yet Another Resource Manager. The resources can be allocated and executed on different cluster nodes.

d) Hive – To access big data, Hive systems support for HSQL.

e) Apache Spark – its otherwise big data analytics framework, academic & industry gained lot off attraction in apache spark, its designed for fast computation

f) ZooKeeper - ZooKeeper is a highly reliable distributed coordination kernel, which can be used for distributed locking, configuration management, leadership election, work queues,...

Zookeeper is a replicated service that holds the metadata of distributed applications. Key attributed of such data; small size, performance sensitive, dynamic and critical.

g) NoSQL - NoSQL technology is new and with wide variety of databases that can stored unstructured data.

h) Cassandra – Data is placed on different machines with more than one replication factor that provides high availability and no single point of failure.

IV. BLOCKCHAIN USECASES IN BIG DATA & RESULTS

Big data analytics process provided in the form of both industry and academia. It operating and managing with own data in Blockchain technology. The blockchain systems ensure that privacy and integrity of data, it predicts large amount of data in big data techniques but focus on validating data, the data to be brought together in decentralized manner and the origin of the data linked in chained manner. The use cases of blockchain technology in big data can be specified that, a) Ensuring trust (data integrity) – verification process ensures its quality, the blockchain record verify authenticity of the documents, b) preventing malicious activities – consensus algorithms to verify transactions and generate the computational power, c) Making Predictions (Predictive Analysis) – predict large sets of data to predict future outcomes and gathered structured data from individual devices, d) real-time data analysis – analysis the data in large scale of organizations to achieve blockchain enabled system, after analyzing observe changes in the data, make possible to change quickly in real time manner, and e) manage data sharing – data can be stored in blockchain network and monertize it, outcome of the network . Stored on the blockchain platform.

V. CONCLUSION

The paper summarizes the overview of blockchain technology and big data systems. In blockchain systems provide the immutable, distributed, decentralized data & also highlight the cryptocurrency and bitcoin systems. Big data systems highlight the importance different enabling techniques revolutionize the way to process and analyze data. Further research to incorporate the different applications and use cases of blockchain technology and Big Data techniques

REFERENCES

1. Voshmgir, S. (2019). What is Blockchain? Retrieved July 2019 from https://blockchainhub.net/blockchain-intro/
2. Xiwei Xu, Ingo Weber, & Mark Staples, Architecture for blockchain applications, Springer-Verlag, 2019
3. Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system". (2008). https://bitcoin.org/bitcoin.pdf
4. Zheng, Z., Xie, S., Dai, H. N., & Wang, H. (2017). An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends, IEEE 6th International Congress on Big Data Congress, June 2017. Hawaii, USA.
5. Tinu, N.S. A Survey on Blockchain Technology - Taxonomy, Consensus Algorithms and Applications. International Journal of Computer Science and Engineering, Volume 6, Issue 5, May 2019. E-ISSN: 2347 - 2693
6. Dev, MLS. Blockchain Architecture Basics: Components, Structure, Benefits & Creation. https://medium.com/@MLSDevCom/blockchain-architecture-basics-components-structure-benefits-creation-beace17c8e77. Mar 7, 2019.
7. Archana Prashanth Joshi, Meng Han, and Yan Wang, A survey on security and privacy issues of blockchain technology, American Institute of Mathematical Sciences, 1(2), 2018, 121 - 147
8. Praveen Jayachandran, The difference between public and private blockchain, https://www.ibm.com/blogs/blockchain/2017/05/the-difference-between-public-and-private-blockchain/,May 31, 2017

9. Deepak Puthal, Nisha Malik, Saraju P.Mohanty, Elias Kougianos, and Gautam Das, Everything You Wanted to Know About the Blockchain: Its Promise, Components, Processes, and Problems, IEEE Consumer Electronics Magazine, Volume 7, Issue 4, July 2018. https://www.researchgate.net/publication/326102908_Everything_You_Wanted_to_Know_About_the_Blockchain_Its_Promise_Components_Processes_and_Problems/download
10. Amit Verma, Top 10 Open Source Big Data Tools in 2019, March 12, 2018. https://www.whizlabs.com/blog/big-data-tools/
11. Muddasir khan , Bharti Kalra. An inspection on big data computing, International Journal of Engineering & Science Research, IJSER/ICRIT 2018, Special Issue, Article No-52/36-329,ISSN 2277-2685
12. Smartereum, Feb 22, 2019. Blockchain and big data use cases: Challenges and opportunities Opportunities with blockchain and data science. https://www.ibm.com/blogs/blockchain/2019/02/blockchain-and-big-data-use-cases- challenges-and-opportunities-with-blockchain-and-data-science/. Published January 5, 2019.