Hardware and Software Infrastructure for Analysis, Processing and Decision Making in Medical Entities Through the Use of Big Data

Miguel-Angel Quiroz-Martinez [1], Christopher-Gustavo Roby-Cevallos [1], Daniel-Humberto Plua-Moran [1], Maikel-Yelandi Leyva-Vazquez[1]

[1] *Computer Science Department, Universidad Politécnica Salesiana,*

Guayaquil, Ecuador

[2]*Facultad de Ciencias Matemáticas y Físicas , Universidad de Guayaquil , Guayas, Ecuador*

ABSTRACT

Information was analyzed from architectures and models generated by big data and IoT research for the medical area. The problem is the lack of proposals to have hardware and software infrastructures based on scientific research that assist in the analysis and processing phases to make decisions in medical organizations based on large volumes of data. The main objective of this research was to propose the hardware and software infrastructure for analysis, processing, and decision making in medical entities using large volumes of data. The development proposal of the following research work uses the analytical, inductive, deductive, observation, and quasi-experimental method that allows us to propose a general IoT and Big Data model and architecture for medical entities. This proposal resulted in a General IoT model in the health sector, a General IoT architecture for the health sector, and a Big Data General Architecture for Health Sector. It was concluded that big data and IoT are complemented by data lifecycle management in the capture, storage, processing, and analysis; this management is a conceptual proposition so that other researchers can deepen the design; the physical components of the architecture influence low performance, in part software components assist in high performance.

Keywords: Decision making, Medical entities, Large volumes of data, Internet of Things, Big Data

INTRODUCTION

Within Information and Communication Technologies (ICT), there is the Internet of Things (IoT) that is used to manage patient data through cloud computing; with this, the data collected becomes intelligent information for patient analysis and diagnosis (Yao *et al.*, 2019). In 2020 there are 35 Zettabytes of information in the medical field generated by the penetration of the Internet in the medical industry; in addition, electronic medical information, hospital systems, images, and other kinds of data are growing; this growth is in public and private health; it is necessary to take into account the security and privacy of medical information because there are reviewed by various levels of professionals; also, the devices that capture and send health data are involved (Lv and Qiao, 2020). Data is a valuable asset in today's organizations; Big Data (BD) is managed by information systems and social sciences; the impact of large volumes of data is seen in social networks, sensor networks, mobile hardware, marketing, education, smart city, production, electronic business, among others; BD is used in banks 43%; technology 14%; Manufacture 8%; energy 8%; consumption 9% and Health 11% (Rahul and Banyal, 2020). The privacy and security of patient information are of great importance when accessing data; the internet, smart devices, and information systems are used by doctors, medical assistants, and providers to generate or visualize information (Lv and Qiao, 2020). Big Data is a high volume of data with high speed; it is a variable information asset with cost-benefit requirements and improvement in its processing to improve decision-making (Chen and Wu, 2020); the main feature of BD is its heterogeneity and continuous growth that requires non-standard guidance for data storage and processing (Shakhovska *et al.*, 2019); aims to transform data into knowledge; BD has three features: volume, velocity, and variety (Nadal *et al.*, 2017); BD is used to process structured and unstructured data and to send it to informational interfaces(Rahul and Banyal, 2020). Big data in health it is also known as big medical data; is important in making decisions from public or private hospitals; in (Chen and Wu, 2020) it is stated that medical centers generate information assets with high volume, speed, and diversity through medical care; diversity is in type and form, there are unstructured and complex data. The healthcare industry is based on the dimensional organization of quality data; the high dimension of medical data is achieved through management and control; analyzing the volume of medical data helps doctors make good decisions (Lv and Qiao, 2020). IoT in hospitals offers the great advantage of having people connected to objects in one environment and takes advantage of distributed environments (Chen and Wu, 2020); IoT is a scheme of strategic interconnections at the computer level, digital and mechanical hardware to capture and transmit data over a network (Pratap, Javaid and Haleem, 2020). Different IoT proposals for medical environments, the data is produced into information for the benefit of patients; here, technology and health converge (Vazquez *et al.*, 2020).

The problem is the lack of proposals to have hardware and software infrastructures based on scientific research that assist in the analysis and processing phases to make decisions in medical organizations based on large volumes of data. Why review hardware and software infrastructures that generate large volumes of data for medical entities and serve analysis, processing, and decision-making? To understand the architectures that generate and capture the data produced by the internet-connected hardware, understand the processing of massive data used in the medical area, understand alternatives for the analysis of medical data.

The main objective of this research is to propose the hardware and software infrastructure for analysis, processing, and decision-making in medical entities using large volumes of data.

MATERIALS AND METHODS

Several methods were used to achieve the objective. *Analytical method:* Allows to analyze the characteristics of Big Data and research in IoT; to learn about work on these technologies that collect and process medical data; in addition to understanding the different models or architectures of other researchers. *Inductive-deductive method:* These logical reasoning strategies, the first is used to go from principles to verifications proposed in this research; the second is used from general principles towards specific outcomes of knowledge generated in this research. *Observation method:* Used to understand Big Data and IoT environments in the medical area and propose an alternative. *Bibliographic Analysis:* Used to conduct this research from 2016 to 2020 from high-impact databases such as IEEE Xplore and Scopus. *Quasi-experimental method:* Used for the proposed architecture, it is based on the components of the IoT and Big Data model; its components include devices, services, databases, database columns, sessions, networks, health professionals, and health indicators.

The following references are work-related to Big Data and software in the medical area: some barriers to implementing BD in health centers (Chen and Wu, 2020); data selection (Shakhovska *et al.*, 2019); BD analysis tool was used in several projects and organizations (Nadal *et al.*, 2017); analytical model in real-time for access to medical services (Xia *et al.*, 2019); a framework for medical professionals (Istephan and

Siadat, 2016); BD architecture has a metadata and data analysis (Holom *et al.*, 2020); analysis for medical data (Pramanik *et al.*, 2020); bioinformatics data management software (Yang, Troup and Ho, 2017); forecast in the acquisition of medical equipment (Xu and Chan, 2019). The following references are IoT-related work in the medical area: In a neural network for diagnosis through analysis of images (Yao *et al.*, 2019); the lifecycle of IoT-generated data is creation, storage, utilization, sharing, archiving, and destruction (Rahul and Banyal, 2020); applications on covid-19 (Pratap, Javaid and Haleem, 2020); an IoT architecture that works with medical sensors (Fouad *et al.*, 2020); IoT is used in areas such as cities, industries, health, households, vehicles, tools, clothing, others (Nižetić, Šolić and Patrono, 2020); the design of a holistic multilayer IoT architecture for healthcare providers (Farahani and Firouzi, 2018); real-time sensor monitoring to provide data to healthcare professionals (Saheb and Izadi, 2019); medical devices that encrypt data (Rao and Clarke, 2020); real-time patient monitoring through sensors and storage computing (Mohanta *et al.*, 2020).

RESULTS

General IoT model in the health sector

A general IoT model was proposed to manage the data produced and the information used for decision making (see Figure 1).

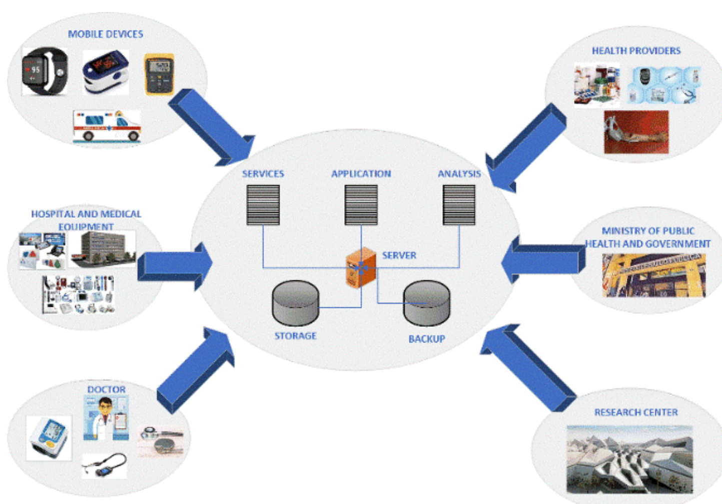

Figure 1. General IoT model in the health sector.

In one segment are mobile devices used by people on continuous movement or are at home; among the devices in this segment are electronic stethoscopes, tensiometer, glucometer, and others are in an ambulance and can send data to the network. In another segment are hospitals and medical equipment. Here are stationary or mobile devices belonging to the hospital such as cardio equipment, electrosurgical equipment, lights, diagnostic equipment, anesthesia artifacts, sterilization artifacts, video resolution equipment, and images. In another segment is the doctor or health professional, here among the devices are the doctor's own devices. In another segment are health care providers; here are private medical centers, health care providers, and health insurance agencies. In another segment is the institution that regulates health services by the government; it manages the political, legal, preventive, and curative part of the health environment in a country; short-term or long-term health policy projects. Another segment is the research center; one or more raises strategies at the national level for disease analysis and public health research. At the center is the cloud that collects, stores, and processes IoT data; services are available for software that requests or delivers resources; applications process data; the analysis performs the transformation of data into management indicators; it is also necessary to maintain support for indicators.

General IoT architecture for the health sector

A five-layer architecture is required for the model to be supported (see Figure 2); to capture, process, and analyze data between different layers; not only are there devices but there is also integration with other layers, applications are cloud-based for online tracking; cloud usage is proposed because it is a distributed architecture, good processing speed, scalable storage and control in information security; some research suggests using Computing Edge although security is still fragile (Mohanta *et al.*, 2020).

Device layer: Here are the mobile or stationary devices of people, hospitals, health centers; these capture patient health data; there are two types of sensors; the first type is physical sensors can be wireless or wired, among these: heart monitors, hemoglobin or glucose, pulse meters, wireless clothing, among other sensors; the second type is virtual sensors that use applications for tracking, consultation, diagnosis and health records (Farahani and Firouzi, 2018). *Network Layer:* Uses interfaces compatible with personal networks and wireless networks such as RFID, BLE, Zigbee, Wi-Fi, 3G/4G, and Ethernet; other wired protocols related to Arduino, serials, and peripherals (Farahani and Firouzi, 2018). *Service and Storage Layer:* MapReduce, Spark, or Hadoop is applied for the analysis of data transmitted online, this is significant in data management and analysis; IoT is leveraged for the analysis of large volumes of data in parallel, distributed, and cloud computing across clusters, parallel processors, and different types of networks (Saheb and Izadi, 2019).

Figure 2. General IoT architecture for the health sector.

Application layer: It can use third-party platforms to have computing applications such as Azure IoT Suite, IBM Watson, Amazon Web Services, Oracle IoT, Kaa, Bevywise, IoTIFY; these are scalable, open-source platforms for processing all the information (Mohanta *et al.*, 2020). *Dashboard layer:* This layer features serialized diagnostics, time analysis, standard image deviations, standard sample deviations, sample ranges, image data segmentation, apply filters, anomaly ranges (Pramanik *et al.*, 2020).

Big Data General Architecture for Health Sector

A proposed general data management scheme for the health sector in four layers (see Figure 3).

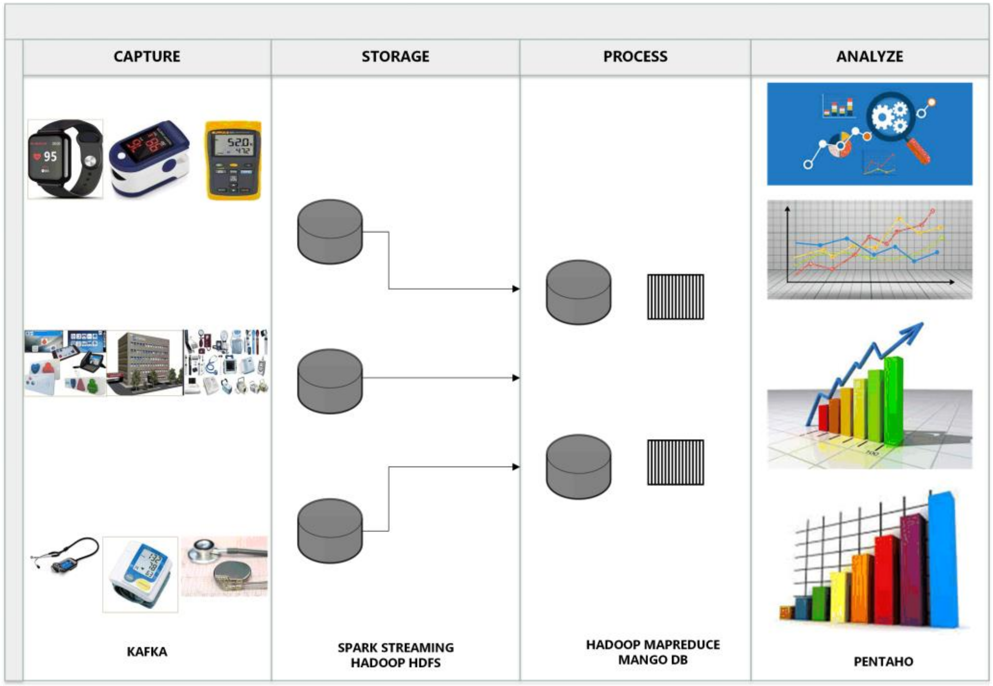

Figure 3. Big Data General Architecture for Health Sector.

The first layer of data capture: Takes data generated by IoT networks from mobile users, homes, hospital devices, healthcare professional devices; these are uploaded to the cloud in different formats according to the device; Kafka software is used here for capturing and sending data flow. Kafka helps in the processing and transmission of data in real-time. That is, it transforms the data into readable or understandable formats; the data of the devices is pre-processed, the sending of the data to the warehouse that is in the cloud is facilitated; it also facilitates channeling and generation of datamart. *Second data storage layer:* Performs the transformation of captured data into data stored in a heterogeneous way; is the cloud space for data occupancy; here, Hadoop HDFS software is used for managing large volumes of distributed data. *Third data processing layer:* Deals with data transformation and datamart collection; here Spark Streaming is used to process data, is scalable, good performance and fault-tolerant; Hadoop MapReduce is used for process management over large volumes of distributed data; the data is stored in MongoDB this is a non-SQL database to be able to save any type of data. *The fourth layer of data analysis:* health indicators are displayed through any means of communication or device; here, Pentaho Data Integration is used to obtain indicators and make a decision according to the results presented in this layer.

DISCUSSION

The IoT model in the health sector contains participants in data generation, data usage, and decision-makers; relates to the IoT architecture because these participants are distributed across one of the layers and have some role or responsibility for the data; the IoT architecture relates to the Big Data architecture at the intersection of patients and doctors, here patients are the data generators, and the doctors are the decision-makers, based on the large amount of data stored and processed. The results of this proposal are widespread; a hardware and software infrastructure can vary between medical organizations due to characteristics such as the physical extension of the organization, number of medical personnel, number of patients served, areas of medical specialties, financing for the project, among others. This proposal does not consider data mining for data volume analysis; no medical device pricing, IoT infrastructure, storage, software license, deployment times are considered. The results of this proposal are consistent with the results of the references in health proposals in Big Data, and IoT are (Yao *et al.*, 2019; Chen and Wu, 2020; Rahul and Banyal, 2020); health proposals in Big Data are (Istephan and Siadat, 2016; Yang, Troup and Ho, 2017; Shakhovska *et al.*, 2019; Xia *et al.*, 2019; Holom *et al.*, 2020; Lv and Qiao, 2020); proposals for health reviews in Big Data are (Nadal *et al.*, 2017; Pramanik *et al.*, 2020). The theoretical consequence of this research may be the optimization of the proposed results for possible implementation in a medical organization.

FUTURE WORKS AND CONCLUSION

As future work, a tool approach on Big Data for decision-making in the medical area was proposed. It was concluded that Big Data and IoT are complemented by data lifecycle management in the capture, storage, processing, and analysis; this management is a conceptual proposition so that other researchers can deepen the design; the physical components of the architecture influence low performance, in part software components assist in high performance. There are challenges in the unstructured data that devices collect, in converting that data for cloud saving, and in improving the way useful information is a presentation for different stakeholders.

Acknowledgments. This work has been supported by the GIIAR research group and the Universidad Politécnica Salesiana.

REFERENCES

Chen, P. T. and Wu, W. N. (2020) ‘Big data management in healthcare’, *Inf Management*, 53(10–2018), p. 102078. doi: 10.1016/j.ijinfomgt.2020.102078.

Farahani, B. and Firouzi, F. (2018) ‘Towards fog-driven IoT eHealth’, *Computer Systems*, 78, pp. 659–676. doi: 10.1016/j.future.2017.04.036.

Fouad, H. *et al.* (2020) 'Analyzing patient health information based on IoT sensor', *Measurement*, 159, p. 107757. doi: 10.1016/j.measurement.2020.107757.

Holom, R. M. *et al.* (2020) 'Metadata management in a big data infrastructure', *Procedia Manufacturing*, 42, pp. 375–382. doi: 10.1016/j.promfg.2020.02.060.

Istephan, S. and Siadat, M. (2016) 'Unstructured medical image query using big data', *Biomedical Informatics*, 59, pp. 218–226. doi: 10.1016/j.jbi.2015.12.005.

Lv, Z. and Qiao, L. (2020) 'Analysis of healthcare big data', *Future Generation Computer Systems*, 109, pp. 103–110. doi: 10.1016/j.future.2020.03.039.

Mohanta, B. K. *et al.* (2020) 'Survey on IoT security', *Internet of Things*, 11, p. 100227. doi: 10.1016/j.iot.2020.100227.

Nadal, S. *et al.* (2017) 'A software reference architecture for semantic-aware Big Data systems', *IST*, 90, pp. 75–92. doi: 10.1016/j.infsof.2017.06.001.

Nižetić, S., Šolić, P. and Patrono, L. (2020) 'Internet of Things (IoT)', *Cleaner Production*, 274, p. 122877. doi: 10.1016/j.jclepro.2020.122877.

Pramanik, M. I. *et al.* (2020) 'Healthcare informatics and analytics in big data', *Applications*, 152, p. 113388. doi: 10.1016/j.eswa.2020.113388.

Pratap, R., Javaid, M. and Haleem, A. (2020) 'IoT applications to fight against COVID-19 pandemic', *Diabetes & Metabolic Syndrome*, 14(4), pp. 521–524.

Rahul, K. and Banyal, R. K. (2020) 'Data Life Cycle Management in Big Data Analytics', *Science*, 173(C), pp. 364–371. doi: 10.1016/j.procs.2020.06.042.

Rao, A. R. and Clarke, D. (2020) 'Perspectives on emerging directions in using IoT devices', *Internet of Things*, 10, p. 100079. doi: 10.1016/j.iot.2019.100079.

Saheb, T. and Izadi, L. (2019) 'Paradigm of IoT big data analytics in the healthcare industry', *Informatics*, 41(February), pp. 70–85. doi: 10.1016/j.tele.2019.03.005.

Shakhovska, N. *et al.* (2019) 'Big data analysis in development of personalized medical system', *PCS*, 160, pp. 229–234. doi: 10.1016/j.procs.2019.09.461.

Vazquez, M. Y. L. *et al.* (2020) 'Design of an IOT architecture in medical environments', *RISTI*, 2020(E33), pp. 188–200.

Xia, T. *et al.* (2019) 'Measuring spatio-temporal accessibility through big GPS data', *Health*, 56(04–2018), pp. 53–62. doi: 10.1016/j.healthplace.2019.01.012.

Xu, S. and Chan, H. K. (2019) 'Forecasting medical device demand', *Manufacturing*, 39, pp. 32–39. doi: 10.1016/j.promfg.2020.01.225.

Yang, A., Troup, M. and Ho, J. W. K. (2017) 'Scalability and Validation of Big Data', *Biotechnology Journal*, 15, pp. 379–386. doi: 10.1016/j.csbj.2017.07.002.

Yao, C. *et al.* (2019) 'Gallstones with big data in medical Internet of Things', *Computer*, 94, pp. 140–147. doi: 10.1016/j.future.2018.11.011.