Structure proposed to deal with sensor-based data

Proposta de estrutura para tratamento de dados oriundos de sensores

Leandro Colevati dos Santos[1,2,*] (iD), Sebastião Gomes dos Santos Filho[2] (iD), Maria Lúcia Pereira da Silva[1,2] (iD)

1. Faculty of Technology of São Paulo, Centro Paula Souza – São Paulo (SP), Brazil.

2. School of Engineering, Universidade de São Paulo – São Paulo (SP), Brazil

Correspondence author: leandro.colevati@usp.br

ABSTRACT

The possibility of greater control over processes, with the increasing use of sensors, leads to the generation of a large amount of data and, consequently, inherent difficulties, also need care. Storage, processing, filtering, become more difficult and the use of cloud computing can be a solution, however, the cloud computing technology also has limitations regarding latency and availability that can be mitigated with the use of an intermediate layer, called fog computing. As the structuring of the fog computing layer is not yet standardized, one of the possibilities is the application of an enterprise service bus (ESB), a structure based on the availability and interconnection of targeted services, for processing, filtering or storage. From an ESB structure such as fog computing, proofs of concept were developed to validate the technology, the first proof of concept being an environment, whose temperature must be controlled, with distributed sensors and their values received, processed and stored by services in the ESB and the second proof of concept, based on soil contaminant data, obtained by sensors and made available in the literature, for which filtering and processing services, on the same bus, were used. The study presents, in its results, the applicability of the fog computing layer, minimizing the limitations of cloud computing.

KEYWORDS: Enterprise Service Bus, Fog Computing, Sensors.

RESUMO

A possibilidade de maior controle sobre os processos, com o uso crescente de sensores, leva à geração de uma grande quantidade de dados e, consequentemente, dificuldades inerentes, que também precisam de atenção. Armazenamento, processamento, filtragem, tornam-se mais difíceis e o uso da computação em nuvem pode ser uma solução, no entanto, a tecnologia do uso da computação em nuvem, também tem limitações em relação à latência e disponibilidade que podem ser mitigadas com o uso de uma camada intermediária, chamada computação de névoa. Como a estruturação da camada de computação da névoa ainda não é padronizada, uma das possibilidades é a aplicação de um barramento de serviços corporativos (ESB, na sigla em inglês), que se trata de uma estrutura baseada sobre a disponibilidade e interconexão de serviços direcionados, para processamento, filtragem ou armazenamento. Partindo de uma estrutura ESB, como computação de névoa, provas de conceito foram desenvolvidas para validar a tecnologia, a primeira prova de conceito é desenvolvida em um ambiente, cuja temperatura deve ser controlada, com sensores distribuídos e seus valores recebidos, processados e armazenados pelos serviços no ESB e a segunda prova de conceito, com base em dados de contaminantes do solo, obtidos por sensores e disponibilizados na literatura, para os quais filtragem e serviços de processamento, no mesmo barramento. O estudo apresenta, em seus resultados, a aplicabilidade da camada de computação em névoa, minimizando as limitações da computação em nuvem.

PALAVRAS-CHAVE: Barramento de Serviços Corporativos, Computação em Névoa, Sensores.

INTRODUCTION

The consolidation of the technology Internet of Things (IoT) allows elements such as sensor circuits or actuators be arranged as nodes in a large and interconnected network and exchanging data[1]. In addition, intelligent and integrated devices, such as smartphones, offer a large number of sensor circuits and they are also intrinsically

nodes in a large network, since they are already designed connected in high-speed packet access (HSPA), long-term evolution (LTE) (also called 3G and 4G) or Wi-Fi networks. Direct consequences of high availability of sensor devices are the high amount of data generated and the difficulties in storing, filtering, analyzing and interpreting this data, letting alone the high latency established by conventional solutions[2]. A similar situation occurs with equipment based on vacuum technology, such as mass spectrometers, itself or coupled to more complex devices, i.e., tandem methods. The equipment parameters, as well as the produced data, can be huge, and a paramount example is the analysis of organic compounds on soil.

This article was applied research base. Most of the steps taken were exploratory, since it was defined by proof of concept, demonstrating the possibility of validating an idea using a prototype and a test script. Therefore, it was a qualitative and quantitative work, because it defined the software performance and their premises. Thus, this work aimed to present a solution to these listed difficulties, establishing a low-cost computational structure and with reduced latency and preprocessing. To achieve this objective, studies based on sensor measurements were developed.

THEORETICAL FRAMEWORK

Considering the characteristics of cloud computing and its main limitations, it was developed the concept of fog computing, which can be defined as a layer at the edge of the cloud, in which there is also a part of processing, communication and data storage[3]. Communication can be done between nodes, that is, computing devices or between any device and the cloud in which this layer is associated[4]. Therefore, an architecture with cloud computing associated with fog computing was defined in three layers: the top layer is the cloud, the middle layer is the edge of the cloud (the fog), and the last layer, in which the computational nodes are, is usually computers and smartphones. However, due to the concept of IoT, the last layer can also be other devices, such as wearable devices, sensors, among others[5].

One of the characteristics of fog computing is that some data does not need to be grouped and, when eventually it is no longer needed, it does not need to be in data centers, which gives greater autonomy and greater control to data owners[3]. Another characteristic is that, because it is closer to the end user, it has less latency in the transmission of data, which in cases of applications that require faster decision making is of paramount importance and mitigates one of the limitations of computing in cloud[5]. Due to these characteristics, the union of fog computing and big data technologies becomes very immediate, with the fog computing layer responsible for "acquisition, aggregation, pre-processing, reduction, data movement and storage"[4].

Big data is not just a large collection of data, but something with "great meaning, complexity and challenge"[6]. The literature describes the main characteristics of big data from the five V: volume, speed, variety, veracity, and value. When referring to volume, it is specifically the amount of data generated, usually measured, in data generated versus time unit. Speed refers to how much data per unit of time can be generated or transmitted. Variety is the characteristic that presents the diverse (and countless) ways that data can be presented in a context. Veracity deals with the quality of the data, which must be as accurate as possible. Value refers to the ability of a big data set to add value to product generation, to improve processes, that is, to the context to which it is applied[7].

Considering a big data context associated with fog computing, it is possible to design applications focused on augmented reality, along with wearable devices, such as smartglasses[3], large-scale environmental monitoring systems[4], and geolocation[5]. Sensors such as IoT nodes, associated with intelligent devices, that is, with computational capacity, will generate a lot of data per second that will need to go through some phases before they are available in a cloud or may be part of an immediate decision-making process, to which the cloud layer will be limited. These phases can be data processing or data transmission[8].

A fog computing structure can be defined in several ways, and the proposal of this work was its architecture based on an enterprise service bus (ESB), which is a way to transform an application into small units with defined activities that are named services[9]. These services are described by metadata that will be understood by any application, or even by other services. Despite being an infrastructure based on service-oriented architecture (SOA), it is not restricted to a single protocol, such as SOAP or HTTP, and can integrate several protocols[10]. The main characteristic of an ESB is an intermediary structure based on messages that offer ways to intercommunicate services and functionality to applications with security and reliability[11]. Therefore, the bus should offer ways to manage, in a heterogeneous environment, the ways of operating and distributing messages[12].

This work presents a proof of concept of a sensing system applied to a fog computing structure, as shown in Fig. 1. The use of proof of concept as a methodological structure was adequate for the development of this research, because it had as premises to verify the standards obtained through the literature and to confront them with ideas and experiments to validate whether there is applicability in a real environment[13]. The main objective, as a proof of concept, was to add, in a fog computing environment, as proposed by Santos et al.[2], services that allow the generation, processing, movement and storage of data.
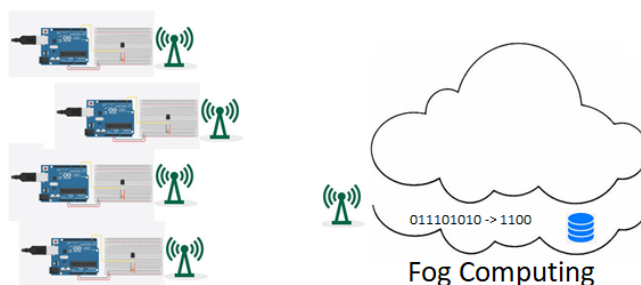

Figure 1: Base structure to work with sensors in fog computing.

For this study, the authors chose contexts in which data needs to be processed or trafficked quickly for possible decision making. The first context concerned an environment of controlled temperature, which cannot suffer great variations and, in case of increase of the temperature in some degrees, equipment can suffer serious damages. The second context concerned the analysis of volatile organic compounds (VOC) in the soil over long periods, in which some sensitive variations in the concentrations of these compounds could impair the activity that would be performed in that soil. For both cases, computers or smartphones that relate to the measurement systems, as well as services that connect people who can make decisions, were designed for the solution[14]. VOCs are pollutants that come from the activities of chemical industries present in lubricants, solvents, among others. It is an organic compound present in photoreaction. Examples of VOCs are methane, ethane, tetrachloroethane and other chlorohydrocarbons and perfluorocarbons. Emissions of this type of compost must be controlled, as they can generate environmental impacts such as the influence on climate changes or the health of animals[15]. Direct impacts on the ozone layer and on the biota that regulates the climate in the oceans are also possible[16].

The fog computing structure proposed by Santos et al.[2] was based on a corporate service bus that already had a data pre-processing and data storage service structure, and it was already foreseen that the data could vary widely.

The first proof of concept structure, considering sensors, was structured to be coupled to the ESB as fog computing, whose main characteristic is to allow the insertion of new services, as shown in Fig. 1. Sensors generated the data in the environment that were inserted, the IoT node that had the sensors connected sent the data to a service that was waiting for this sending in full time, and a data aggregation and reduction service did the pre-processing and sent it to a data storage service.

The second proof of concept was based on the study done by Schumacher et al.[14]. The structure was based on the measurement of VOC made with sensors – or sample collection –, arranged at different depths of a soil for a period. The measurements were redone in short periods, and the concentration of the soil contaminant was stored in a dataset. It is a measure with several characteristics, similar to the study carried out by Santos et al.[2]. Therefore, the ESB was suitable for pre-processing, since Schumacher et al.[14] had a dataset greater than 15k measurements.

RESULTS AND DISCUSSION

The first proof of concept was developed based on a matrix of thermal sensors, placed in a room whose temperature should be 18°C and cannot exceed 20°C. The matrix was connected to a Wi-Fi network, and each sensor had a unique identification. For this proof of concept, NTC sensors were used and chosen by their high sensitivity to temperature rise. The matrix was connected to a microcontroller-based system, with a Wi-Fi module to guarantee the connection.

In order to connect to the ESB defined by Santos et al.[2], this work proposed a service that was listening, in real time, to data coming from the sensors, a pre-processing service that makes a data analysis and, in case of non-compliance with the determined standards already trigger, an alert service, and a database storage service.

The web scraping service, looking for a reliable source for the external temperature, was necessary, because the average room temperature could need to be adjusted if the temperature outside was too high. The service based on web scraping deals with the communication with a recognized and reliable website that provides the local temperature, for several regions of Brazil and the world, with update in a short time. This service scrapes the website, which is to read its code and to extract the raw data, disregarding aspects and graphic elements that are usually implemented to make content available to end users. The schematic representation is shown in Fig. 2. To validate if the proposed environment was efficient in this configuration, an application connected to the fog computing environment was developed to present the temperature variation in real time, as well as the external temperature. Fig. 3 shows an application connected to the fog computing environment.
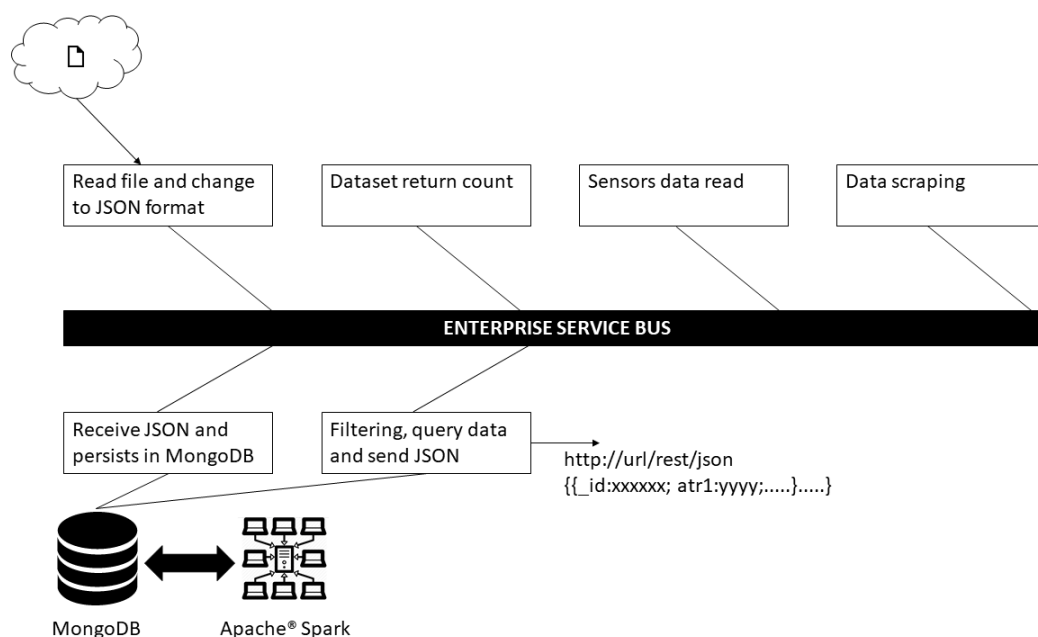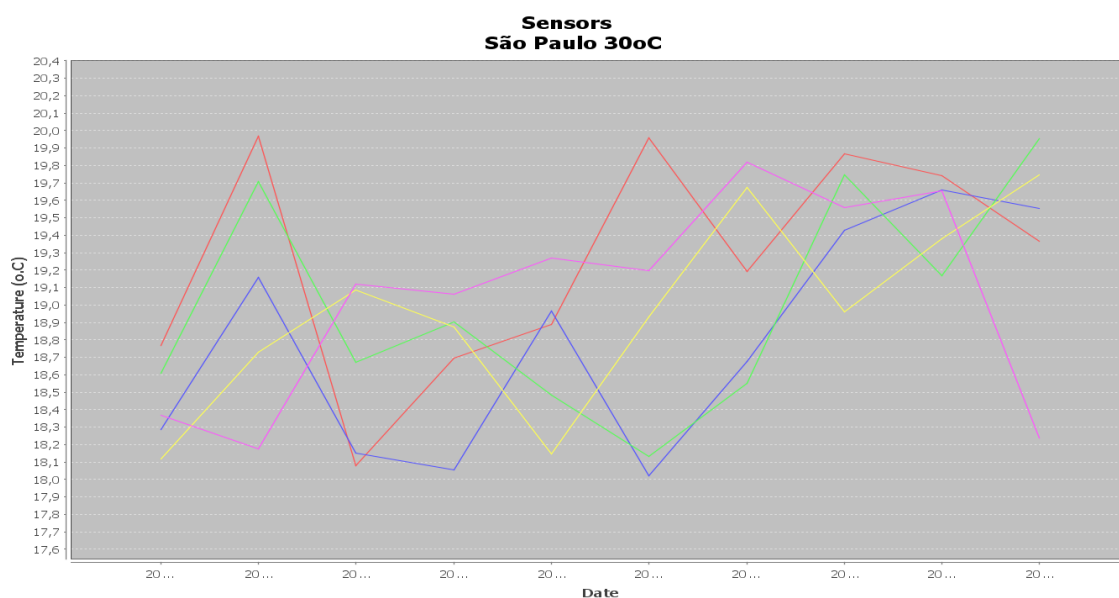
Figure 2: Fog computing structure.

Figure 3: Third party application connected in fog computing.

The first proof of concept showed that the structure was consolidated, also to deal with a mass generating environment of data in real time, and provided a solution with quick response to immediate decision-making needs.

The second proof of concept carried out with the finality of confirming the validity of results on the first proof of concept and had the same principle, that is, data from sensors from a potentially distant environment that need pre-processing. In this case, the greatest relevance was in the pre-processing, since the data needed to be evaluated and eventual distortions had to be discarded, so the set of measures that can feed a database in the cloud met the expectations of veracity and value, in addition to the volume, to facilitate later analysis as by artificial intelligence.

The analysis was performed using the dataset of Schumacher et al.[14] and the parameters of Santos et al.[2] to simulate a sensor reading. The data filtering service of a file, already available at ESB, was used to simulate the reading of sensors. A service for pre-processing the data was coupled to the bus to filter by sensor (organic compound in a certain depth of the soil), its concentration and, in the sequence, to reduce it to be able to send it to the data storage service. The measurement data was structured. Therefore, it was necessary to have a structured data storage service, differently from the one used in the first proof of concept. As expected, some measures could give non-standard values, due to situations that can get out of control. Then, they had to be discarded already in the pre-processing phase. A service to provide filtered data, as in Santos et al.[2], was also coupled to the ESB, so that third-party applications consume and use it in graphs, tables or reports. Figure 4 shows a consumption of an MS Excel® spreadsheet connected to the data storage service, filtering through the element 1-1-1 tetrachloroethane in the soil in a period.
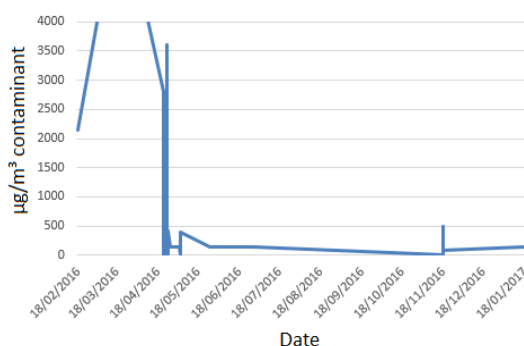

Figure 4: MS Excel® application connected in fog computing, data about 1-1-1 tetrachloroethane in soil.

CONCLUSIONS

Considering as the main limitations of a cloud computing environment the latency in data transmission and the processing and analysis of data in an environment closer to the end user, or to provide results in a short time, or to filter and make data reduction, fog computing presents itself as an adequate, intermediate option, focused on solving these limitations. As it does not have a defined structure for its implementation and as it presents itself as an intermediate structure for exchanging messages, the ESB also becomes an adequate way to structure a fog computing layer.

The first proof of concept can demonstrate practical applicability for structured fog computing in an ESB for the need to acquire large amounts of data, pre-processing and communication with low latency. The data is already filtered and can be used for any application that wants to work with that data.

The second proof of concept presents the capacity, and, therefore, the practical applicability for fog computing in an ESB for the need for pre-processing, analysis and filtering of data from an environment that generates a large amount of data in a short period. Even if it is a group of interrelated data, with wide variations, the analysis close to the end user makes the grouping of data and the cleaning of data that are far out of expectations increases the veracity and value of the whole set.

Finally, it can be said that the solution is versatile, since it presents good results for measuring temperature and contaminants.

AUTHORS' CONTRIBUTION

All the authors contributed equally to this study.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable.

FUNDING

REFERENCES

1. Sehrawat D, Gill NS. Smart sensors: analysis of different types of IoT sensors. In: Annals of the 3. ICOEI; 2019. 2019.

2. Santos L, Simões EW, Silva MLP, Santos Filho SG. Proposta de estrutura computacional decentralizada, de fácil uso, para aplicação em seleção de dados ambientais dispersos e não estruturados. In: Annals of the 21. Engema; 2019. 2019.

3. Yi S, Hao Z, Qin Z, Li Q. G computing: Platform and applications. In: Annals of the 3. IEEE Workshop on Hot Topics in Web Systems and Technologies (HotWeb); 2015. 2015.

4. Yi S, Li C, Li Q. A survey of fog computing: concepts, applications and issues. In: Proceedings of the 2015 Workshop on Mobile Big Data; 2015. 2015.

5. Vaquero LM, Rodero-Merino L. Finding your way in the fog: towards a comprehensive definition of fog computing. Comput Commun Rev. 2014;44(5):27-32. https://doi.org/10.1145/2677046.2677052

6. Ward JS, Barker A. undefined by data: a survey of big data definitions. arXiv:1309.5821. 2013.

7. Nguyen TL. A framework for five big v's of big data and organizational culture in firms. In: IEEE International Conference on Big Data (Big Data); 2018. 2018.

8. Santos LC, Silva MLP, Santos Filho SG. Proposal for decentralized computational structure for selection of dispersed and non-structured data. Int J Dev Res. 2020;10.

9. Maréchaux JL. Combining service-oriented architecture and event-driven architecture using an enterprise service bus. IBM Developer Works. 2006;1269-75.

10. Schmidt M-T, Hutchison B, Lambros P, Phippen R. The enterprise service bus: making service-oriented architecture reality. IBM Systems J. 2005;44:781-97.

11. Menge F. Enterprise Service Bus. In: Free and Open Source Software Conference; 2007. 2007.

12. Keen M, Acharya A, Bishop S, Hopkins A, Milinski S. Patterns: Implementing an SOA using an enterprise service bus. IBM Redbooks. 2004;336:20-8.

13. Canele RC. Construindo ontologias a partir de recursos existentes: uma prova de conceito no domínio da educação [thesis]. São Paulo: Universidade de São Paulo; 2009.

14. Lutes CC, Holton CW, Truesdale R, Zimmerman JH, Schumacher B. Key design elements of building pressure cycling for evaluating vapor intrusion—a literature review Ohio Wiley-Blackwell; 2019.

15. Khan FI, Ghoshal AK. Removal of volatile organic compounds from polluted air. J Loss Prev Process Ind. 2000;13(6):527-45. https://doi.org/10.1016/S0950-4230(00)00007-3

16. Mellouki A, Wallington TJ, Chen J. Atmospheric chemistry of oxygenated volatile organic compounds: impacts on air quality and climate. Chem Rev. 2015;115(10):3984-4014. https://doi.org/10.1021/cr500549n