ENABLING DATA DEMOCRACY IN SUPPLY CHAIN USING BLOCKCHAIN AND IOT

Deepika Sachdev

Vice President Delivery, Knowesis
UPES, Dehradun, India

ABSTRACT

Data Analytics in today's Demand Driven Supply chain Management (SCM) involving trusted as well as untrusted fulfilment partners will see exponential improvements with large volumes of reliable, trusted data generated automatically through IoT and stored in immutable format in the Blockchain. This paper examines the challenges which Supply Chain Partners in a Highly Digitized ecosystem have in getting access to good quality reliable data. It examines Use Cases and benefits of current IoT and Blockchain integrated frameworks for SCM. It also explores limitations of the traditional IoT and Blockchain integration patterns and presents an architectural framework for a permissioned Blockchain that can be used to overcome these limitations.

Key words: Blockchain; Internet of things; SCM; Privacy; DPB; Scalability

Cite this Article: Deepika Sachdev, Enabling Data Democracy in Supply Chain using Blockchain and IoT, *Journal of Management*, 6(1), 2019, pp. 66–83.

<http://www.iaeme.com/JOM/issues.asp?JType=JOM&VType=6&IType=1>

1. INTRODUCTION

1.1. Supply Chain Management (SCM)

Supply Chain Management refers to the Business and Technical Process Integration between direct and indirect stakeholders including consumer, retailer, wholesaler, distributor manufacturer and end supplier. We are witnessing an increased volatility in SCM due to shortened product life cycles and uncertainty due to competition in the global economy. Due to the risk involved with slow moving pipelines, organizations are looking at restructuring their Supply Chain operations. In this era of time-based competition organizations are looking at innovative ways to ensure on-time delivery and offering custom made products tailored to the long tail of the consumer market. In fact, we could be moving from a supply chain, to what Michael Casey, Senior Advisor to MIT's Digital Currency Initiative, calls a demand chain. The demand chain is a customer driven model with fast turnarounds, lower costs, and greater efficiency. In this model, producers respond to the demand of customers, and products are collectively produced from an amalgam of suppliers

The Highly Digitized Demand Chains have an enhanced need of IoT & Blockchain since:

- Environments with links to open marketplaces, social components & automation will be vulnerable to hacking, fraud & manipulation
- Development in Robotics, 3D manufacturing, unmanned / social fulfilment process will demand provenance & require immutable proof on hand-offs between Supply chain actors
- Distributed data access & digital contracts will require enhanced Access Control Mechanics

1.1.1. DPB (Data Science, Predictive Analysis & Big Data) for SCM

The term DPB referring to Data Science, Predictive Analytics, and Big Data has been coined by Waller et. Al [19] in *Journal of Business Logistics*. It refers to the scientific application of qualitative as well as quantitative methods on SCM data to enable faster information access as well predictive analysis to forecast trends for consumer demand.

Based on a detailed study conducted in 2012, McAfee and Brynjolfsson [14] conclude that objective and financial results of an organization increase significantly as the organization invests in data analytics to investigate its outcomes. Companies which were in the top 1/3rd of their industry for data-driven analysis showed 6% higher profitability and 5% higher productivity as compared to the market [19].

Figure 1 Big Data Sources in SCM

1.1.2. Data Industrialisation for SCM

As per a study by Chase in 2016 [2], all aspects of business including supply chain has been impacted by the Digital Revolution. Data generated by smart unmanned delivery drones/vehicles, by robot managed 3D manufacturing hubs located close to markets, by biometric authentication systems, by digital contracts & provenance mechanisms will be massive in volume. Machine learning & predictive analytics will play a massive role in optimizing these supply chains, preventing fraud and in enhancing consumer choice & consumer service.

A study on potential benefits and challenges with IoT [7] indicates that SCM will be significantly impacted by the digital era where almost all devices will be interconnected via the internet. Additionally, as part of Industry 4.0 study by PwC on 2000 respondents, more than one third of the respondents have confirmed that digitization in SCM has started in there organization to enable Data Industrialisation.

1.1.3. Data Transparency in SCM

Bilateral information transparency between partners is a key reason of concern for Supply Chain Managers impacting organization effectiveness. Key reasons for this are social factors such as Customer Domination as well as technology related factors such as lack of a trusted information sharing system. The same has been highlighted in “The Vantage Point Syndrome” [11] where it is suggested that in many cases organizations behave that they are the sole owners of end to end supply system information, hence suppliers and sometimes customers also follow there directions.

A very interesting metaphor of the Transparency in Supply Chain between partners by Lamming et. al[11] is presented which uses the analogy of light shining through a mineral to explain various frameworks of Information Sharing across partners:

Table 1 Supply Chain between partners by Lamming

	Opaque	Translucent	Transparent
Geological Use Case of minerals	No Light can pass through the surface	Light can pass through the mineral but gets distorted during transmission	Light can pass through easily across the surface
SCM Use Case	Not even basic information passes through the Partners. Such relationships are normally observed when one of the Partners is extremely dominant and sees no Value in Data Sharing.	Only outline information is shared between partners similar to Black Box level of sharing. This is useful for Operational enhancement to ensure reductions in Supply Chain wastages. However, this shall not lead to benefits of knowledge sharing and enrichment between partners	Selected and justified information is shared across partners. This leads to collaborative abilities and knowledge sharing.

1.1.4. Information Sharing Ecosystem in SCM

There are two broad categories of Industry Use Cases for Supply Chain Information Sharing:

1.2. Explicit Information Share with Supply Chain Partners

This refers to information share from Order to Delivery with the members of the Supply Chain. Increased Order lead times from Customer to Retailer to Wholesaler to Manufacturer results in loss of opportunity loss and customer satisfaction impact in a Supply Chain. A prime reason could be that all key players are on their disparate centralized ordering systems and information sharing across members is constrained due to geographical, technological or communication reasons.

1.3. Tacit Information Share with Supply Chain Community

With the advent of Digitization, there is multi fold increase in the amount of information that is generated by Supply Chain members. One of the pioneers in Data Driven SCM, Amazon performs predictive analytical models running on historical transactions as well as real time customer clicks. These models proactively predict customer purchase patterns for “Anticipatory Shipping” to reduce order to delivery timeframes.

The plethora of information in SCM can be used to tune Predictive Models to a high degree of accuracy, if the information is shared across enterprises at multiple tiers of the SCM hierarchy. However in today's world, information is not shared across Trusted members due to Privacy concerns arising due to lack of Trust in the information sharing process. Partners do not have the confidence that their information will be used by rightful users only. In addition, the huge volumes of data make it difficult for information to be shared. Partners are incentivized to share information as a part of Data Monetization by applying varying Tariff Rates based on the timeliness, accuracy and content of the shared data.

1.4. Internet of Things

The Internet of Things (IoT) is a networking paradigm that is relevant to all physical entities around us that are digitally enabled and can exchange data over the internet. The promise of IoT to SCM is to create a transparent ecosystem by enabling a wide variety of physical and virtual devices including infrastructure such as vehicles, fridges, Burglar alarms as well as constrained devices such as sensors, thermostats, valves and switches to communicate with each other. For e.g. while transporting vegetables on a truck, sensors from the IoT device can transmit the location, quality, size and even temperature of the goods being transferred to provide monitoring of transport conditions as well as accurately predict time of arrival.

Some of the key characteristics of an IoT based solution are:

1.4.1. Veracity in Data Content

IoT is enabled by sensors embedded in physical devices that detect changes in state and communicate the same over standard Communication protocols to the external world. Individual object level interaction contributes to collective intelligence in the complete SCM ecosystem. Some standard protocols used for IoT communication are MQTT (Device to Server), XMPP (Device to People) and AQMP (Server to Server). The challenge in Machine to Machine communication is that there is no single standardized unification protocol, hence the need for a robust middleware in the framework that can support the vast variety of data formats.

1.4.2. Volume

With the enormous potential of IoT enablement, the number of IoT enabled devices is expected to grow exponentially in the coming years. As per 2015 Gartner report predicted that every day 5.5 million new devices will be connected every day and by 2016 there will be 6.4 billion connected things. The number of connected devices is expected to reach 20.8 Billion by 2020. Hence IoT frameworks support horizontal as well as vertical scaling.

1.4.3. Security

The IoT devices have a web interface built in to allow users to communicate, however the same interface makes the device very vulnerable to security threats also. Cross Site Scripting, SQL Injection and Weak User credentials are some of the possible reasons for attacks. In addition network vulnerabilities including Denial of Service attacks and Open Ports also result in attacks from external networks. In addition since IoT data is traditionally stored in the cloud the probability of Data Leaks is significantly higher when the data is not properly encrypted or insufficient User Authorization is performed. All the above lead to Privacy Issues, Data Leakage and Third Party attacks, especially in an SCM environment where there can be Data exchange across untrusted partners.

1.5. Blockchain – The Decentralized Protocol

The Blockchain protocol created by Satoshi Nakamoto [16] can be defined as a distributed database for storing time series transactions. Unlike a traditional database it is append-only with very restricted or no update or delete capability. Instead of storing individual transactions, it stores blocks of transactions which are cryptographically chained together. Blockchain is a decentralized protocol aimed to solve the Byzantine Generals problem. Interestingly, even though Blockchain was initiated by Satoshi for a Finance Use Case its applications are now spreading to all sectors. Due to its promise of True Decentralization and removal of the Middle Man, SCM is identified as one of the “Killer Applications” for Blockchain.

Figure 2 Simplified Blockchain View

Some key Blockchain characteristics and their significance for a Supply Chain system would be as follows:

Table 2 Supply Chain system

Definition	Significance
Transparent Public Lock Box	It is a secure means to transfer money, however it is not private
A digital ledger for recording chronologically and publicly transactions made in a Blockchain	There is Transparency in transactions stored so that they can be accessible to all for reconciliation
The Blockchain is defined as a gossip protocol where any ledger state modification is broadcast to all participants	There is a significant drag in the ability of all the Global Transactions to be replicated on the Blockchain hence Real Time is not possible with current protocols
The Blockchain is a decentralized asset database powered by Smart Contracts	Blockchain has the ability to automatically execute scripts for automating rule set execution without intervention of a centralized authority.

In 2016 by Abeyratne et. al[1] it was highlighted that a low hanging fruit for logistics and SCM improvement is transparency between direct as well as indirect stakeholders of the SCM ecosystem. At the same time in 2016 Tapscott & Tapscott highlighted in “Blockchain revolution: how the technology behind bitcoin is changing money, business and the world” that Blockchain promises to create transparency, by ensuring that all members of the network are able to access the data ensuring that there is a single point of truth. Keeping the two statements together, Blockchain has enormous potential for supply chain transformation in the way

products are ordered, produced, transported, delivered and consumed. It has truly been said that Blockchain will be the next holy grail for the enterprise.

2. CURRENT INDUSTRY USE CASES FOR SCM, IOT & BLOCKCHAIN INTEGRATION

Over the last few years, the Supply Chain Industry has become increasingly aware of the benefits of the increased power of integration between IoT and Blockchain. We see multiple Use Cases where an integrated architecture is used to overcome the issues of transparency, trust and Provenance.

2.1. Walmart

In December 2017, IBM, Walmart and JD.com (one of China's largest retailers) have started an initiative for food tracking in Supply Chain using Blockchain. This effort is aimed at bringing safer food supply to China, and an extension of the work initiated by Walmart and IBM in August 2017 in the US. It is important to note that Walmart had done earlier trials involving barcodes technology, however these required a centralized server and database and hence were not reliable. The consortium is tracking pork from Chinese stores and produce from Latin America. The provenance data shall be used to track down the origins of the food reliably in the advent of a foodborne disease outbreak.

The ultimate objective of this consortium is to use this data to reduce food wastage. As per FAO of the United Nations in up to one third of food is perished in transit in today's world. Maintaining optimum temperature conditions can significantly reduce damage of fresh vegetables. IoT can provide real time information monitoring status of temperature conditions of the food as well as can also be programmed to adjust the temperature to significantly reduce these wastes.

2.2. Alibaba

To combat food fraud using Blockchain technology, a consortium of Alibaba, AusPost, Blackmores, and PwC has been created. Supported by Australia Post and Blackmores who will provide information from their existing Supply Chain, The Food Trust Framework will help improve integrity and traceability on food supplies from Australia to China. Historically, Alibaba has faced lawsuits where they were sued for illegal watches sold by partners from their site. To avoid such legal issues resulting in defamation and money losses in future, Alibaba's long term vision is to extend Blockchain to ensure Provenance of retail items.

In a food based disease outbreak, the suppliers have a hard time identifying the Provenance of the food [18]. In the current eco system it can take up to weeks to identify the origins of the food and that results in a slower capture of the outbreak. With the Food Trust Framework in place, Suppliers can be expected to track the provenance and in turn ban supplies from the source in a matter of hours and not weeks.

2.3. DeBeers

In December 2017, De Beers has made investments in a Blockchain-based diamond tracking platform that it hopes will augment supply chain transparency and diamond traceability to avoid conflict diamonds. De Beers' Forevermark diamonds which come with ID inscription that can be searched on a special website, as well as various interventions that have emerged recently to make use of the Blockchain technology to create tamper-proof supply chains. Blockchain creates an immutable digital register for recording all transactions of the diamond in the value chain.

2.4. Maersk

Maersk and IBM started their Blockchain trailers for tracking Supply Chain freight using Blockchain in the summer of 2016. The first venture was to track a container of flowers that sailed from Mombasa, Kenya to Rotterdam, Netherlands. Since then they have partnered with Tetra Pak, Dupont, Dow Chemicals and various ports and custom office to test the delivery of supply chain tracking through Blockchain using tamper proof digital records stored in an immutable ledger. The objective is to increase efficiency and provenance of goods movement in the Supply Chain.

2.5. Kontrol Energy

In January 2018, Kontrol Energy announced a tie-up with IOTA called bIOTAsphere to solve the greenhouse gas emission challenges faced by multinational corporations. Blockchain will be used to immutably and irrevocably store the Carbon credits and offsets of organizations. After the Paris Commission, many organizations have aspired to become Carbon neutral by 2050. An organization that has carbon credits, can sell it to other organizations through a Carbon offset. Effectively this will create a Carbon Marketplace that will be used to transparently and fairly monitor as well transfer Carbon Credits by a decentralized organization.

3. KNOWN BENEFITS & CHALLENGES OF BLOCKCHAIN IOT INTEGRATION FOR SCM

Figure 3 IoT and Blockchain in SCM

As discussed above, SCM needs faster and collaborative information sharing for enhanced Analytics. IoT helps track changes in device state as well as geospatial attributes by using RFID, GPS tags, barcodes, sensors and chips. Hence use of IoT in SCM is bringing increased profits by reduction of expired products as well as on time deliveries for suppliers. As per Li et. al[12] faster growth and competitive advantage will be achieved by SCM's that adjust and upgrade to use IoT technology.

The first of Block chain's direct benefits to SCM is that device state tracking data is immutable and decentralized, resulting in increased trust and reduced communication costs [9]. Additionally Blockchain also provides a possible solution to identity management in a decentralized manner.

Hence the world has started seeing benefits of the IoT Blockchain Integration framework for SCM today with reduced Order to Market time, as well as more efficient Predictive Models

due to availability of larger datasets. Below section explains issues in legacy SCM Analytics, benefits derived from IoT Blockchain integration and also known gaps in the integrated system.

3.1. Data Transparency with Trusted and Untrusted Partners with Access Control

3.1.1. Issue

In a supply chain ecosystem, information sharing between members is not regularized leading to the bullwhip effect [3] which states that as one moves upstream in the supply chain the demand signals get amplified across the retailer, manufacturer, supplier and commodity providers. The key reason for this is that data is in siloes in private cloud databases owned by Private Enterprises. Enterprises which own the data, protect the Data for Privacy, Regulatory as well as security reasons. A World Economic Forum¹⁴ report in 2012 highlighted that organizations are hesitant to share data for risk of violating user trust and confidentiality. Privacy and security concerns must be addressed before personal data can be more widely shared. But even for non-personal data, enterprises have not opened the gates.

It is often seen that centralized platforms initially share data with partners to make their services more valuable by making data accessible to partners. However, as platforms move up the adoption S-curve, their power over users and 3rd parties grows and they develop a tendency to reduce data openness.

For Analytical Models in SCM to be effective they need to be exposed to data from multiple geographies, economic sectors as well as cultural backgrounds. For example in Air Quality Monitors, environmental data sensors track the air quality in a specific region, and the same is compared with results from other geographies. Currently, the data from those sensors is segmented, controlled, and even manipulated, by governments and other organizations. Putting that data on a Blockchain would present us with a complete and transparent picture of the air quality around the world, not just snapshots from different regions. Even more critical for Data sharing in enterprises is in Health Care, where the DNA, genome and medicinal response patterns of Clinical Patients can be lifesaving and can even be used to do predictive analysis.

3.1.2. Benefits

With the increasing adoption of IoT devices there is an exponential rise in volume, and veracity of data. Data is now captured in every step of the Supply Chain from Demand to Supply and available to the enterprise for analysis.

Blockchain solves a missing piece of the supply-chain automation puzzle that has long hindered end-to-end visibility providing a trusted way for competitors to share data. It offers a decentralized ecosystem where information can be shared with partners according to McDermott [4].

3.1.3. Gaps

- **Security Concerns:** SCM transactions are private and the User wishes to keep his identity known only to Trusted partners. However, the Bitcoin Blockchain is a public ledger whose transactions can be viewed by all members. Bitcoin is called to be a pseudo anonymous system since even though the address of the owner is used in the transaction, the true identity of the user can be identified through “taint analysis”. Associations between addresses in a Bitcoin transaction can be evaluated through taint analysis.
- **Role Based Access Control:** In a permissioned environment, where new partners keep getting added in the Supply Chain, it should be possible to defined Write and Update Access to the Blockchain at a Group User Level.

- **User Defined Meta Data for Search:** IoT automation shall enable up to Peta Bytes of data stores for Supply Chain Partners. However in a Data Market Place, it is important that the data is tagged using User defined Meta Data so that it is easily discoverable by other Users.

3.2. SCM Device Identities

3.2.1. Issue

For any devices to be tracked in the IoT network, it needs a decentralized mechanism for endpoint discovery and identity. In the Internet world, this problem is solved by using the Domain Name System which scales up to millions of interconnected nodes. However with billions of IoT devices in the market, we need to evolve to a scalable and decentralized system.

3.2.2. Benefits

Based on Zokoo's Triangle it is very difficult to have an underlying Identity Management system which supports all three - Human Meaningful, Decentralized and Secure. Hence the Identity Management System is constrained to be Decentralized and Secure Only.

Namecoin, Blockname and Blockstack have created decentralized registry systems where the IoT device can register itself directly to the network, as against a centralized DNS assigning addresses based on a hierarchical context. For IoT Device Identity, Human Meaningful is not considered as a requirement and hence can be ignored.

3.3. Demand Management in a Digital Ecosystem

3.3.1. Issue

Another critical problem is demand management in the highly Digitized economy where consumers could order online, and suppliers fulfil orders through near shore production using 3-D printers with minimal turnaround time. Deliveries could be automated through drones or enabled through social delivery models or through other innovative methods.

Segmented Data in such an environment, will not allow the supply chain to leverage the true power of big data and automation. Instead, the Supplier would need to trust prediction models based on their limited data which might not be very accurate [10].

3.3.2. Benefits

IoT has enabled automation in Demand ordering and tracking. Blockchain provides a decentralized Data Store which can be accessed by trusted partners in real time. In addition to Authentication rules, the system provides an additional layer of security by using Encryption.

Blockchain also enables Decentralized Market Place which helps to share anonymized data with authorized partners which will help ensure that the Predictive Demand Models are strengthened by data from diverse geographies.

3.3.3. Gaps

Scalability: Bitcoin scales up to 7 TPS at peak time. One of the key reasons for this is Block size limit which has been set at 1MB currently, which limits the throughput and does not allow the system to scale in peak times. In addition all participating nodes participate in the Consensus protocol, hence as the number of nodes increases, system latency also increases. BIP101 was proposed to increase the Blockchain size steadily over time but was opposed since miners felt that it was very aggressive. Hence currently the system is limited by the number of transactions it can handle in a finite time. Technologies like IOTA, Lightning Network and Raiden are

coming up which shall increase the scalability of Blockchain and ensure that the system can scale beyond the 7 TPS [12].

Audit Trail: The information producer needs an immutable audit log of all consumers who have accessed information generated by them. In addition to trust in the system, in a Data Marketplace this is essential for ensuring that the rightful owners get paid for the data that they have generated.

3.4. Data Provenance

3.4.1. Issue

Provenance in SCM is a specialized form of Meta Data used to uniquely describe the source of entities, activities and people. Due to its importance in origin tracking its security needs are significantly higher compared to standard meta data.

Provenance of high value items such as Diamonds is currently stored on paper and can be easily stolen or tampered with. Similarly, in the medical sector counterfeit drugs with altered provenance can lead to death of patients. The same lack of Provenance Trust extends to expensive wine, watches, or handbags. As per the TE-Food [5] White Paper, even data coming from a third party like existing traceability system seems suspicious to the members.

3.4.2. Benefits

Blockchain Provides an Immutable, Secure and Decentralized Framework for Provenance Tracking for the Supply Chain. Since the Blockchain is an append only database, it lends a unique characteristic to ensure that the provenance of the product is tamper proof. Organizations like PROV-DM are using a modelling based approach using ontologies to help in automatic verification and inference of provenance management.

The start-up Everledger records 40 data points to uniquely identify a diamond in the Blockchain, since the Serial Number can be easily cut. A diamond buyer can use these data points to identify the origins and genuine owner of the diamond. Everledger plans to extend the concept Blockchain based fraud detection from diamonds to many other high value items [8]. Similarly by using barcodes technology Blockchain can be used to verify the provenance of a drug by medical patients. Blockchain has been applied in pharmaceutical SCM to provide patients with customized drugs using gene therapy.

3.4.3. Gaps

Storage Limitations: Blockchain is a decentralized ledger that is replicated on all full nodes that participate in the Consensus protocol. Data on a Blockchain is stored permanently and hence the volume keeps growing constantly over time. Hence the cost of storage on the Blockchain is prohibitively high. The fees for storing a 256 bit word is 20k gas as per the Ethereum Yellow paper [20].

3.5. Gaps Summary of existing Block chain and IoT Ecosystem for SCM

To summarize, based on review of existing systems, following would be key requirements of the enhanced IoT Block chain architecture to ensure faster adoption in SCM:

- Provide Privacy for SCM-IoT Transactions in the Pseudo anonymous ecosystem of Block chain
- Ability to cater to High IoT Throughput and Data Storage requirements in the low TPS, high cost storage ecosystem of Block chain
- Provide Role Based Access Control for trusted data sharing in a Permissioned Block chain

- Enable User Based Meta Data Tags for ease of search
- Immutable Audit Log for Access Verification and Revenue Settlement in Data Marketplace

4. DATA CHARACTERISTICS FOR BLOCKCHAIN AND IOT

Before we provide a recommendation architecture, to overcome limitations of the current IoT and Blockchain integration framework for SCM, let us review key characteristics of IoT Data for SCM as well as Blockchain capabilities as a Data store.

4.1. IoT Data Characteristics

Below are some sample scenarios where IoT Data is transmitted from a device and characteristics attributes of such data storage:

Table 3 IoT Data

Device	Data Characteristics
Data from Autonomous trading algorithms from Stock Market	Data transmitted at regular intervals during the day Reduced Data in night time from a Geography Data storage needs to store the change in value of a stock from previous time interval
Data from Smart Home Monitor such as Burglar alarm and Temperature Regulators	Data feed is spread at regular intervals through the day as well as night Can be some random spurts of data such as Alexa Feeds Typically Data store will need to store the incremental transition of the data e.g. change in temperature from 7:00 PM to 7:10 PM
Data from Self Driving Cars	At fixed intervals, example every minutes, the device transmits information of location, fuel spend and neighbouring vehicles Data will be send at periods when the Vehicle is on the road and there will be stationary periods when no data is transmitted.

A common attribute amongst all these Use Cases is that these applications get data feeds over pre-defined intervals of time and need to store the state change of the data across the time zones. Since time is a primary attribute for the data store, we can summarize that IoT Data is traditionally Time Series Data. Some other examples of Time Series Data would be Web Streaming Data, Satellite data for weather conditions and Agricultural Sensors to measure crop growth in a farm. To summarize, common attributes of the IoT data in SCM are:

- **Large Volumes of Data:** For e.g. a single connected car will transmit up to 25 GB of data per hour
- **Arrives at predefined intervals:** The time interval has a pattern of variation, example day and night traffic load, for e.g.:
- Vehicle Data Feed would be more in Day Time and significantly reduced in the night
- Agricultural Data Feeds would be more dominant from 7:00 AM – 7:00 PM
- Feeds Manufacturing Plants would be increased during working hours
- **Small Record Size:** Data Payload per feed is small since it has predefined attributes
- **Insert Only Data:** Data Streams are insert only with no update or delete operations

4.2. Blockchain as a Data Store

The Blockchain is a Distributed Ledger to immutably store transaction records. Specific characteristics of a Blockchain compared to traditional database that would apply to an IoT Data Store in SCM are as follows:

Table 4 Blockchain as a Data Store

Attribute	Traditional Database	Blockchain
Concurrency Control	Typically two-phase commit is implemented to avoid concurrency control	Since the Blockchain is built to handle more hostile environment such as Byzantine attacks, protocols such as Proof Of Work, Proof of Stake or Proof of Activity and Proof of Burn. These have higher computational overheads compared to two-phase commit and hence the system has low TPS
Event Chaining Structure	By default, transaction's in a traditional database are not interlinked by the database by default. It is possible for the developer to create Chains across entities by using Foreign Keys across entities.	Transactions in a given time frame are chained in Blockchain by using Chaining mechanisms such as Merkle Tree or Hash Chain.
Immutable Data Stores	Traditional Databases don't provide immutable data stores. It is possible for the admin to revoke update or delete privilege's for group of Users, however this is under Centralized Control and can be tampered with due to fraudulent or Use Case requirement	The Blockchain is an immutable data store due a combination of Event Chaining as well as Consensus Algorithms. The Chaining ensures that updates in a single block need to be reflected in all upward Blocks. And the Consensus mechanism ensures that transmission of Chaining information is approved by the Majority of Users.

5. ARCHITECTURAL COMPONENTS

Keeping the above summary requirements as objectives, and based on the deeper insight into IoT and Blockchain data attributes, following are the Architectural Principles that would lay the foundation for an efficient IoT Data Storage using Blockchain for SCM Analytics framework:

5.1. Dynamic Data Chunking

Figure 4 Dynamic Data Chunking

As discussed in IoT Data characteristics, IoT Data is Time Series based streaming data [17]. To store time-series data efficiently the data stream is split into dynamic time based chunks [6]. Key attributes of Dynamic Data Chunking would be:

- The application would extract IoT Data as a stream of time-tag-value record Chunks
- The Data chunks would be cryptographically chained together using hash pointer based chaining
- To cater to variations in Data Feeds volume the system will cater to multiple time slicing parameters that can be dynamically varied based on incoming Data Source Volumes
- Indexes would be a composite key of time as well as tags. The index would then be decoupled from the data to support efficient queries, to reduce the overhead of search queries on time based chunked data as compared to search run on granular records.
- In addition the data would be encrypted as well as compressed to increase security as well as efficiency.

5.2. Off Chain Data Store

As discussed above in “Block Chain as a Data Store”, the cost of storing high volume IoT Data in Bitcoin or Ethereum would be extremely high. Hence the solution should support Off Chain Data store. There are two categories of Off Chain Data Stores which shall be used based on functional requirements:

- **Inter Planetary File System (IPFS)** is a distributed file system technology build on DHT and Bit Torrent protocol. It is used to create a permanent and decentralized method to store and share files. For File Based stores such as Vector Image Files, the Blockchain shall store in IPFS. Saving data on IPFS creates a unique hash, which shall be stored on the Blockchain
- **BigchainDB Scalable Blockchain Database-** The BigchainDB [15] design starts with a distributed database , and adds Blockchain characteristics including decentralized control and immutability. It has scalability up to 1 Million writes per second. Data can be published via Smart Contracts to BigchainDB as Transactions using HTTP POST. The hash id of the transaction is returned as output which can be stored in the Blockchain.

Figure 5 Off Chain Data Store in Blockchain

5.3. Security through Role Based Access Control

Role-based access control (RBAC) is used for defining system access privileges for users based on defined roles. In centralized SCM systems, a hierarchy of roles and permissions are defined, and Users can be assigned to one or many roles based on their organization and job function, to defined access control permissions.

Figure 6 RBAC in BigchainDB

RBAC for Data writes can be simulated in a Blockchain by:

- Treating Roles as an “Asset” in the BigchainDB. App is the topmost asset in the hierarchy
- Categories of Roles are defined as Types, e.g. in a Vehicle Supply Chain, Vehicle, Retailer, and Dealer can be defined as Types
- Create hierarchies between Types by using “Link” operator which will help define relationships across assets. E.g. in a Vehicle SCM, if a Dealer can add Vehicles to the Inventory, the Dealer Asset will be “linked” to Vehicle Asset to indicate create privilege. A “Link” can be internally implemented by transferring special tokens to the Assets Wallet.
- Each New User in the enterprise created is an instance of the Role Asset
- Admin Users who can define association across Assets will have “Can Link” asset in their Wallets. The Can Link asset will be transferred to their Wallet based on Consensus which can be implemented through Proof of Stake.

Note: The above RBAC framework is applicable for Write Controls only. Read privilege to different sets of Users will be provided through cryptographic access.

5.4. Meta Data for Easy Search

Figure 7 User Defined Meta Data

A critical requirement for Big Data Analytics is the ability to perform large data set queries on User defined Metadata. Ethereum Smart Contract supports Metadata definition, however it does not currently support iterations for search on large contracts. Hence the Metadata definition is proposed to be stored as part of the BigchainDB. Assets are created through transactions and stored as JSON documents in BigchainDB.

5.5. Immutable Audit Trail

All Read requests from Trusted as well as Non Trusted partners to the SCM Transaction store shall be tracked by Ethereum Smart Contracts and an immutable audit log including Document and Reader ID shall be stored in the Audit repository.

6. IMPLEMENTATION DETAILS

Blockchain technology is currently going through a process of maturity and cost scalability. There are emerging technologies that promise complex capabilities, however the cost of using them for large scale Data Stores in a scalable and economical fashion is prohibitive. Hence the proposed technical stack promises to optimize the cost to performance factors of the SCM solution by distributing the Storage requirement across multiple layers based on Functional requirements:

- **Smart Contracts:** Ethereum shall be used to host the Producer Smart Contracts including User Access Rights. The smart contracts would store what's relevant for business & economic processing, and the rest is stored in the cheaper, faster metadata layer. The smart contracts primarily consist of User Access rights, schema verifiers, and factories. Each authorized User and Producer will have unique persistent identifiers.
- **Dynamic Time-Based Chunking:** would be implemented on the Edge on the IoT Device using Node.JS scripts. Node.JS has a web3 interface to Smart Contracts and hence enables easy communication with the Blockchain. To enable Dynamic Time-based chunking an adaptive approach shall be used based on a simplistic "FlexCap" algorithm that uses median block size strategy. To determine the time slice, it shall compute the median time slice needed to aggregate "N" Transactions over historical samples in the last "X" months, and a variable multiplier of "Z" shall be applied over the same to compute the time slice. Values of N, X and Z shall be optimized using Machine learning as the system operational results are accumulated.

Figure 8 Solution Component View

- **Vector File Hosting:** At the storage layer, files themselves need to be treated similarly. Services such as IPFS & Swarm can provide this backbone. The solution shall use IPFS to host the image files, as well as to interact with Ethereum.

- **Supply Chain Transaction, Audit Log & Meta Data Hosting:** Additionally, the solution has a proprietary back-end build on Blockchain DB technology. Given the current economic concerns around cost of storage BigchainDB is proposed to be the enabler for the Immutable Database to store transactions. An indexing/searching system across this layer would be enabled to ensure faster data access.
- **Data Pricing:** Data will be free and accessible to all, but at the same time it shall also be available at a price. Data will be charged through a Tariff Model implemented in the Smart Contracts based on Data price and Fungibility in the Data Marketplace.
- **User Defined Meta Data:** BigchainDB provides out of the box support for defining User Defined Meta Data. Sample Code for User Defined Transaction Metadata in BigchainDB:

```
"metadata": {
  "any": "MyMetaData",
  "timestamp": 1448362607
}
```

In order to perform query from the Smart Contract into BigchainDB the application shall use Data carrier oracles that relay query results from a trusted data source (BigchainDB in our case) to a smart contract.

Note: As with full-text search of asset data, the full-text search of metadata only works with MongoDB. That's because it uses MongoDB's built-in full-text search capabilities.

- **Front End GUI:** The front-end, shall use Truffle to handle contract management & testing, and Webpack for build management.

The result would be a platform where:

- Every Producer and Administrator will have a persistent identity with Smart Contract Based Authentication
- An immutable Data Store for Vector Files, Audit Trails as well as DB Based Profile Information
- An Easy to use GUI with Meta Data Search for Supply Chain Partners

7. DIRECTIONS FOR FUTURE RESEARCH

7.1. Sharding

In sharding the overall state is separated into different “shards,” and each part of the state is stored and processed by different nodes in the network. Each shard only processes a small part of the state and does so in parallel. Blockchain sharding is similar to sharding in the traditional database world, except with the extra hard challenge of needing to maintain security and authenticity amongst a decentralized set of nodes. There are several early projects such as Swarm, Storj and Decent that are using various strategies for splitting up the data into shards and storing it in a distributed fashion across participating nodes (i.e. distributed storage). The limitation with current sharding protocols is that only transaction processing load is reduced [13]. The complete Blockchain transaction history still has to be replicated on the validator nodes, resulting in operational overheads.

7.2. Legal Ownership of Data

Blockchain can ensure Data Provenance and Integrity, once the information has been truthfully published it will be immutable. However, it cannot ascertain the truth regarding the legal ownership of the data. This results in either incorrect data from the source or illegal owners claiming rights to data in a Data Marketplace. Some known issues are:

- Farmers can tamper agricultural IoT feeds by putting physical barriers, example if an umbrella is put on the sensor then climate information can be tampered
- There may be cases where there is genuine doubt on the owner of the data, e.g. if a sensor measures tire pressure, then is the data owned by tire company, car manufacturer, insurance company or car driver.

7.3. Bigchain DB is not Byzantine Fault Tolerant (BFT) Proof

At the time of writing this paper, BigchainDB does not support BFT, since all the nodes writes to a single cluster of the underlying Database, which could be MongoDB or RethinkDB. However, the BigchainDB team is planning to migrate to TenderMint for consensus, which will make it BFT compliant.[15]

8. CONCLUSIONS

It can be safely concluded that the Blockchain IoT framework will be a key driver to take SCM to the next level of Analytics resulting in enhanced efficiencies as well as predictive models by truly enabling Data Democracy. However, upon reviewing the Integrated Architecture, there are known gaps before the solution becomes financially and technologically feasible to be implemented at a global scale in SCM. The reference architecture provides guidelines which need be implemented by product vendor as well as solution integrator to ensure System Scalability as well as Security.

REFERENCES

- [1] S.A. Abeyratne, R.P. Monfared, Blockchain Ready Manufacturing Supply Chain Using Distributed Ledger, *Int. J. Res. Eng. Technol.* 5 (2016) 1–6.
- [2] B.C.W. Chase, The Digital Revolution Is Changing the Supply Chain Landscape, *J. Bus. Forecast.* (2016) 2016.
- [3] F. Chen, Z. Drezner, J.K. Ryan, D. Simchi-Levi, Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times, and Information, *Manage. Sci.* 46 (2000) 436–443.
- [4] K. Christidis, M. Devetsikiotis, Blockchains and Smart Contracts for the Internet of Things, *IEEE Access.* 4 (2016) 2292–2303.
- [5] TE Food, Making business profit by solving social problems, (2017) 1–50.
- [6] T. Gupta, R.P. Singh, I. Nsdi, Bolt : Data Management for Connected Homes This paper is included in the Proceedings of the, (2014).
- [7] A. Haddud, A. DeSouza, A. Khare, H. Lee, Examining potential benefits and challenges associated with the Internet of Things integration in supply chains, *J. Manuf. Technol. Manag.* (2017) 00–00.
- [8] E. Kaku, Using Blockchain To Support Provenance in the Internet of Things, (2017).
- [9] W. Kersten, T. Blecker, C.M. Ringle, N. Hackius, M. Petersen, Published in: Digitalization in Supply Chain Management and Logistics Blockchain in Logistics and Supply Chain: Trick or Treat?, 9783745043 (2017).
- [10] D.M. Lambert, M.C. Cooper, J.D. Pagh, Supply Chain Management: Implementation Issues and Research Opportunities, *Int. J. Logist. Manag.* 9 (1998) 1–20.
- [11] R.C.. Lamming, N.D.. Cadwell, D.A.. Harrison, W. Phillips, Transparency in supply chain relationships: concept and practice, *J. Supply Chain Manag.* 37 (2001) 4–10.

- [12] W. Li, A. Sforzin, S. Fedorov, G.O. Karame, Towards Scalable and Private Industrial Blockchains, Proc. ACM Work. Blockchain, Cryptocurrencies Contract. - BCC '17. (2017) 9–14.
- [13] L. Luu, V. Narayanan, C. Zheng, K. Baweja, S. Gilbert, P. Saxena, A Secure Sharding Protocol For Open Blockchains, Proc. 2016 ACM SIGSAC Conf. Comput. Commun. Secur. - CCS'16. (2016) 17–30.
- [14] A. McAfee, E. Brynjolfsson, Big data: The Management Revolution., Harv. Bus. Rev. 90 (2012) 61–67.
- [15] T. Mcconaghy, R. Marques, A. Müller, D. De Jonghe, T. Mcconaghy, G. McMullen, R. Henderson, S. Bellemare, A. Granzotto, BigchainDB: A Scalable Blockchain Database (DRAFT), BigchainDB. (2016) 1–65.
- [16] S. Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System, Www.Bitcoin.Org. (2008) 9.
- [17] H. Shafagh, L. Burkhalter, A. Hithnawi, S. Duquennoy, Towards Blockchain-based Auditable Storage and Sharing of IoT Data, (2017) 25–30.
- [18] F. Tian, An Agri-food Supply Chain Traceability System for China Based on RFID & Blockchain Technology, 2016 13th Int. Conf. Serv. Syst. Serv. Manag. (2016) 1–6.
- [19] M.A. Waller, S.E. Fawcett, Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management, J. Bus. Logist. 34 (2013) 77–84.
- [20] G. Wood, Ethereum: a secure decentralised generalised transaction ledger, Ethereum Proj. Yellow Pap. (2014) 1–32.