PROPOSAL OF AN ARCHITECTURE FOR DATA INTEGRATION AT AGRICULTURAL SUPPLY CHAINS, CONSIDERING THE IMPLEMENTATION OF IOT, NOSQL AND BLOCKCHAIN TECHNOLOGIES

Roberto F. Silva, Gustavo M. Mostaço, Fernando Xavier, Antonio Mauro Saraiva, Carlos E. Cugnasca

Department of Computer Engineering and Digital Systems,

Escola Politécnica da Universidade de São Paulo (USP), Brazil

roberto.fray.silva@gmail.com, gmostaco@usp.br, fxavier@usp.br, saraiva@usp.br, carlos.cugnasca@usp.br

ABSTRACT

Agricultural supply chains produce a huge amount of data related to traceability, production processes, environmental monitoring, among others. These are very important for the decision-making processes of the different supply chain stakeholders. With the implementation of technologies related to the Internet of things (IoT), the quantity and variety of data generated will increase even further. In order to extract useful information in an efficient way, it will be important to consider many aspects related to data management, such as security, processing, storage, transfer, etc. In this paper, we gather the requirements to implement IoT on agricultural SCs and propose an architecture that uses local databases to store raw and confidential data, a NoSQL database on the cloud to aggregate data that is important for decision-making by different stakeholders, and a blockchain that aggregates and safely stores the information for two main users. They are (i) consumers, considering aspects that guarantee product quality, together with what is relevant for them to choose between different products or brands; and (ii) government, which is related to data used for inspections, quality control, customs processes, and certifications. We conclude this paper by presenting at which level each of the functional, non-functional and domain-specific requirements will be fulfilled, and its main advantages in comparison with other architectures.

Keywords*:* agri-food, supply chains, IoT, big data management, blockchain

1. INTRODUCTION

A supply chain (SC) can be defined as a group of agents that are responsible, directly or indirectly, for supplying the demand of customers (Chopra, Meindl, 2013). Agricultural SCs deal with agents and products related to the agri-food domain, and they normally involve links such as production (at farms), warehousing (of raw materials and end products), processing (at industries), distribution centers, wholesalers, retailers and consumers.

An SC generates a considerable amount of heterogeneous and non-standardized data from different sources, which are processed by different software at the agents' systems and stored on different databases (Hribernik et al., 2010; Corella, Rosale, Simarro, 2013). In the case of agricultural SCs, these are related to three main systems: (i) traceability and product identification; (ii) environmental monitoring; and (iii) process monitoring. These will be further explored in Section 2.

The different stakeholders have different needs in terms of information availability, ease of use, confidentiality, among others. With the implementation of technologies related to the Internet of Things (IoT), these needs tend to increase. Nevertheless, new technical problems tend to rise, mainly related to the lack of interoperability and the generation and processing of huge amounts of data, demanding innovative ways to manage and organize the data flows.

In this sense, the objective of this research is to propose an architecture that uses local databases to store raw and confidential data in a NoSQL database, stored on the cloud, to aggregate data that is important for decision-making by different stakeholders, and a blockchain that aggregates and safely stores the information that is relevant for consumers and government officials.

2. DATA IN AGRICULTURAL SUPPLY CHAINS

There are three main types of systems that generate data on agricultural SCs (Silva et al., 2015; Pang et al., 2015; Verdouw et al., 2013; Chopra, Meindl, 2013; Verdouw et al., 2016):

1. **Traceability and product identification** → related to data generated during activities such as product and batches identification, tracking and tracing. The main technologies involved are barcodes, QR codes and radio-frequency identification (RFID). Data capturing can be automatic or semi-automatic. Product location is also part of this system;
2. **Environmental monitoring** → related to data generated while monitoring environmental variables that affect the product's quality. Ranges from the production at the farm, through processing at the industry, transportation or warehousing, all the way to the wholesalers and retailers shelves. The main environmental variables monitored for agricultural SCs are temperature, relative humidity, light, and pressure;
3. **Process monitoring** → related to data generated while monitoring production processes, mainly in the industrial environment. These are highly heterogeneous and can vary considerably among agents in the SC and even inside the same link. Most of the data generated on this system are not open for other agents in the SC and can contribute little for their decision-making processes. Nevertheless, some data, such as which ingredients and raw materials used, should be made public, as well as the transformation processes.

System 1, related to traceability, is very important for agricultural SCs and food products, as it is crucial to identify product contaminations, batches with problems, and the location of products in the SC, among others. Traceability can be defined as the tracking and tracing of product batches throughout the SC, allowing for the identification of critical quality control points (Juran, Godfrey, 2000).

As several agents adopt different technologies to allow product traceability throughout the SC, a series of interoperability problems can be generated. One way to reduce the occurrence of these problems is to utilize a common and widely accepted paradigm that describes how each technology should interact on the SC. One such paradigm is the IoT, which will be described in the next section.

3. RELATED WORKS

According to Atzori, Iera and Morabito (2010), IoT is a paradigm in which technologies related to product traceability, identification and monitoring - such as RFID, Wireless Sensor Networks (WSN), Global Positioning System (GPS), among others - interact with each other. This can be used, for example, to remotely monitor the weather at a vineyard, while connecting grapes harvesting data in each plot to an RFID tag, achieving traceability from the farm link to the end consumer. Several kinds of research use this paradigm for the generation, collection, processing, and storage of data from agricultural SCs (Gubbi et al., 2013; Pang et al., 2015; Corella, Rosale, Simarro, 2013).

Lack of interoperability is a very important problem for the implementation of the IoT paradigm. Currently, there is no universally accepted standard for this. According to Harris, Wang and Wang

(2015), two main factors cause problems on the adoption of IoT technologies on SCs: (i) lack of interoperability between existing systems and (ii) long implementation periods associated with ICTs. To the best of our knowledge, most architectures, such as the IoT-A (Carrez, 2013), do not fulfill all the requirements from agricultural SCs.

To address the problem of excessive volume, variety, and velocity of data generation, the NoSQL database was proposed. Its main advantages over traditional database models are a fast read and write operations, easy expansion, support for mass data storage and low cost. Moreover, the use of NoSQL databases is driven by other performance reasons, such as the avoidance of unneeded complexity, high throughput and horizontal scalability.

While traditional databases are mostly based on the relational model, NoSQL databases use other data models, such as key-value or column-oriented documents, each one presenting its related advantages and disadvantages. The study of Kamilaris, Kartakoullis and Prenafeta-Boldú (2017) reviews the Big Data technologies used in agriculture, with the use of cloud computing, machine learning and NoSQL databases in several studies. As highlighted by these studies, the increased availability of data collection devices represents both an opportunity and a challenge to turn this massive volume of heterogeneous data into useful information that could be addressed by the NoSQL model. Nevertheless, it is also important to make relevant information accessible to the agents in the supply chain. Among these, consumers and the Government demand data that could be easily accessible and trustworthy. The blockchain technology is an option to address this need.

A blockchain can be described as a sequence of information blocks that are connected with each other using cryptographic hashes. Each block contains information in its header, data from the present transaction, besides the unique ID of the previous block (Zheng et al., 2017). The header information is used to ensure data consistency within the chain by a cryptographic hash. In the blockchain, there are mechanisms for authorization and validation of the read and write operations of the blocks, ensuring security in the transactions (Hackius, Petersen, 2017; Tian, 2016).

Using a blockchain, data control is completely decentralized, bringing more transparency, without losing control of issues such as access security, and data integrity (Hackius, Petersen, 2017). In addition, with fewer intermediary elements on the transactions, there may also be performance gains. Decentralization, security and integrity are labeled as fundamental advantages of blockchain technology in relation to other technologies, due to the use of other concepts already established in Computer Science, such as distributed computing, cryptography and fault tolerance mechanisms.

4. METHODOLOGY

The methodology used in this research is based on two steps:

1. Identification of the main functional, non-functional and domain-specific requirements, through an extensive and in-depth literature review using the Scopus database;
2. Proposal of an architecture to address the requirements identified on the first step, considering real-world scenarios, as well as the increase of the IoT paradigm adoption, without standardization of all its technologies.

5. REQUIREMENTS IDENTIFICATION

An in-depth literature review was conducted to identify the main requirements for data integration at agricultural SCs, considering the IoT paradigm. A series of proposals and frameworks for the implementation of IoT on SCs were analyzed. For means of this research, it was discarded models that: (i) considered only specific links in the SC; (ii) did not consider interoperability problems; (iii) did not consider the requirements of the different agents in the SC; and (iv) did not consider product quality monitoring.

The main requirements were identified and separated into three groups: functional (6 requirements), non-functional (4 requirements), and domain-specific (7 requirements). Functional requirements are related to the existence of services, functionalities, techniques, and tools to manage different hardware and software resources and aspects such as energy management. Non-functional requirements are related to performance, quality of service level, security, availability, and trustworthiness. Domain-specific requirements are related to implementation in a particular domain. Those requirements are presented in Table 1.

6. PROPOSED ARCHITECTURE

The proposed architecture is illustrated in Figure 1. It is composed by:

- Data generation and collection at each SC agent. Data should be standardized, if possible;
- Raw data storage at local databases. Each agent should store data that is not yet suited for decision-making on the SC level. Nevertheless, the agents should, as much as possible, treat their data to eliminate errors, outliers, null values, etc.;
- Transmission of non-confidential local data from the three systems to the cloud. The agent should, if possible, label the data to help on fusion and processing activities. If not, unsupervised machine learning methods, such as k-means could be used in the cloud to create labels for these data. In the cloud, two processes will happen: (i) data processing and fusion using a middleware; and (ii) storage of processed data in a NoSQL database;
- Transmission of only relevant data for consumers and the Government to the blockchain, which is openly available. This would contain data related to product batches origin (System 1), and the processes and ingredients that were used during its transformation (System 3)

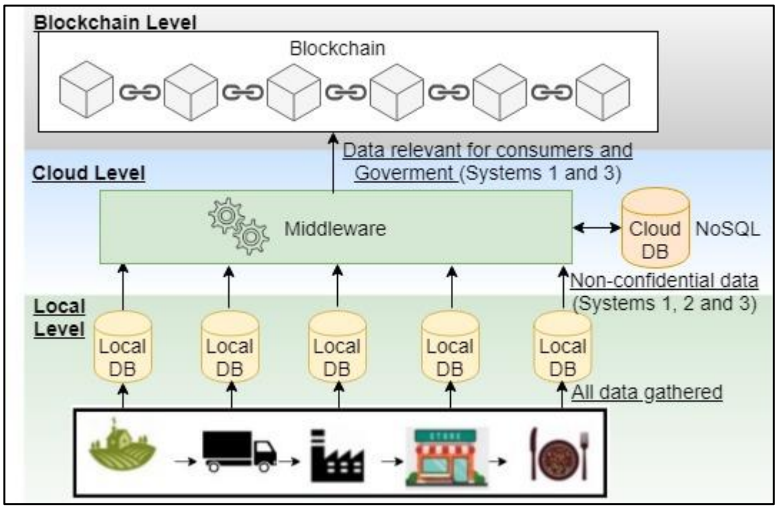

Figure 1. Proposed architecture for data integration at agricultural SCs

Table 1 contains the descriptions of how the proposed architecture addresses the functional, non-functional and domain-specific requirements previously identified. The functional and domain-specific requirements will be fulfilled at the local, cloud and blockchain levels, while the non-functional requirements will be fulfilled at the cloud and blockchain levels.

Table 1 - Proposed architecture and fulfillment of functional requirements

Requirement*	**How it will be fulfilled in the proposed architecture**
F1. Automatic identification	**1. Local level:** a) Standardizing the agent system as much as possible

F2. Resources management **F3.** Data management **F4.** Events management **F5.** Interoperability **F6.** Context identification	**2. Cloud level:** a) Functionalities developed on top of the IoT-A framework; b) Middleware for processing, NoSQL for data storage **3. Blockchain level:** a) Consumers and Government visualizations
NF1. Autonomy **NF2.** Middleware scalability **NF3.** Resilience **NF4.** Privacy and security	**1. Cloud level:** a) Functionalities developed on top of the IoT-A framework; b) Middleware that uses parallel computing **2. Blockchain level:** a) Fault tolerance and authorization
D1. Technologies coexisting **D2.** Creation of virtual SCs **D3.** Distributed traceability **D4.** Virtual object quality control **D5.** Heterogeneous data **D6.** Authorization, reliability **D7.** Information fusion	**1. Local level:** a) Standardizing the agent system as much as possible; b) Guaranteeing that the agent follows traceability rules **2. Cloud level:** a) Functionalities developed on top of the IoT-A framework; b) Middleware for data processing; c) Authorization levels **3. Blockchain level:** a) Fault tolerance and authorization

*Groups are: functional (F), non-functional (NF) and domain-specific (D).

Unlike the model proposed by Tian (2016), in proposed architecture, the data does not need to follow standards to be used for decision-making. The use of a middleware on the cloud will allow both a faster implementation (as the agents will not need to adopt standards before they participate in the system), but it will also allow for coexisting technologies. In addition, to the best of our knowledge, the proposed architecture takes the lead on clearly identifying the systems that will generate most of the data in an agricultural SC.

7. CONCLUSIONS

With the implementation of IoT technologies, agricultural SCs will generate a huge amount of heterogeneous data. The proposed architecture can be considered more adequate in this context, in which huge amounts of data becomes a challenge to extract useful information. In order to store, process and present relevant information for consumers and Government officials, the requirements were identified and an architecture was presented for implementing IoT on agricultural SCs. This was encompassed by the use of local databases at the agents to store confidential data and the use of a central cloud database with NoSQL technology. They are connected by a middleware, capable of parallel processing and that receives important data for the SC, and the use of a blockchain to encrypt and present relevant data.

The main difficulty on developing this architecture was the lack of open real-world data for the requirements identification, as companies rarely share their data. Overcoming this limitation demanded a thorough literature review. Nevertheless, several points may be developed on future research, such as a more thorough description of each level in the architecture, a description of how it can be implemented together with the IoT-A framework, and the development of the middleware that will allow all the data to be processed and stored on the NoSQL database.

ACKNOWLEDGMENTS

This work was supported by Itaú Unibanco S.A. through the Itaú Scholarship Program, at the Centro de Ciência de Dados (C²D), Universidade de São Paulo, Brazil, by the National Council for Scientific and Technological Development (CNPq), and also by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001.

References

Atzori, L., Iera, A., Morabito, G. (2010) 'The internet of things: A survey', Computer Networks, 54(15), pp.2787-2805.

Carrez, F., Bauer, M., Boussard, M., Bui, N. (2013) 'Final architectural reference model for the IoT v3. 0', EC FP7 IoT-A Deliverable, 1.

Chopra, S., Meindl, P. (2013) 'Supply chain management: Strategy, planning, and operation', 5th ed. New Jersey, USA: Pearson Education, 528pp.

Corella, V.P., Rosalen, R.C., Simarro, D.M. (2013) 'SCIF-IRIS framework: a framework to facilitate interoperability in supply chains', International Journal of Computer Integrated Manufacturing, 26(1-2), pp.67-86.

Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M. (2013) 'Internet of Things (IoT): A vision, architectural elements, and future directions', Future Generation Computer Systems, 29(7), pp.1645-1660.

Hackius, N., Petersen, M. (2017) 'Blockchain in logistics and supply chain: trick or treat?', In: Proceedings of the Hamburg International Conference of Logistics (HICL), pp. 3-18.

Hribernik, K.A., Warden, T., Thoben, K.D., Herzog, O. (2010) 'An internet of things for transport logistics–an approach to connecting the information and material flows in autonomous cooperating logistics processes', In: Proceedings of the 12th international MITIP conference on Information Technology & Innovation Processes of the Enterprises, pp. 54-67.

Juran, J. M., Godfrey, A. B (2000) 'Juran's Quality Handbook', 5th ed. USA: McGraw-Hill Publisher, 1872pp.

Kamilaris, A., Kartakoullis, A., Prenafeta-Boldú, F.X. (2017) 'A review on the practice of big data analysis in agriculture', Computers and Electronics in Agriculture, 143, pp.23-37.

Pang, Z., Chen, Q., Han, W., Zheng, L. (2015) 'Value-centric design of the internet-of-things solution for food supply chain: value creation, sensor portfolio and information fusion', Information Systems Frontiers, 17(2), pp.289-319.

Silva, R.F., Praça, I., Yoshizaki, H., Cugnasca, C.E. (2015) 'Proposal of a traceability model for the raw Brazilian sugar supply chain using RFID and WSN', In: Production and Operations Management Society, POMS 26th Annual Conference, Washington DC, USA.

Tian, F. (2016) 'An agri-food supply chain traceability system for China based on RFID & blockchain technology', In: 13th international conference on service systems and service management (ICSSSM), pp. 1-6.

Verdouw, C.N., Vucic, N., Sundmaeker, H., Beulens, A. (2013) 'Future internet as a driver for virtualization, connectivity and intelligence of agri-food supply chain networks'. International Journal on Food System Dynamics, 4(4), pp.261-272.

Verdouw, C.N., Wolfert, J., Beulens, A.J.M., Rialland, A. (2016) 'Virtualization of food supply chains with the internet of things'. Journal of Food Engineering, 176, pp.128-136.

Zheng, Z., Xie, S., Dai, H., Chen, X., Wang, H. (2017) 'An overview of blockchain technology: architecture, consensus, and future trends', In: 2017 IEEE International Congress on Big Data (BigData Congress), pp. 557-564.