[[1]](#footnote-1)

Blockchain Analysis (April 2019)

Yumei Bennett, *Graduate Student, Southern Methodist University*, Daniel Byrne, *Graduate Student, Southern Methodist University*, Benjamin Wilke, *Graduate Student, Southern Methodist University*, Brian Coari, *Graduate Student, Southern Methodist University*

*Abstract*— Blockchain technology is at the heart of bitcoin and is increasingly being developed for use in a broad range of financial applications. The goal of this project is to provide a survey and tutorial on Blockchain technology. The project and writeup will include an evaluation of an experimental Blockchain database.

*Index Terms*— Computer network management, Cooperative communication, Cryptographic protocols, Data transfer, Encryption, Distributed databases, Document handling, Fault tolerant control, Identity management systems, Information exchange, Internet, Network function virtualization, Network security, Peer-to-peer computing

# INTRODUCTION

T

HIS paper will explore the distributed ledger technology known as Blockchain. In this paper we will cover what Blockchain is, the theories and technologies behind Blockchain, its strengths and shortcomings, as well as its current and possible future uses. We will explain the different types of Blockchains (public, private, and Federated), and the pros and cons of each. This paper will conclude with a demonstration of an emerging application of Blockchain, [BigchainDB](https://www.bigchaindb.com/), which is a NoSQL database augmented with the core strengths of Blockchain, its decentralized nature and its immutability.

In this paper we will explain that Blockchain is, at its heart, simply a system that tracks a history of events called a [digital ledger][1]. We will define terms such as Blocks, the Blockchain Distributed Database, Block Time, and the Blockchain Community. We will examine, in detail, how every event knows a unique identifier of the event that came before it, and when the event is approved by the Blockchain Community the event is given a timestamp and added to its chain. When an event is added to a chain, every user that is part of the Blockchain Community adds that event to their copy of that chain simultaneously, which creates the [Blockchain Distributed Database][2]. This helps prevent tampering with the events in a chain, since an attacker would need to edit every chain from every user in the Blockchain Community for a change to be accepted, which means the attacker would need access to the private databases of everyone in the Blockchain Community. By this use of the Blockchain Distributed Database, Blockchain ensures trust in the history of its transactions and makes it an attractive fit for any process that values a [trusted history.][1]

We will also explore the different types of Blockchains: Public, Private, and Federated. We will explain how public blockchains allow anyone to submit transactions to the blockchain and become part of the Blockchain community, how private blockchains restrict access to submit and verify transactions, and how Federated Blockchain is like a Private Blockchain but allows many organizations to have nodes on the network. We will give some examples and pros and cons of each of these [blockchain types.][3]

At the end of this paper the goal is for the reader to not only have a thorough understanding of why Blockchain exists and what solutions it can provide, but also to be able to join an existing Blockchain instance or to start his or her own Blockchain with some confidence that it can provide the trust missing from so many independent and private transactions.

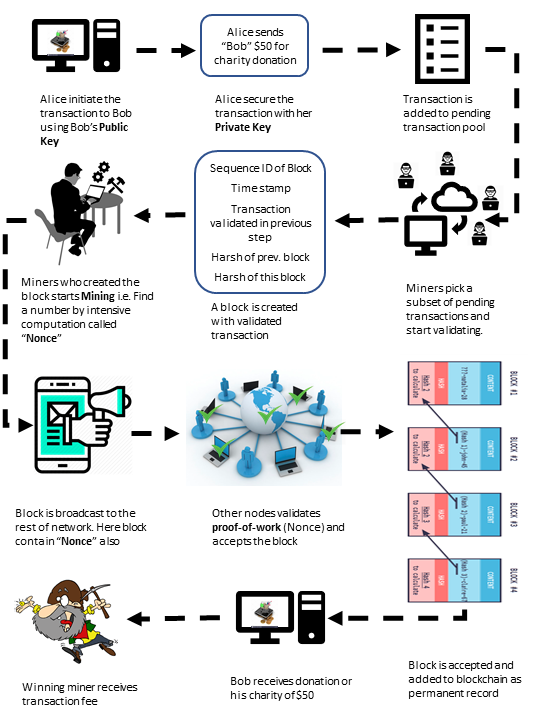
# Blockchain technology

Blockchain technology is probably the best invention since the internet itself. It allows value exchange without the need for trust or for a central authority. For the first time, blockchain enable a way for one Internet user to transfer a unique piece of digital property to another Internet user, such that the transfer is guaranteed to be safe and secure, everyone knows that the transfer has taken place, and nobody can challenge the legitimacy of the transfer. The consequences of this breakthrough are hard to overstate.

Blockchain is a distributed database or a Ledger (as used in Cryptocurrencies) containing all the records linking the history to the very initial record, that can be used in various applications like Currency, Supply Chain, Identity Management, Assets Tracking etc. Distributed because it is not authored or controlled by a centralized entity like a Financial Institution (Banks) or Government, rather, each participating node has a copy of the most updated Ledger to prevent any tampering of data in the Ledger.

The Ledger is divided into Blocks of equal sizes and each block is connected to its previous block using the Cryptographic Hash of the Data in Previous Block, so if someone tries to change the value of a record in any Block, all the Blocks ahead of it are invalidated, thus providing an indication that the data has been tampered.

Here are the steps involved:



1. On a Blockchain, the transactions are happening between different nodes on the network, all the time. In our example, Alice initiate a transaction to Bob.

2. Alice’s wallet encrypt the content with Bob’s public key and signs the transaction with her private key.

3. The transaction will not be added to the block database (ledger) right away, it is added to pending transaction pool.

4. Special nodes, called Miners, pick up several transactions (size varies depending on application, up to 1MB in size for Bitcoin network).

5. Miners first validate all the transactions to see that all the transactions are fair and valid and does not contain any fraudulent transaction with problems like enough resources/fund in Alice wallet, duplicate transaction etc. (Validation rules are application specific.) These validation rules are called transactional consensus.

6. Once the transactions are validated, the miners create a Block of the transactions he picked. This Block contains a sequential ID, the Transactions picked by the miner, SHA256 Hash of all the values and Hash of the Previous Block.

7. The Blockchain Consensus requires the miners to resolve a computational-intense hard mathematical problem to sign the generated block before broadcasting it to the network.

8. The other miners validate the solution of the problem and if validated accepts the block.

9. Once over 50% of the Miners on the blockchain network accepts the block, the block is added to the ledger (block database) and distributed to all nodes on the network.

10. Bob finally sees $50 in his charity donation wallet.

11. The winning miner who created the accepted block is awarded with transaction fee.

# Blockchain Applications

In the year 2017, Blockchain became the second most popular search word on Gartner's website, and Distributed Ledger Technology (DLT) will continue to gain utmost significance across many industries.

Blockchain will help address several modern-day security concerns, including issues with contracts, identity, and fraud management. Blockchain-based lists will allow online retailers and financial organizations to conveniently vet their customers and fight against fraudulent activities.

The following are top 20 real-world blockchain applications across the industry according to CoinSwitch desk.

***Cybersecurity:***

*Guardtime* - a keyless signature which is currently used to secure the health records of one million Estonian citizens.

*REMME* - a decentralized authentication system which aims to replace logins and passwords with SSL certificates stored on a blockchain.

***Healthcare:***

*Gem* - working with the Centre for Disease Control(CDC) to put disease outbreak data onto a blockchain.

*Simply Vital Health* - ConnectingCare which tracks the health progress of patients, and Health Nexus, which aims to provide decentralized blockchain patient records.

***Banking and Financial:***

*Barclays* - Barclays has launched a number of technological initiatives which involves tracking financial transactions, combating fraud and compliance.

*Augur* - Allows the creation of blockchain-based predictions markets for the trading of derivatives and other financial instruments in a decentralized ecosystem.

***Manufacturing:***

*Provenance* - provide a blockchain-based provenance record of transparency within supply chains.

*TransActivgrid* - allowing members to locally produce and sell energy, with the goal of reducing costs involved in energy distribution.

***Government:***

*Dubai* - 30 government departments formed a committee dedicated to investigating opportunities across health records, shipping, business registration and preventing the spread of conflict diamonds.

*South Korea* - Samsung is creating blockchain solutions for the South Korean government which will be put into use in public safety and transport applications.

***Charity:***

*Bitgive* - aims at providing greater transparency to charity donations and reduce operation overhead.

***Retail:***

*OpenBazaar* - a decentralized market where no middleman is involved in trading of goods and services.

*Loyyal* - universal blockchain-based framework which aims at allowing customers to combine trade and loyalty reward points in newer ways so that retailers can offer more sophisticated loyalty packages.

***Real Estate:***

*Ubiquity* - a blockchain-driven system for tracking the complicated legal process which creates friction and expenses in real estate transfer.

***Transportation and Tourism:***

*Arcade City* - An application which aims at beating Uber at their own game by moving ride sharing and car hiring onto the blockchain.

*Webjet* - an online travel portal, which is developing a blockchain solution to allow stock of empty hotel rooms to be efficiently tracked and traded.

***Media:***

*Kodak* - developing a blockchain system for tracking intellectual property rights and payments to photographers.

*Ujo Music* - record and track royalties for musicians and allows them to create a record of ownership of their work.

***Smart Contracts:***

*Slock* - an Ethereum based internet-of-things (IoT) platform, which is using this application so that people can rent anything from bicycles to apartments by unlocking a smart lock after both parties agreed on the terms of the contract.

# Blockchain Database

Blockchain databases are different from traditional centralized databases. They are distributed in nature. For a blockchain database, each participant maintains, calculates and updates new entries into the database. Blockchain databases consist of several decentralized nodes. Each node participates in administration: all nodes verify new additions to the blockchain. For an addition to be made to the blockchain, the majorities of nodes must reach consensus. This consensus mechanism guarantees the security of the network, making it difficult to tamper with.

Blockchain databases are able to keep information that is relevant now, but also all the information that has come before. Blockchain technology can create databases that have histories of themselves. They grow like ever-expanding archives of their own history while also providing a real-time portrait.

The choice of database for blockchain platform or application is somewhat confusing. Some consider blockchain technology is another category of database like relational and document storage. The current blockchain application or platform are built on top of semi-custom database.

For example, Ethereum is using Leveldb, Leveldb was originally designed by Google to be a disk backed memory database. As long as most of it can fit into memory, it performs very well, but as more and more disk access is required, performance can degrade. Facebook's RocksDB tries to address some of these issues with multi-threaded background computation, but it’s still a fairly new thing and being in C makes portability hard. Parity is using RocksDB as it's database. The reason for choosing RocksDB is performance. Faster than LevelDB in their tests. Lately though they have hit some performance issues and a lot of corruption issues with RocksDB and are developing a proprietary DB to better suit Parity, the ParityDB.

So, what are the characteristics of a database that is suited for blockchain technology application development?

***Distributed:***

It is designed to be distributed and synchronized across networks, which makes it ideal for multi-organizational business networks. It also encourages organizations to come out from behind their firewalls and share data.

***Permanent Record:***

All confirmed transactions are stored as permanent record, you can’t just do whatever you want to the data. The types of transactions one can carry out are agreed between participants in advance and stored in the blockchain as “smart contracts,” which helps give confidence that everyone is playing by the rules.

***Majority Rules:***

Before one can execute a transaction, there must be agreement between all relevant parties that the transaction is valid. For example, if you’re registering the sale of a cow, that cow must belong to you or you won’t get agreement. This process is known as “consensus” and it helps keep inaccurate or potentially fraudulent transactions out of the database.

***Mutual Trust:***

Immutability of the data. Once you have agreed on a transaction and recorded it, it can never be changed. You can subsequently record another transaction about that asset to change its state, but you can never hide the original transaction. This gives the idea of provenance of assets, which means that for any asset you can tell where it is, where it’s been and what has happened throughout its life without any central authority.

Taken together, these four characteristics give organizations a high degree of trust in the data and the business network. That level of trust makes blockchain important for the next generation of business applications.

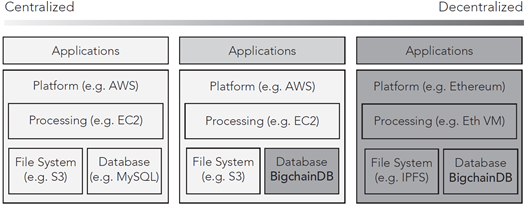
Next, we will explore one of the rising popularity database, BigChain DB who is adapting itself towards supporting blockchain applications.

# BigChain Database

February 10, 2016 — ascribe GmbH has announced the release of BigchainDB, a scalable blockchain database that can process one million writes per second and their first licensed user, Everledger.

BigchainDB achieves one million writes per second with sub-second latency and petabytes of capacity in a federated blockchain database. BigchainDB allows developers and enterprises to deploy blockchain applications and platforms at scale, formerly a limiting factor in getting sandbox proof-of-concepts into production systems. The throughput capacity of BigchainDB exceeds PayPal by 10,000 times, Visa by 20 times and the peak volume of Nasdaq by 10 times.

BigchainDB was a direct result of seeing scalability as a fundamental barrier to wider adoption of blockchain. BigchainDB is complementary to decentralized storage, processing and communication building blocks. It can be used side by side with higher-level decentralized computing platforms and applications, and protocols for identity, financial assets, intellectual property and sidechains. BigchainDB fills a gap in the decentralized stack.



As a blockchain database, BigchainDB is complementary to other decentralized systems, such as decentralized le storage (e.g. IPFS), decentralized data exchange protocols (e.g. Ocean Protocol), smart-contract blockchains (e.g. Ethereum or Hyperledger Fabric), and decentralized processing (e.g. TrueBit). BigchainDB works with centralized computing systems. From a base context of a centralized cloud computing ecosystem (left), BigchainDB can be added as another database to gain some decentralization benefits (middle). It also fits into a fully-decentralized technology stack (right).

BigchainDB is software that has blockchain properties (e.g. decentralization, immutability, owner-controlled assets) and database properties (e.g. high transaction rate, low latency, indexing & querying of structured data). BigchainDB version 2.0 makes significant improvements over previous versions. In particular, it is now Byzantine fault tolerant (BFT), so up to a third of the nodes can fail in any way, and the system will continue to agree on how to proceed. BigchainDB 2.0 is also production-ready for many use cases.



BigchainDB is designed for use cases requiring high-throughput, lower latency and more storage capacity such as streamlining clearing and settlements in the financial sector, tracking supply chains such as the tracking of diamonds, spare parts or consumer goods, or tracking records of provenance for physical or digital assets.

Diagram below illustrates the main components in a four-node Big-chainDB 2.0 network, and how they communicate with each other.



BigchainDB 2.0 uses Tendermint for all networking and consensus. Each node has its own local MongoDB database, and all communication between nodes is done using Tendermint protocols.

***Full Decentralization and Byzantine Fault Tolerance***

Any node can fail and the rest of the network will continue to operate. In fact, up to one third of the nodes can fail1 in arbitrary ways, and the rest of the network will continue to

work, i.e. the non-faulty nodes will agree on how to proceed.

***Immutability***

BigchainDB uses several strategies to achieve practical immutability. The simplest one is that there are no BigchainDB-provided APIs to change or erase stored data.

Another strategy is that every node has a full copy of all the data in a standalone MongoDB database (i.e. there is no global MongoDB database). Even if one node gets corrupted or destroyed, the other nodes won't be affected and will still have a copy of all the data.

Another strategy is that all transactions are cryptographically signed. After a transaction is stored, changing its contents will change the signature, which

can be detected (unless the public key is also changed, but that should also be detectable because every block of transactions is signed by the a node, and the public keys of all the nodes are all known).

***Owner-Controlled Assets***

Like most blockchains, BigchainDB has a concept of owner-controlled assets. Only the owner (or owners) of an asset can transfer that asset. (The owners are the holders of a particular set of private keys.) Not even a node operator can transfer an asset.

In most blockchains, there's only one built-in asset (e.g. Bitcoin or Ether), but BigchainDB allows external users to create as many assets as they need. However, it's worth noting that a user can't create assets that appear to be created by someone else.

***High Transaction Rate***

Despite its strong guarantees, Tendermint provides exceptional performance. In benchmarks of 64 nodes distributed across 7 data-centers on 5 continents, on commodity cloud instances, Tendermint consensus can process thousands of transactions per second, with

commit latencies on the order of one to two seconds. Notably, performance of well over a thousand transactions per second is maintained even in harsh adversarial conditions, with validators crashing or broadcasting maliciously crafted votes.

***Low Latency and Fast Finality***

Tendermint-based networks (such as BigchainDB networks) take only a few seconds (or less) for a transaction to be included in a new committed block. Once that happens, there's no way it can be reverted or considered defunct in the future, because Tendermint doesn't do forking.

***Indexing & Querying Structured Data***

Each node in a BigchainDB 2.0 network has its own local MongoDB database. That means that each node operator has access to the full power of MongoDB for indexing and querying the stored data (transactions, assets, metadata and

blocks, all of which are JSON strings). Each node operator is free to decide how much of that power they expose to external users. One node operator might decide to index geospatial data and offer optimized geospatial queries via a REST API, whereas another node operator might decide to offer a GraphQL API.

By default, BigchainDB 2.0 creates some MongoDB indexes and the Big-chainDB HTTP API includes some endpoints for doing basic queries.

***Sybil Tolerance***

Some blockchain networks (such as Bitcoin) allow anyone to add their node to the network. That brings the concern that someone could add so many nodes that they effectively control the network: a Sybil attack. Bitcoin makes Sybil attacks unlikely by making them prohibitively expensive. In a BigchainDB network, the governing organization behind the network controls the member list, so Sybil attacks are not an issue.

# BigChain DB Demo

# Blockchain outlook

Several analysts expect blockchain will first prove its value in supply chain management, where Cisco and IBM are both working on applying the technology.

Still, industry experts project on average that the technology will take 5.9 years to gain widespread adoption, according to a Cowen survey.

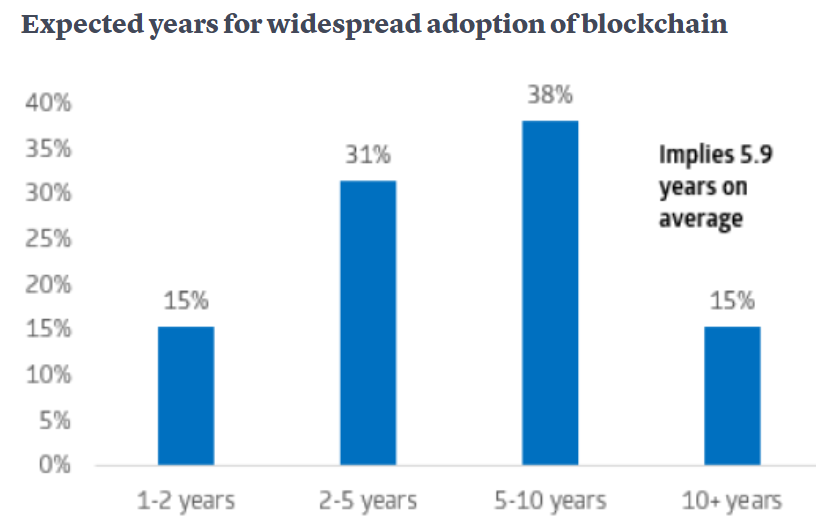
Many blockchain enthusiasts also say the technology will work best not as a new system within private enterprise, but in its pure form — an open-source, publicly accessible platform powered by a global network of users.

When it comes to proving the tangible benefits of the blockchain technology that many expect to change the world, it is still very early days and the earliest viable applications will likely be in supply chain management, rather than financial services, executives say.

In fact, Cisco stopped researching blockchain applications in financial services after 18 months because it will take a while for the many players in the complex markets to get up to speed, Anoop Nannra, head of the company's blockchain initiative, told CNBC in May. Instead, for the last year or so he said the team has been investigating blockchain applications in supply chain management.

Even still, their research indicates it will be roughly 10 years before blockchain really becomes prevalent in supply chain management. For financial services, Nannra predicts somewhere from five to 25 years.

"It's anyone's guess," Nannra said. "When you're trying to drive a mindset shift it's almost a generational thing."



<https://www.cnbc.com/2018/06/04/for-all-the-hype-blockchain-applications-are-still-years-even-decades-away.html>

<https://radiostud.io/eight-blockchain-platforms-comparison/>

<https://www.quora.com/What-kind-of-databases-are-used-by-Blockchain-platforms>

But if you are developing a blockchain-based application yourself, you'll probably need to provide support and maintenance for it. Research shows that the cost of support and maintenance for software is at least as great as the cost of developing it. This explains why software vendors charge so much for product support and maintenance. A good rule of thumb: If you develop an application that uses blockchains yourself, expect to pay at least as much throughout the application lifecycle as you do during development.

References

1. Judd Bagley (2018 , Sep . ). What is Blockchain Technology? A Step-by-Step Guide For Beginners. Blockgeeks. [Online]. Available: <https://blockgeeks.com/guides/what-is-blockchain-technology/>
2. Collin Thompson (2016 Oct.). How does the Blockchain Work? (Part 1). The Blockchain Review. [Online]. Available: <https://medium.com/blockchain-review/how-does-the-blockchain-work-for-dummies-explained-simply-9f94d386e093>
3. Collin Thompson (2016 Oct.). Blockchains & Distributed Ledger Technologies. BlockchainHub. [Online]. Available: <https://blockchainhub.net/blockchains-and-distributed-ledger-technologies-in-general/>

1. “This work was submitted as part of the MSDS 7349 Section 404 “Data and Network Security” course at Southern Methodist University as part of the Spring 2019 semester.

   Yumei Bennett (e-mail: yumei.bennett@oracle.com), Daniel Byrne (byrned@mail.smu.edu), Benjamin Wilke (e-mail: bwilke@mail.smu.edu), and Brian Coari (e-mail: bcoari@mail.smu.edu) are graduate students in the Data Science program at Southern Methodist University, PO Box 750100, Dallas, TX 75275 [↑](#footnote-ref-1)