CAPCO

Operational

Overview of Blockchain Platforms and Big Data

Guy R. Vishnia, Gareth W. Peters

Journal

THE CAPCO INSTITUTE JOURNAL OF FINANCIAL TRANSFORMATION

APEX 2016 AWARD WINNER

FINANCIAL TECHNOLOGY

Download the full version of The Journal available at CAPCO.COM/INSTITUTE

#44
11.2016

EMPOWERING THE [FINANCIAL] WORLD

Pushing the pace of Financial Technology, together we'll help our clients solve technology challenges for their business – whether it's capital markets in Mumbai or community banking in Macon.

We leverage knowledge and insights from our clients around the world:

20,000	clients in towns everywhere are becoming more efficient, modern and scalable.
27 billion	transactions processed help solve clients' challenges — big and small.
$9 trillion	moved across the globe in a single year empowers our clients' communities to build storefronts, homes and careers.
55,000	hearts and minds have joined forces to bring you greater capabilities in even the smallest places.

Empowering the Financial World

FISGLOBAL.COM

FIS

Journal

The Capco Institute Journal of Financial Transformation

Recipient of the Apex Award for Publication Excellence

WHAT ARE THE DRIVERS AND DISRUPTIONS THAT DETERMINE INNOVATION AND PROSPERITY?

CAN EVERY PROBLEM BE SOLVED WITH A QUESTION? YES, BUT NOT EVERY QUESTION HAS A SINGLE ANSWER.

The Munk School's Master of Global Affairs program is developing a new class of innovators and problem solvers tackling the world's most pressing challenges.

> Tailor-made, inter-disciplinary curriculum delivering the best of both an academic and a professional degree.

> Access to world-leading research in innovation, economic policy and global affairs.

> International internships with top-tier institutions, agencies and companies that ensure students gain essential global experience.

COME EXPLORE WITH US

BE A MASTER OF GLOBAL AFFAIRS

MUNKSCHOOL.UTORONTO.CA
MGA@UTORONTO.CA

MUNK SCHOOL OF GLOBAL AFFAIRS

UNIVERSITY OF TORONTO

Financial Technology

Operational

Transformational

Overview of Blockchain Platforms and Big Data

Guy R. Vishnia – Department of Computer Science, University College London
Gareth W. Peters – Department of Statistical Science, University College London

Abstract
An emerging trend in industry and research is the need to deal with increasing complexity and volume of data when performing analytics. This has led to the rise of the topic of "big data" systems within the financial technology sector, which we explore in this paper within the context of emerging blockchain technologies. Both big data and blockchain technologies have witnessed significant innovations, emerging new concepts, and use cases in a relatively short time. We discuss in this article these technologies in general, survey some projects and products that combine the two areas, and present a use case for these technologies in coming regulatory requirements for auditing and reporting under MiFID II.

BIG DATA

The term "big data" initially appeared around the mid- to late-1990s,[1] and has since come to represent a label given to areas that require analysis, processing, modeling, or analytics on huge datasets that cannot be processed by traditional database systems. One complication with such a generic label is that it has often been questioned exactly what constitutes big data, as any generic definition will by its very nature tend to be subjective and may vary between industries and disciplines. Rather, we prefer to think of big data as a set of new techniques and analytical practices that can be applied to large sets of data that result in insights and deeper understanding of the relevant topic.

Some authors and technology experts have attempted to define key attributes of big data. For instance, in 2001, in a research note by Doug Leany from the Meta Group, the term "3Vs" was introduced to describe the three basics of big data, which included:

- **Volume:** the quantity of the data.
- **Velocity:** speed of data generated, usually for real-time availability.
- **Variety:** the different sources of data.

Later, two other Vs were added:

- **Variability:** the consistency of the data
- **Veracity:** the quality of the data.

Currently big data trends continue to emerge in a variety of new fields, from marketing to social media and especially within the financial services industry. Most are used for data analytics and insights. There are tools and specific in-memory databases like, for example, KDB+[2], designed to process and analyze billions of records in real-time. In this context, we particularly focus on a specific, but important, topic within the financial services sector that is becoming much more relevant today as a result of MiFID II regulations, and making this a more of a big data problem to solve; namely, the capture and storage of trading events, trade reports, and transaction reports for five years, and be able to report this data back to the regulator on demand.

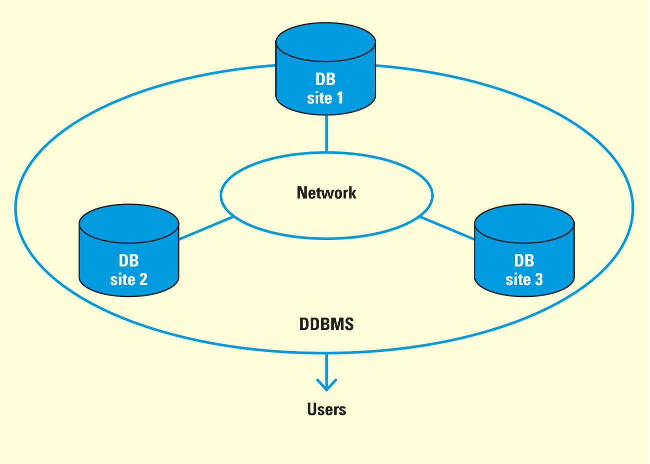


Figure 1 – Distributed database architecture

DISTRIBUTED DATABASE SYSTEMS

Distributed databases are a subset of the distributed systems field in computer science, where components on different physical locations interact via a network in order to achieve a common goal. A distributed database system is a collection of databases that adhere to the above, the databases each reside at physically separated locations and communicate with each other over a common network. Each node is managing its own set of data via DBMS independently of the other nodes, and all databases are managed by a distributed database management system (DDBMS), which is responsible for synchronizing between the nodes, ensuring all nodes have the full data and the integrity of the data, and loading balancing between the databases for data retrieval. To summarize such functionality, the DDBMS handles all databases as if they are all stored on a single location in a completely transparent way for the end user [Özsu and Valduriez (2011)].

BLOCKCHAIN

Discussions on blockchain technology are provided in Peters et al. (2015) and Peters and Panai (2015). In general, the terminology of this new field is still evolving, with many using the terms block chain (or blockchain), distributed ledger, and shared ledger interchangeably. Formal definitions are unlikely to satisfy all parties, but for the purposes of this article the key terms are as follows.[3] A blockchain is not a database but it can conceptually be thought of as acting like a database in the sense that it is a ledger that takes a number of records and puts them in a block (rather like collating them on to a single sheet of paper). Each block is then "chained" to the next block, using a cryptographic signature. This allows blockchains to be used

[1] For example, in 1997 it was referred to at the Institute of Electrical and Electronics Engineers (IEEE) conference on visualization.

[2] http://www.kx.com

[3] These will be discussed in greater detail below.

like a ledger, which can be shared and corroborated by anyone with the appropriate permissions. There are many ways to corroborate the accuracy of a ledger, but they are broadly known as consensus (the term "mining" is used for a variant of this process in the cryptocurrency Bitcoin). If participants in that process are preselected, the ledger is permissioned. If the process is open to everyone, the ledger is unpermissioned (see discussions below). The real novelty of blockchain technology is that it is more than just a database – it can also set rules about a transaction (business logic) that are tied to the transaction itself. This contrasts with conventional databases, in which rules are often set at the entire database level, or in the application, but not in the transaction.

A blockchain is not exactly a database

In terms of applications of blockchain technology, one could argue that we are still in the exploration phase. It is prudent to be cautious about claims that this technology, particularly in its "permissioned blockchain" form when being used in fields as diverse as banking, insurance, or accounting. In particular, it would be useful to explore exactly what advantages blockchains have compared to well-understood transaction recording technologies, such as databases. In fact, one could think of a blockchain as a technology for creating structured repositories of information, often termed a ledger in blockchain parlance. This can be strongly linked to similar understanding of a database, for instance, when talking about a ledger for financial assets. This, of course, could be represented in a database table, where in the simplest form each row represents one asset type owned by one particular entity. It has a number of attributes, one per column indicating information such as the owner's identifier, an identifier for the asset type, and the quantity of that asset.

We can think of blockchain in the simplest form as a technology that allows for such ledgers to be managed with multiple participants. In simple forms of blockchain technology, each participant will in some cases also run "nodes" in the blockchain network which hold a copy of the database. Their role is then to transmit transactions to other nodes in a peer-to-peer fashion. These transactions, from multiple participants, can occur in a blockchain typically without requiring the trust of all the participants. This brings us to considerations such as those discussed in Peters and Panai (2015), which considers data integrity and governance issues via the blockchain technology's ability to offer disintermediation.

So, we learn that a blockchain is a technology that allows us to utilize a database with multiple non-trusting participants, but does not necessarily require a trusted intermediary. Versions of the blockchain architecture, such as those developed in Bitcoin, remove the requirement for trusted intermediaries by extending the definition of a transaction, i.e., a modification to the database entry, to include a proof of authorization and proof of validity. This relates to the data integrity protocols discussed in Peters and Panai (2015), as several approaches can be adopted to achieve this in blockchain technologies. Upon this extended definition of a transaction it allows for the removal of intermediaries, since now transactions can be independently verified and processed by every node in the network that maintains a copy of the database.

To move beyond this simple description and understand further the differences between blockchain and standard database technologies, we first discuss the types and capabilities of modern databases. Depending on the nature of the data one is storing, there are five genres of databases [Redmond and Wilson (2012)]:

- **Relational databases**, such as SQL and variants, which are based on set theory and implemented as two-dimensional tables.
- **Key-value stores**, which store pairs of keys and values for fast retrieval.
- **Columnar databases**, which store data in columns, and can have more efficient representations of sparse tables compared to relational databases.
- **Document databases**.
- **Graph databases**, which model data as nodes and relationships.

Databases can be centralized (residing at a single site) or distributed over many sites and connected by a computer network. We will focus on the latter, given the closer proximity to the blockchain concept.

Distributed databases and blockchain

A number of emerging blockchain platforms are beginning to utilize connections between the blockchain ledger and some version of a distributed database for secure off-chain data storage. It is, therefore, useful to recall the difference between a blockchain and a distributed database.

A distributed database is a database in which portions of the database are stored in multiple physical locations and processing is distributed among multiple database nodes.

A centralized distributed database management system (DDBMS) integrates the data logically so that it can be managed as if it were all stored in the same location. The DDBMS synchronizes all the data periodically and ensures that updates and deletes performed on the data at one location will be automatically reflected in the data stored elsewhere.

Distributed databases can be homogenous or heterogeneous. In a homogenous distributed database system, all the physical locations have the same underlying hardware and run the same operating systems and database applications. In a heterogeneous distributed database, the hardware, operating systems, or database applications may be different at each of the locations.

The objective of a distributed database is to partition larger information retrieval and processing problems into smaller ones, in order to be able to solve them more efficiently. In such databases, a user does not, as a general rule, need to be aware of the database network topology or the distribution of data across the different nodes. It should also be noted that in a distributed database, the connected nodes need not be homogeneous, in terms of the data that they store.

Because of the design of these databases and the replication of data across different nodes, such a database has several advantages [Elmasri and Navathe (2014)]: (1) better reliability and availability, where localized faults do not make the system unavailable; (2) improved performance/ throughput; and (3) easier expansion.

In every distributed database, however, there is the issue of how modifications to the databases are propagated to the various nodes that should hold that data. The traditional approach is a "master-slave" relationship, where updates to a master database are then propagated to the various slaves. However, this means that the master database can become a bottleneck for performance. In multi-master replication[4] modifications can be made to any copy of the data, and then propagated to the others. There is a problem in this case also, when two copies of the data get modified by different write commands simultaneously.

A blockchain could be seen as a new type of distributed database that can help prevent such conflicts. In the same way that the Bitcoin network will reject a transaction where the Bitcoin balance to be transferred has already been "spent," a blockchain can extend the operation of distributed databases by rejecting transactions, such as delete a row, that have already been undertaken by a previous transaction (where a modification is a deletion, followed by the creation of a new row).

A second difference between blockchains and distributed databases lies in the ability to create self-enforcing contracts that will modify the blockchain's data. Many permissioned blockchains have a built-in virtual machine, such that one can execute pieces of computer code on the network. If this virtual machine is Turing-complete, this means that the machine can potentially solve a very large set of problems, which is very useful for executing more complex transactions on the network, possibly conditional on the state of certain off-chain variables.

The proliferation of databases as data stores has spawned considerations regarding data-related aspects, such as security, confidentiality, and integrity. We argue that discussions around these issues will be important for blockchain technologies too, if they are to be successful in a business enterprise setting. In the following section we discuss these security aspects in depth and comment on blockchain attributes with regard to them.

So far we can conclude that blockchains are a sensible technology when we wish to consider a set of databases that are to be shared by multiple participant contributors all of whom can modify the database directly, in an environment in which no trust is required between members of the network. Furthermore, we can see that blockchains further differentiate themselves from direct database solutions when we begin to consider transactions of multiple participants that interact or have dependencies on transactions of other blockchain member participants in non-trivial manner with each other.

BLOCKCHAIN TYPES

There are several types or "flavors" of blockchain, and in this section we will provide a short review of each.

Permissionless ledgers

A blockchain with no single owner, such as the one used in Bitcoin, is defined as unpermissioned or permissionless ledger. This type of ledger allows anyone to contribute to the chain, i.e., no one has the power to prevent others from adding data to the chain, and everyone holds the exact same copy of the ledger. The integrity of the chain is, therefore, determined by the consensus of all participants. However, this makes the ledger challenging to govern.

Permissioned ledgers

A blockchain with one or many owners, where a limited number of participants have the power to approve a new record added to the ledger, is a permissioned ledger. The governed structure of this type of ledger makes the consensus process much simpler and these ledgers are usually faster than unpermissioned ones.

Distributed ledgers

Distributed ledgers are like a distributed database and are spread across multiple sites and networks. Records are added continuously one after the other, and not by blocks. This type of ledger requires more trust in the validation of the operation over the ledger. The global financial transactions system, Ripple,[5] for example, uses a list of trusted validators in order to prevent transaction fraud.

[4] http://www.multichain.com/blog/2015/07/bitcoin-vs-blockchain-debate

[5] https://ripple.com/

Public blockchains

A public blockchain, generally considered to be "fully decentralized,"[6] is a blockchain that anyone in the world can read, add transactions to, and participate in the consensus process. These chains are secured by cryptoeconomics – the combination of economic incentives and cryptographic verification, where the influence on the consensus process is aligned to the size of economic resources a participant brings to the chain.

Shared ledgers

A shared ledger is a term coined by Richard Brown, formerly of IBM and now Chief Technology Officer of the Distributed Ledger Group, and typically refers to any database and application that is shared by an industry or private consortium, or that is open to the public. It is the most generic and catch-all term for this group of technologies. A shared ledger may use a distributed ledger or blockchain as its underlying database, but will often layer on permissions for different types of users. As such, "shared ledger" represents a spectrum of possible ledger or database designs that are permissioned at some level. An industry's shared ledger may have a limited number of fixed validators, who are trusted to maintain the ledger. The face that a number of trusted participants can validate transactions can offer significant benefits.

Fully private blockchains

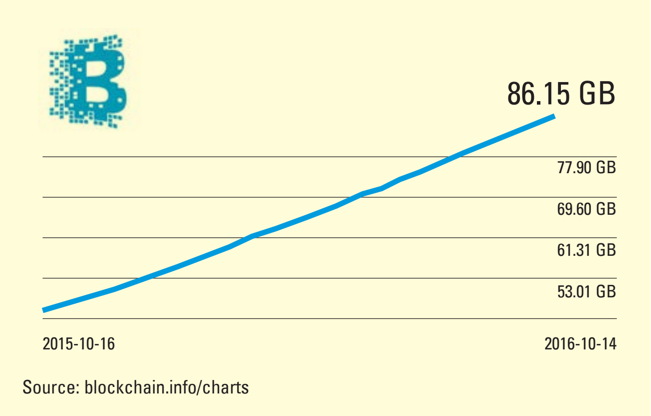
A ledger where all write permissions are controlled by one organization is considered a private ledger. Read permissions can be made public. These types of ledgers are useful for auditing purposes, as we see in our usage case presented later in the article.

Smart contracts

Smart contracts are contracts whose terms are recorded in a computer language instead of legal language. It can be designed to enact legal contracts or regulations. Smart contracts can be automatically executed by a computing system, such as a suitable distributed ledger system in response to changes in the ledger, in real time.

THE ROLE OF BLOCKCHAIN TECHNOLOGY IN BIG DATA

Big data in finance usually describes Petabytes (1 Petabyte is 1000 Terabytes) of trading data that are used for analytics generation. For example, the first level of the order book for all European markets [trades, bid and ask, see Gould et al. (2013)], will generate a capture file of about 5 GB a day. To put that in perspective, the size of the Bitcoin transaction database is a bit less then 85GB, with average block size of around 0.75MB (as of October 11, 2016) (Figure 2).[7] Blockchain as it currently stands is not built for large datasets and big fast data insertion and queries. Several solutions are now emerging, which use blockchain features for very large sets of data, for example by extending a distributed database functionality or by offloading data to an offline data storage.



Source: blockchain.info/charts

Figure 2 – Blockchain size

Big data blockchain solutions

Certain problems require the save and storage of large amount of data; from real time marketing analysis online, to trading monitoring and indication systems. Usually, databases, either classic or distributed, and data warehouses will be used for this task, as they provide the capacity, latency, and scalability needed from a big data solution. As shown above, a blockchain is not a database, but it carries with it a data storage and some interesting characteristic such as immutability, which can be a necessity for some applications. In addition to immutability, there are all the SQL language features that are part of the traditional database systems, the ease and fast insert operations, and a timely select function with different filters. A usable blockchain big data solution will need to provide all these basic properties as a prerequisite.

There are several approaches for this problem, we will review the main ones with a representative for each of the options.

Blockchain on top of distributed database

One interesting and innovative idea is using an existing distributed database technology with added blockchain functionality. The distributed database is by nature an excellent big data storage. It can scale horizontality and increase capacity and throughput by adding

[6] https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains
[7] http://www.blockchain.info

shards (or nodes), it has a rich query language, like T-SQL (Transact-SQL) or NoSQL, and a built-in permissioning management. Adding blockchain features seems very natural as both share a distributed architecture, and blockchain can add immutability, the option to have a decentralized control mechanism, and a common well-known way of handling trail of digital assets. We will demonstrate a potential usage for this technology in our architecture for event capture data storage. The main features will include:

- High throughput and capacity
- Low latency
- Permissioning mechanism
- Querying language
- Decentralized system
- Immutability

BigchainDB

BigchainDB[8] is a big data solution that takes a distributed database and adds blockchain properties on top of it. It features a full NoSQL query language and aims for a performance of 1 million writes per second, which should meet known financial systems requirements. Classic blockchain performance is not in the same bracket as this type of database, as it can handle only a few transactions per second, and confirmations can take up to ten minutes. On a modern distributed database, capacity and throughput are a given with the scalability of the system. McConaghy et al. (2016) describe in detail BigchainDB, its performance, and case studies.

Modern applications collect huge amount of data from users in real time. Think about Amazon, Facebook,[9] Google and the like, which collect, analyze, and monitor huge amounts of data within minutes. Financial companies also save large amount of datapoints for trading analyses and reporting. This will increase immensely in the near future with the MiFID II regulation. Distributed databases store petabytes (1,000,000 GB) of data and can be easily extended. The Bitcoin blockchain, on the other hand, currently stores only 85GB of data, which for some people in the community seems too big. BigchainDB uses RethingDB[10] as the base for its distributed database.

Blockchain and offchain distributed hash tables

A distributed hash table (DHT) provides a look-up service similar to a key-value hash table, but does so in a decentralized distributed manner. The key-value pair can be stored in any participating node and the key-value mapping is then maintained by all nodes. This allows a DHT to scale on a very large number of nodes. DHT, which was in part originally motivated by peer-to-peer (P2P) systems, can be used to build complex infrastructures like distributed file systems, P2P file sharing, and content distribution. There are three key properties of a DHT:[11]

- **Autonomy and decentralization:** all the nodes construct the system without any centralized governance.
- **Fault tolerance:** the system will continue to work as usual when nodes are added/removed or suffer a from a failure.
- **Scalability:** the system can scale up to millions of nodes and continue to work as normal.

Enigma

Enigma[12] is a new decentralized computation cloud platform from MIT with guaranteed privacy [Zyskind et al. (2015)]. Enigma offers privacy by distribution of data between nodes, where no node has access to the data in full. Computation is run on the nodes without the need to reveal the full information to other nodes. Since data is not replicated on each node, it gives the platform the ability to scale horizontally.

Some of the main features that Enigma offers users as a blockchain platform are privacy and scalability.

In terms of privacy, this is achieved in Enigma through its use of a secure multi-party computation model. In this framework, queries are done in a distributed way without a governing trusted third-party being required. Furthermore, the computation is split between different nodes and no single node has access to the other nodes' data. Each node only sees part of the data that has no value or meaning on its own.

With respect to scalability, this is achieved by the fact that data is not being replicated to every node in the network. The computation is being done on a small subset of nodes that hold different parts of the data. This enables Enigma to run more demanding computations and require significantly less storage requirements.

The off-chain nodes feature allows Enigma to store large sets of data, and in a way constructs a distributed database in which each mode has its own distinct view on the data.

One of the possible applications for Enigma is as a distributed personal data store, which fits our usage case for personal trader data information.[13]

[8] https://www.bigchaindb.com

[9] https://code.facebook.com/posts/229861827208629/scaling-the-facebook-data-warehouse-to-300-pb/

[10] https://www.rethinkdb.com

[11] http://www.cs.princeton.edu/courses/archive/spr11/cos461/docs/lec22-dhts.pdf.

[12] http://enigma.media.mit.edu/

[13] See discussion in Section 8.7 of the Enigma document, where it states: "Store and share data with third parties while maintaining control and ownership. Set specific policies for each service with private contracts. Identity is truly protected since the decision to share data is always reversible - services have no access to raw data, all they can do is run secure computation."

Enigma design

The framework of Enigma offloads private and intensive computation work from an existing blockchain to an off-chain network. It also provides a scalable Turing complete scripting language for handling private contracts (with private information). An Interpreter will break down the execution of this private contract, which in addition to privacy also improves the run time of the code.

Through the use of off-chain processing and storage it is possible for Enigma to solve data capacity problems; it enforces the privacy of computation by allowing each node to execute code without leaking data to the other nodes and solves scalability problems when heavy computation is needed on the chain. Enigma performs the heavy computation on the off-chain and broadcast the results to the blockchain.

The off-chain storage creates a distributed database, where every node has a distinct view of the data. It is possible to store large public data in the off-chain and link it to the blockchain. The distribution is based on a Kademlia DHT protocol, which was modified for Enigma.

In Enigma, the blockchain acts as an interface between the off-chain DHT architecture that stores references to data in a decentralized manner and the actual data of interest, which is first encrypted on the client side before storage and access protocols are enacted in the blockchain or on off-chain distributed data-bases. However, since the Enigma blockchain does not replicate the data over all nodes in the network, instead only requiring a small subset of such nodes to perform each computation over different parts of the data, it achieves efficiency gains. The off-chain storage of data occurs with off-chain nodes constructing a distributed database.

Zyskind et al. (2015) explain how Enigma offers a combination of off-chain storage and blockchain storage for data. In this structure, each node will have a specific unique view of what they term “shares” in the total data (a portion of the total data) as well as the encrypted data, where the share is set up in such a manner as to guarantee privacy preservation and fault tolerance. In addition, this architecture also allows for large public data storage that may be linked to the blockchain and unencrypted for all participants to access. The manner that this is achieved in a network architecture is known as Kademlia DHT [Maymounkov and Mazieres (2002)] with enhancements for the Enigma use case.

USAGE CASE – EVENT CAPTURE ARCHITECTURE

In recent years, financial firms have seen enhanced scrutiny and oversight by the regulators that are stepping up their demands for trade and transaction reports, transparency and best execution proof, order trail, and auditing data. Trading venues and brokers are obligated to provide reports with many more fields, capture a lot more events, and store the collected data for a period of five years in an accessible secure manner. This is already producing massive datasets and the increasing requirements has led many analysts to suggest that such data requirements for storage of trade activity is likely to continue to grow. The security and data integrity of these records is also a critical feature to be considered [see discussion on these matters in the context of blockchain in Peters and Vishnia (2016)].

The classic way to achieve the above reporting and storage requirements would be to store the data in a database; relational or a NoSQL database [Tauro at al. (2012)] like MongoDB [Chodorow (2013)] via a big data warehousing solution. We will present here an architecture for storing the event capture data in a secure immutable way using blockchain technology. This new architecture will provide the regulator with easy access, on demand data queries without the risk of data being tempered or lost, and for the reporting entities a common simple manner for storing the data and replying to regulator queries.

Reportable events and data points

Trading firms and venues will need to provide to the regulator, on demand, under the European Securities and Market Authority Regulations, all relevant event capture data. This data can be an order event like Ack, fill, cancel, etc., market data points like bid/ask for best execution proofing, algorithmic trading decisions, order initiator, and

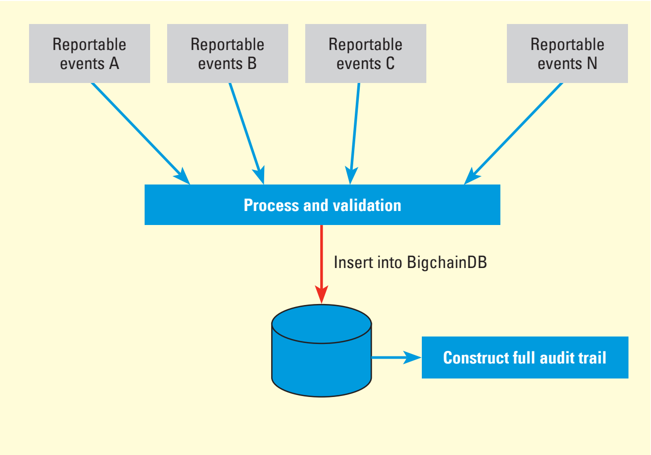


Figure 3 – Event capture storage architecture with BigchainDB

many more trading life cycle events (at time of writing a full event list is yet to be confirmed). The data also needs to be time stamped and synced [RTS 25, Article 4; ESMA (2015)] with precision of at least one millisecond. This is a hard demand to follow when aggregating trading data from different systems. Of course, each trading entity will have different amounts of data to collect and store, but even for a medium-sized company this sums up to a very large set of data. Needless to say, all this data should be stored safely and securely and yet be accessible rapidly on demand to select groups, such as the regulator and the event capturing entity.

Data storage

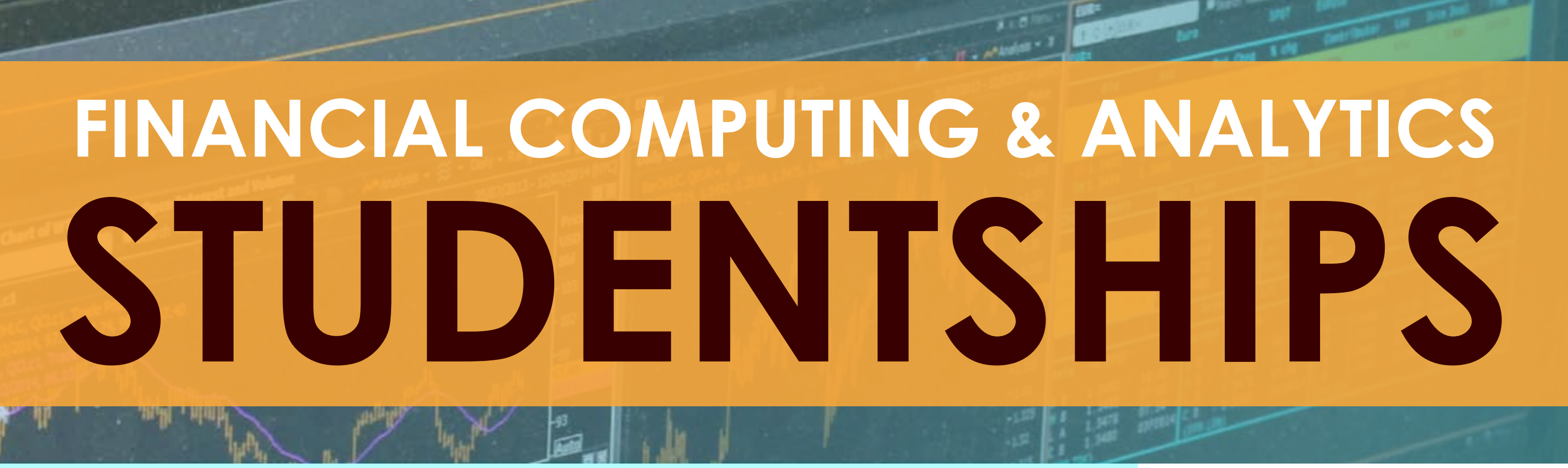
We suggest, for example, using a blockchain database such as BigchainDB in order to store all event capture data in distributed block chain database. Trading events occur very fast, with high throughputs in random times of the day. A tradition distributed DB will be sufficient to handle these events and provide rich and easy-to-use query capabilities. However, for the regulations we also need to make sure the data is immutable, and to be maintained by different market participants. We also want to have a decentralized control with a read-only user (the regulator). A "classic" blockchain will not be able to cope with the amount of data, size, and throughputs, that this challenge presents. However, a combination of blockchain immutability property and the decentralized nature of a distributed database give us a clean solution to meet the demands of the regulators.

CONCLUSION

In this article we present the case of usage for blockchain technologies with big data. We think that in the coming future a big data solution will have to cater for blockchain features in order to be considered complete in its offering. Several possible implementations are already out in the market by smaller FinTech companies, but we think in the long term bulge bracket database companies will start offering blockchain features integrated within their products, whether as part of the core product or as an add on.

REFERENCES

- Back, A., M. Corallo, L. Dashjr, M. Friedenbach, G. Maxwell, A. Miller, A. Poelstra, J. Timón, and P. Wuille, 2014, "Enabling blockchain innovations with pegged sidechains," white paper, Blockstream
- Chodorow, K., 2013, "MongoDB: the definitive guide," O'Reilly Media, Inc.
- Elmasri, R., and S. B. Navathe, 2014, Fundamentals of database systems, Pearson
- European Securities and Market Authority, 2015, "Regulatory technical and implementing standards – Annex I," September 28, ESMA/2015/1464
- Maymounkov, P., and D. Mazieres, 2002, "Kademlia: a peer-to-peer information system based on the XOR metric," International Workshop on Peer-to-Peer Systems 2429, 53-65
- McConaghy, T., R. Marques, A. Müller, D. De Jonghe, T. McConaghy, G. McMullen, R. Henderson, S. Bellemare, and A. Granzotto, 2016, "BigchainDB: a scalable blockchain database," white paper, BigChainDB
- Özsu, M. T., and P. Valduriez, 2011, Principles of distributed database systems, Springer Science & Business Media
- Peters, G. W., and E. Panayi, 2015, "Understanding modern banking ledgers through blockchain technologies: future of transaction processing and smart contracts on the internet of money," working paper
- Peters, G. W., and G. Vishnia, 2016, "Blockchain architectures for electronic exchange reporting requirements: EMIR, Dodd Frank, MiFID I/II, MiFIR, REMIT, Reg NMS and T2S, Dodd Frank, MiFID I/II, MiFIR, REMIT." Available at SSRN: https://ssrn.com/abstract=2832604
- Peters, G. W., E. Panayi, and A. Chapelle, 2015, "Trends in crypto-currencies and blockchain technologies: a monetary theory and regulation perspective," working paper
- Redmond, E., and J. R. Wilson, 2012, Seven databases in seven weeks: a guide to modern databases and the NoSQL movement, Pragmatic Bookshelf
- Tauro, C. J., S. Aravindh, and A. B. Shreeharsha, 2012, "Comparative study of the new generation, agile, scalable, high performance NOSQL databases," International Journal of Computer Applications 48:20, 1-4
- Zyskind, G., O. Nathan, and A. Pentland, 2015, "Enigma: decentralized computation platform with guaranteed privacy." Available at: http://web.media.mit.edu/~guyzys/data/enigma_full.pdf



FINANCIAL COMPUTING & ANALYTICS STUDENTSHIPS

Four-Year Masters & PhD for Final Year Undergraduates and Masters Students

As leading banks and funds become more scientific, the demand for excellent PhD students in **computer science, mathematics, statistics, economics, finance** and **physics** is soaring.

In the first major collaboration between the financial services industry and academia, **University College London, London School of Economics,** and **Imperial College London** have established a national PhD training centre in Financial Computing & Analytics with £8m backing from the UK Government and support from twenty leading financial institutions. The Centre covers financial IT, computational finance, financial engineering and business analytics.

The PhD programme is four years with each student following a masters programme in the first year. During years two to four students work on applied research, with support from industry advisors. Financial computing and analytics encompasses a wide range of research areas including mathematical modeling in finance, computational finance, financial IT, quantitative risk management and financial engineering. PhD research areas include stochastic processes, quantitative risk models, financial econometrics, software engineering for financial applications, computational statistics and machine learning, network, high performance computing and statistical signal processing.

The PhD Centre can provide full or fees-only scholarships for UK/EU students, and will endeavour to assist non-UK students in obtaining financial support.

UCL

LSE

Imperial College London

financialcomputing.org

INDUSTRY PARTNERS

Financial:

Barclays
Bank of America
Bank of England
BNP Paribas
Citi
Credit Suisse
Deutsche Bank
HSBC
LloydsTSB
Merrill Lynch
Morgan Stanley
Nomura
RBS
Thomson Reuters
UBS

Analytics:

BUPA
dunnhumby
SAS
Tesco

MORE INFORMATION

Prof. Philip Treleaven
Centre Director
p.treleaven@ucl.ac.uk

Yonita Carter
Centre Manager
y.carter@ucl.ac.uk

+44 20 7679 0359

Layout, production and coordination: Cypres – Daniel Brandt, Kris Van de Vijver and Pieter Vereertbrugghen

De Kleetlaan 6, B-1831 Machelen

Imperial College Business School | Centre for Global Finance and Technology

Centre for Global Finance and Technology

The Centre for Global Finance and Technology at Imperial College Business School will serve as a hub for multidisciplinary research, business education and global outreach, bringing together leading academics to investigate the impact of technology on finance, business and society.

This interdisciplinary, quantitative research will then feed into new courses and executive education programmes at the Business School and help foster a new generation of fintech experts as well as re-educate existing talent in new financial technologies.

The Centre will also work on providing intellectual guidance to key policymakers and regulators.

"I look forward to the ground-breaking research we will undertake at this new centre, and the challenges and opportunities posed by this new area of research."
– Andrei Kirilenko, Director of the Centre for Global Finance and Technology

Find out more here:
imperial.ac.uk/business-school/research/finance/centre-for-global-finance-and-technology/

CAPCO

BANGALORE
BRATISLAVA
BRUSSELS
CHICAGO
DALLAS
DÜSSELDORF
EDINBURGH
FRANKFURT
GENEVA
HONG KONG
HOUSTON
KUALA LUMPUR
LONDON
NEW YORK
ORLANDO
PARIS
SINGAPORE
TORONTO
VIENNA
ZÜRICH