

A Review on Big Data and Blockchain Technology

Rajesh Mothe¹, S Tharun Reddy², Y.Nagender³, A.Harshavardhan⁴, Seena Naik Korra⁵
Department of Computer Science and Engineering, S R Engineering College, Warangal^{1,2,3,4,5}
Member, Center for Embedded Systems & Internet of Things⁵
mraj1210@gmail.com

Abstract:- *Blockchain technology is the procedure of advancement of bitcoin, the blockchain technology as a distributed ledger of cryptocurrency exchanges for digitized, decentralized, trusted and secure way. The standard of blockchain technology is bitcoin, bitcoin idea made with record of each and every transaction, transaction considers hashing mechanism to check a lot of data. Big data task necessitates that huge amount of computational space, to produce the terabytes of data for guaranteeing the effective dataprocessing techniques. The significant effect on big data analytics requires increasingly number of data and created data can be contingent on various areas from various organizations. This paper presents a state of definition, qualities, transaction procedure, and applications, alongside talk of big data analytics are presented. In blockchain technology covers the defects of big data in productive relationship, with the elements of security, decentralization, transparency, decentralization and adaptability, so data to be analyze in various and effective ways for organizations all sizes in dataanalytics structure.*

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

I. INTRODUCTION

Big Data is utilized for largedata volumes, these data volumes are created from heterogeneous sources. Its for the most part alludes to large volumes of complex, high speed, variable data, assorted types of data, to empower the diverse method, and examine the data, at that point the data to be overseen in viable way. The sizes of big pieces of data; memory limit of the human mind about 2.5 peta bytes, inside two minutes 13 peta bytes measure of data that could be downloaded from the web. A few instances of large data are interpersonal organizations, web logs, video and sound files, military observation, stargazing, environmental science, genomics, business, sensor systems, RFID, biogeochemical and so forth. Blockchain innovation is a record of exchange, the record of exchanges is partaken in confided in condition, at that point the exchanges and records are kept up and developing consistently in circulated way, the exchange can be review with everybody except no single client control. The data are put away in squares, and squares are blinded together, at that point square is fixed by hash work, these sorts of frameworks are brought together frameworks. Blockchain work like open record, a record has various passages of exchanges, new sections of exchanges are added as far as possible of the square, in the event that you make change in any hinders the hash of every single resulting square additionally change. On the off chance that the exchanges are done in a similar request in their records, any reflection is there, additionally finished with all hubs. The records are appropriated and synchronized, requirement for accord systems, this accord instruments a convention which makes every one of the hubs concede to the exchanges and their request. Another sort of blockchain is conveyed record, dispersed record and blockchain are frequently utilized conversely. Hyperledger Fabric and R3 Corda is a case of circulated record innovation.

II. BLOCKCHAIN

Blockchain technology is arrangement of blocks, basically a trustless, shared and constantly developing database of records. Every one of the blocks are requested and recorded in public ledger, no parent in first block of blockchain; this block is named as genesis block, and the following block focuses to the hash of the past block with one parent block. Each block of transactions focuses to Merkle root. The parts of blockchain are block header and transactions; block header is called metadata about a block; the metadata comprises of a) previous block hash, b) mining insights used to develop the block and c) Merkle tree root. Every one of the blocks are associated in linked list design, each block having some header, the header interfaces with the transactions in block, if any exchange to be changed the header of block likewise changed. Figure 1 shows that grouping of blockchain diagram, and the list of ledgers to be associated in arrangement.

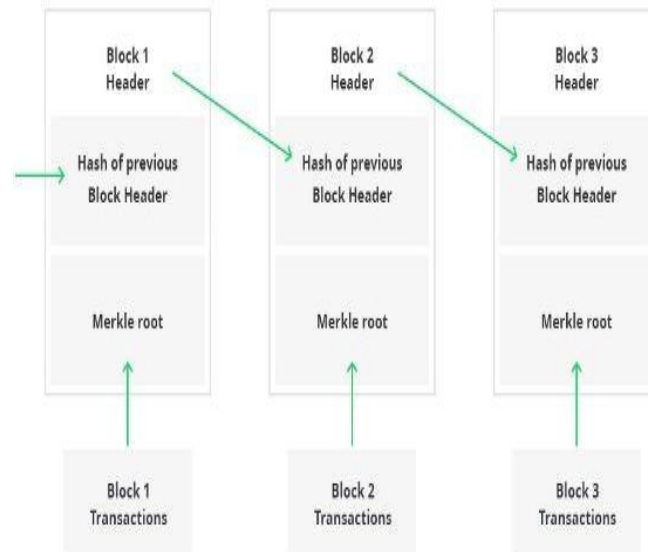


Figure.1. Blockchain Sequence

A. Bitcoin

The principal use of blockchain is bitcoin systems, the system structure that distributed system (programming that sudden spikes in demand for machines of all partners to shape the system), consent less (no personality - no compelling reason to join anyplace to utilize; no entrance control – anybody can take an interest in any job) and totally decentralized systems (no focal gathering for requesting or recording anything). Bitcoin system discharges the token is called BTC, bitcoin is the main digital money made in 2009, the bitcoin system executes the blend of open exchanges and PoW Consensus capacities. It's a decentralized system with no focal expert for exchange, nobody can include in to the system for direct exchanges. Bitcoin systems, it performs two expansive activities; exchange the executives (move of bitcoins starting with one client then onto the next) and Money Issuance (manage the financial base). Kinds of hubs in bitcoin are; Users with wallets - wallets convey client's key sets, track possession and validate exchanges, Miners - contend with one another to include new square of exchanges the blockchain by illuminating riddles in PoW system, and Exchanges - entryways where clients can purchase or sell Bitcoin tokens (BTC) in return for fiat or different digital currencies.

Table1. Cryptocurrency Applications onBlockchain

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 Berkeley 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
Vericoïn	Autoexchange to bitcoin for payment - 2014	PoS
Fericoïn	Implements a monetary policy - 2012	PoS (SHA2256)

B. Cryptocurrency

Digital money is the application on blockchain; it's the decentralized cash substitution that goes about as to conventional fiat money (government and focal budgetary foundations). It can have two jobs; the money to speak to estimations of products and ventures, different as a speculation instrument to the market stock. Digital currency can be characterized into two sorts; altcoins and tokens. Altcoins run their own blockchains, can be partitioned into, in light of bitcoin blockchain, (for example, Litecoin and Dogecoin) and not founded on bitcoin blockchain (wave or Ethereum), and different sorts are Tokens; it's typically speak to any advantage that is tradable, dwells over another blockchain, for example, non-fungible (used to make undeniable advanced shortage), the models are cryptokitties, decentralized and fungible tokens (genuine computerized money), in fungible most tokens are go about as a cash, ex., Bitcoin, ether, Zcash. The different digital money areas are social insurance, mining, eCommerce, Identity, Storage, Gambling, Privacy, Entertainment, Computing, steady, computer generated experience, monetary, Crowd Funding and so forth., The crypto resources can be grouped into framework (different frameworks can be based upon effectively), cash (stores the incentive in client cash substitution) and administrations.

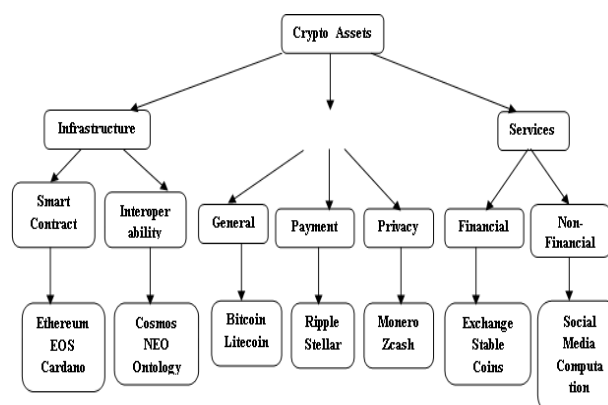


Figure.2. Crypto Assets Classification

C. Building Blocks of Cryptocurrency

Cryptographic money framework is an encryption strategy, to create and confirm the units of cash and move of assets. It has three principle highlights of cryptographic money, a protected blockchain, wallets, and mining. Digital currency having three primary structure obstructs; the structure squares are dispersed record, accord calculations, and money.

- 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 andetc.,
- 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 doublespent

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.

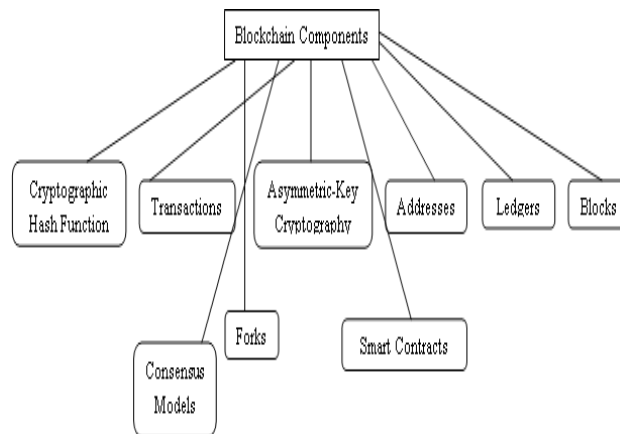


Figure.3. Blockchain Components

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 1Millisecond.

- 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’.

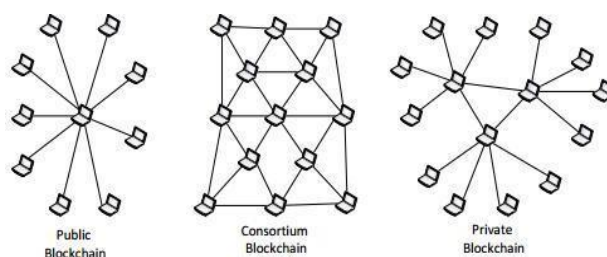


Figure.5. Public vs Consortium vs Private Blockchain

III. BIG DATATECHNIQUES

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 dataapplications.

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.

- Apache Hadoop – Apache hadoop systems are open source framework in big data. It has different basic parts: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 FaultTolerant
- MapReduce - Programming model that simplifies parallel programming. (Map-> apply (), Reduce-> summarize ()) and Google used MapReduce for Indexing websites, and other parts of Hadoop systemsare;
- YARN – A Programming model for processing big data. Flexible scheduling & resource management over HDFS. Yahoo uses YARN to schedule jobs over 40,000 servers
- a) 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 beproduced.

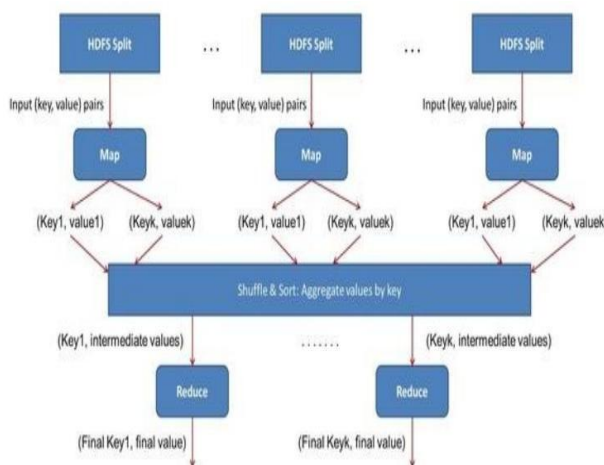


Figure.6. MapReduce Function

- b) YARN – YARN is a Yet Another Resource Manager. The resources can be allocated and executed on different clusternodes.
- c) Hive – To access big data, Hive systems support forHSQL.

- d) Apache Spark – its otherwise big data analytics framework, academic & industry gained lot off attraction in apache spark, its designed for fastcomputation
- e) 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 andcritical.
- f) NoSQL - NoSQL technology is new and with wide variety of databases that can stored unstructureddata.
- g) Cassandra – Data is placed on different machines with more than one replication factor that provides high availability and no single point offailure.

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 monetize it, outcome of the network . Stored on the blockchainplatform.

V. CONCLUSION

The paper outlines the diagram of blockchain innovation and enormous information frameworks. In blockchain frameworks give the unchanging, conveyed, decentralized information and likewise feature the digital money and bitcoin frameworks. Large information frameworks feature the significance distinctive empowering procedures reform the best approach to process and break down information. Further research to consolidate the various applications and use instances of blockchain innovation and Big Data strategies

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. Rajesh Mothe, S Tharun Reddy, Kanegonda Ravi Chythanya, YSupraja Reddy, Challenges, Open Research Issues and Tools in Bigdata Analytics, International Journal of Recent Technology and Engineering, Volume-8, Issue-2S11, September 2019
5. 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.
6. 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

7. G.Sunil,MohdSallauddin,Kanegonda Ravi Chythanya, The Current Status And Research In Industrial Big Data Analysis In Smart Intelligent Systems, International Journal of Advanced Research in Computer Science
8. Dev, MLS. Blockchain Architecture Basics: Components, Structure, Benefits & Creation.<https://medium.com/@MLSDevCom/blockchain-architecture-basics-components-structure-benefits-creation-beace17c8e77>. Mar 7,2019.
9. 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
10. M. Rajesh, B. Vijay Kumar, G.Sunil,Big Data Technology Applications In AutomatedHealth Care Management System , International Journal Of Advanced Research InComputer Science, Vol. 8-9, DEC 2017.
11. 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
12. *G.Roopa,D.Ramesh, Designing a Collaborative Detection System for Detecting the Threats to the Cyber Security in Big Data, Indian Journal of Public Health Research & Development, 9-11,NOV, 2018.*