Big Data Integration with IoT to Achieve the Challenges

V. Bhagyasree[1], K. Rohitha[2], K. Kusuma[3] and S. Kokila[4]
[1,2,3&4]Assistant Professor, Department of Computer Science and Engineering,
Sreenivasa Institute of Technology and Management Studies, Chittoor, Andhra Pradesh, India
E-Mail: sreev.bhagyasree@gmail.com

Abstract - **The Internet of Things anticipates the combination of physical gadgets to the Internet and their access to wireless sensor data which makes it useful to restrain the physical world. Big Data convergence has many aspects and new opportunities ahead of business ventures to get into a new market or enhance their operations in the current market. The existing techniques and technologies is probably safe to say that the best solution is to use big data tools to provide an analytical solution to the Internet of Things. Based on the current technology deployment and adoption trends, it is visioned that the Internet of Things is the technology of the future; while to-day's real-world devices can provide best and valuable analytics, and people in the real world use many IOT devices. In spite of all the advertisements that companies offer in connection with the Internet of Things, you as a liable consumer, have the right to be suspicious about IOT advertisements. This paper focuses on the Internet of things concerning reality and what are the prospects for the future**.
Keywords: **Internet of Things, Big Data**

I. INTRODUCTION

Convergence wireless communications, Digital electronic devices, and Micro-electro-mechanical systems (MEMS) technologies led to the rise of the Internet of Things. Internet-connected objects like computers, smart phones, tablets and Wi-Fi de- vices, sensors, wearable devices and household appliances are all the objects of the IOT components [1] and considered as "Things". The Internet of Things means the production of tremendous amount of data and a collection of substantial different data bulk that has not seen so far. Big data and producing smart data are the research interests of the companies which produces these data. The vast volume of data generated by the Internet of things are handled by application of big data analytics and adds to the overhead of any organization and considered to be one of the most significant obstacles towards the deployment of this technology. The explosive growth of data has a raised in Internet usage, along with smart phone and social programs and machine-to-machine (M2M) communication [3].

The primary objective is the design of a model to analyze big data. In other words, we need to change our view about the blocks created by the Internet of Things [4]. Rather of looking at data as a data ware house, we should look at the supply chain. Since the tools are enabling the extraction of various unstructured and new data sources, the lack of sufficient data issue will reveal soon. Therefore, must overcome the following two fundamental problems [5]. Primarily not to miss the data that truly needed and as well, make sure not to spend time on unnecessary data. Despite the supply chain, organizations can fill in existing gaps in any way they need [6] .To this end, companies can take advantage of these three strategies:

1. Design interfaces for applications that created before.
2. Inquiry helps from partners or third parties who can provide the required data.
3. Develop data with the commission of the physical environment around the business.

Possibly, getting the accurate data is not the only problem with organizations. The other test is to acquire the necessary skills in the field of analytical analysis to deal with the big data. Popular skills in the domain of data analysis on the Internet of Things have not been practical. Companies need people who know analysis, as well as understanding then meaning of the initial data for a specific industry. Along with the growing trend in data generation and analysis required for these big data, organizations are forced to prepare devices that connect customers and objects at any time and any place [7].

One of the critical supports required by active companies in the Internet of Things is having the culture of data-driven decision-making. The Internet of Things, in aspect, provides a flow of accurate data derived from the real world. These data must go beyond the process of transforming data into information, then knowledge and awareness, and ultimately wisdom, using traditional analytical skills in this area to be meaningful. For example, in the field of agriculture, an experienced scientist needs to know how much irrigation required producing a product under different climatic conditions [8].

In intermittent IoT, the ability to collect weather information, field, and product in- formation is automatic and accurate. Anyhow, when the data collected, the actions to be taken on data depend on the expert opinion. Coupled with the growing amount of data and analysis needed, companies need to be ready for a range of devices that connect consumers and objects at any point and any time. Those who accept the idea of data supply chain will go through these waves of information without dwelling in detail. Apart from the cases mentioned, the contribution required in the field of sensors, analytical capabilities, and

data security and support are among the other obstacles faced by the technology of the Internet of Things [9].

Hitherto, all things considered the existing techniques and technologies, it is probably safe to say that the best solution is to use big data tools to provide an analytical solution to the Internet of Things. Based on the modern technology deployment and adoption trends, it is envisioned that the Internet of Things is the technology of the future; while today's real-world devices can provide absolute and valuable analytics, and people in the real world use many IoT devices. Despite all the advertisements that companies offer in connection with the Internet of Things, you as a liable consumer, have the right to be suspicious about IoT advertisements. The fundamental question is: What is the promise of the Internet of things concerning reality and what are the prospects for the future?

The core value of an IoT system is the ability to analyze the data needed and achieve practical and use full insight without making any mistake. Hence for two reasons, creating a communication medium is not easy. Firstly, it must have the ability of scalable analysis. Secondly, must make comprehensive usage of this ability regarding the volume and speed of the IoT devices that generate their data [10]. In this research, we discussed different solutions to help everyone to stay away from a series of issues on how to develop an ideal analysis of big data.

II. INTERNET OF THINGS AND ITS IMPACT ON BIG DATA

Current, with the help of the Internet of Things technology, the ability to connect each object to the network is provided. The Internet of Things offers a chain of connected people, objects, applications, and data over the Internet for remote control, interactive, services integration and management. Hence this network is overgrown; we need a platform that can collect and store the data generated by IoT devices.

Some of the progressive in Internet of Things services require a mechanism to collect analyze and process raw data from sensors to be used as operational control information. Some types of sensor data may have every huge volumes because of the signify cant number of sensors in the Internet of Things ecosystems. Possibly, we would ee a data flow coming from these devices. Appropriately, we need new technologies or architectural patterns in the area of data collection, storage, processing, and data retrieval [11].

Databases designed and implemented to attempt with the Internet of Things have their specific conditions and characteristics. The generation of NoSQL technologies can be considered as an indication that the management of the Internet of Things requires the use of novel approaches in administering and utilizing databases. The arrangement of cloud computing platforms based on the internet of things

eases the opportunities to enter this arena and take advantage of its achievements and services for many businesses of various dimensions [12]. Moreover, there will be a need for IoT data analysts of an acquaintance and entry into the fascinating world of big data.

The Internet of things changes people, processes, data, and things.
1. *People:* More objects can be monitored and controlled, and subsequently in- creased individual's abilities.
2. *Processes:* Users and more machines will be capable to interact with each other in real time. Therefore, very complicated tasks can finish in less time as the percentage of engagement and participation in doing a job is far more significant.
3. *Data:* The capability to collect data at a higher frequency and reliability provided which can lead to a correct decision making.
4. *Things:* The capability to control things more accurately. Hence, the value of objects such as mobile devices will be more and can help with much more than the current situation.

Big data concurrence, super-efficient networks, social media, low-cost sensors, and a new generation of advanced analytics have provided countless new opportunities for business enterprises to use the to either enter an market or strengthen their operations in the current market. The Internet of Things is one of these new markets that can afford countless opportunities for businesses in different fields. Important changes happen by a slight difference, and the Internet of Things can be the source of millions of changes in different areas over the next few years. Examine the Internet of Things as one of the causes for generating data which its impact on IT infrastructure and the use of advanced methods in data analysis are among its exceptional and vital opportunities in this regard[13].

A selection, preparation, and analysis of large volume of data will not be an easy task. Essentially, the amount of data can be doubled in several months and secondly, the gendered nature of this kind of data has its particular complexities. The collection in the template or the format of this type of data is extensive and often includes hundreds of pseudo-structured forms or unstructured formations.

Mostly to achieve a broad view of the sensor data, it should be possible to analyze and manage every structured and unstructured data. An analysis based on a specific data format can significantly limit the created potential insight. Data analysis, disregarding of its composition, is centralized and side by side, which provides a comprehensive analytical perspective to decision makers of business enterprises.

It expresses the consideration of the limitations and stock of traditional enterprise data and current business intelligence software and design. Organizational data ware- houses are not able to focus on disorganized data. Accordingly, we

need to look for explanations that enable unstructured data storage and analysis. If we want to convert the unstructured data into a specific structure by defining a particular structure and using relational database tables, we will lose time.

Consequently, that will surely be possible with the condition of not having the technical limitations. The use of any technology to create an analytical infrastructure that has some limitations can reduce the ability to analyze and, in practice, minimizes the potential for possible value creation. Analyzing of big data with the help of related technologies can be one of the leading actors in this field. Analysis of big data in some cases can help us

1. Combine, integrate and analyze all structured, semi-structured and unstructured data regardless of source, type, size, and format.
2. A quick and cost-effective analysis of the high volume data to create an appropriate insight into a decision making process.

The Internet of Things has agreed to influence various industries from insurance companies and banks to telecom and other business enterprises. Organizations need to correct data analysis methods so that they can collect, clean, prepare and analyze RFID sensor and tag data in the short time possible. Deciding on the actual data and in the shortest reasonable time is the confidentiality of a business firm in today's highly competitive environment. With a proper big data management and the creation of an appropriate atmosphere for their analysis, an organization's vision for proper decisions making increases [15].

III. BENEFITS OF IOT BASED ON BIG DATA

In literature, different structures for big data analysis and IoT proposed, which can manage the challenges of storage and analysis of high volume data from intelligent buildings. The first given structure consists of three main components which are big data management, IoT sensor, and data analysis.

These analyzes use are in the real- time management of oxygen level, fatal gases/soot and the amount of ambient light in smart buildings. Along with smart building management, IoT devices and sensors for receiving traffic information can be used in real time traffic management with low cost and examine the strengths and weaknesses of existing traffic systems.

In smart city management, the big data used in the search of data which obtained from different sensors such as water sensors, transportation network sensors, monitor- ing devices, smart home sensors and smart car park sensors. These data are developed and processed in a multi-stage model and ultimately reached a decision-making stage. These steps are data generation, data collection, data integration, data categorization, data processing and decision making [16].

Sometimes it is essential to pay attention to the concepts of web technology in particular proposed framework to investigate the analytical results obtained from the big data in the Internet of Things. In the literature, this topic has formulated, and a conceptual framework has been proposed consisting of 5 layers

1. *Data Collection Layer:* collected data from different sources, the input layer is the proposed framework.
2. *Extract-Transform-Load (ETL) Layer:* Implements the ability to change the format of information received from different types of sensors into a defined format.
3. The semantic reasoning rules layer - an inference engine that acts on the in- formation received from the ETL layer
4. *Learning Layer:* From the data made to measure to the existing extraction data, extract the various specifications and attributes, and finally, Machine learning-based models provided.
5. *Action Layer:* executes a set of predefined actions by the outputs of the learning layer.

Other applications of IoT help with geographic information analysis, cloud computing flow processing, big data analysis, and storage, cloud computing security, cluster- ing mechanisms, health, privacy security, performance evaluation of monitoring algorithms, manufacturing systems, and energy development[17].

IV. BIG DATA COLLECTION AND STORAGE

Various protocols make it possible to receive events from IoT devices, especially at lower levels. It does not matter if the device connected to a Bluetooth network, cellular network or Wi-Fi, or communicates through a hardware connection, is enough to send a message from a broker using a designated protocol. One of the most famous protocols for large IOT applications is MQT elementary Transport (MQTT). MQTT assigns to the transmission of messages through remotes sensing and queuing which is an M2M IOT connection. This protocol designed as a very slight request/ response (point- to-point) messaging transfer. MQTT is practical and useful for connecting to distant locations that require low memory or low network bandwidth; For example, this protocol used in sensor communication through a satellite link with a broker in dial-up connections with health care providers at different times and in a range of home automation and small devices. Its design principles are to minimize network bandwidth and resource requirements, and at the same time, it also guarantees trust and confidence in the message delivery [18]. There are also other substitutes, such as the limited application protocol, XMPP, and other protocols. Constrained Application Protocol (COAP) is a software protocol that used in straight forward electronic devices which provides communication through the Internet interactively. Constrained Restful Environments (CORE) group along with Internet Engineering Force (IETF) did the main standardization work of this protocol.

Extensible Messaging and Presence Protocol (XMPP) is a connection protocol for the Extensible Markup Language (XML). The XMPP set up exchanging close to real time between structured data but expandable between either any two or more net- work entities. We recommend starting with MQTT due to it's the availability and extensive coverage and the availability of a large number of client applications and open source brokers unless having a convincing reason to choose another protocol.

Mosquitto is one of the MQTTs extensively used open source and will be the definitive choice for applications; if this concept should be proved based on limited budgets and want to avoid the cost of committed devices ,the fact that Mosquitto is open-source brokerage is precious. Regardless of what protocol to choose, it will eventually receive messages that represent events and observations from devices connected to the Internet.

As in considering the message received by a broker (Mosquitto), you can send it to the analytics system. The finest way is to store source data before any transfer or making any changes to them; this is, of course, worthwhile when debugging problems occurs at the conversion stage.
There are certain ways to store IoT data which many uses Hadoop and Hive in their projects. Newly researchers are and successfully working with NoSQL databases such as Couch base. Couch base provides a appropriate combination of high performance and low latency indicators. Couch base is a Document-Oriented Database which lacks a layout that covers a significant amount of data and also can add a variety of new events quickly [19].

Direct data writing on HDFS is another good solution, especially if using Hadoop and batch analysis is considered as part of the analytics workflow. For writing source data in a permanent storage location, the direct code can be added manually to the message broker at the level of the IoT protocol (for example, if you use MQTT, Mosquito Broker). The other way is to add messages to a medium-sized middleware broker like Apache Kafka and use different Kafka users to transfer messages to different parts of the system [20]. One of the well-established patterns is to place messages in Kafka in two user groups based on the subject, where one of the groups writes raw data to their permanent storage, while another, transmits data into area time processor engine like Apache Storm. If Storm is used instead of Kafka, a Bolt processor can fix in topology, which does nothing except sending messages to a permanent storage location. If MQTT and Mosquitto are used, sending direct messages to the Apache Storm topology via the MQTT's spout source is an easy way to link things together.

V. CONCLUSION

The advancement of IoT devices, smart phones, and social media provides decision makers with opportunities to extract valuable data about users, anticipate future trends and fraud detection. With the creation of clear and usable data, big data can create the organizations' values, make the changes clear and expand their performance. The use of data produced from the IoT and the analytical tools creates many opportunities for organizations. These tools use predictive modeling technologies, clustering, and classification to provide data mining solutions.

IoT improves the decision-making bits of decision-makers. The development of IoT and related technologies, such as cloud computing, provides the ability to remove data sources in different domains. Commonly, any data is considered useful in the domain itself, and data on shared domains can be used to provide different strategies. Machine learning, deep learning, and artificial intelligence are crucial technologies that are used to provide value-added applications along with IoT and big data in addition to being used in a stand-alone mode. Before the advent of IoT and cloud computing, the use of these technologies was not possible due to the high amount of data and required computational power. Different data analysis platforms, Business intelligence platforms and analytical applications are emerging platforms that have been introduced to help industries and organizations in transforming processes, improving productivity, and the ability to detect and increase ability. It is foreseen that the speed of technological progress in the next ten years, will be equal to the past thirty years. Therefore, we have to use all our attempts to update our lives to the Internet of Things technology regarding hardware and software.

REFERENCES

[1]   Guoru Ding, "Big Data Analytics in Future Internet of Things", *Research Gate*, 2013.
[2]   Sunghae Jun, "Technology Analysis for Internet of Things using Big Data Learning", *International Journal of Research in Engineering and Technology*, Vol. 03, No. 12, 2014.
[3]   Cyril Cecchinel, Matthieu Jimenez, S'ebastien Mosser, and Michel Riveill, "An Architecture to Support the Collection of Big Data in the Internet of Things", *IEEE, 2014.*
[4]   Mereena Thomas, "A Review paper on BIG Data", *International Research Journal of Engineering and Technology*, Vol. 02, No. 09, 2015.
[5]   D. P. Acharjya, "A Survey on Big Data Analytics: Challenges Open Research Issues and Tools", *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 2, 2016.
[6]   Vandana Sharma, and Ravi Tiwari, "A review paper on "IOT & It's Smart Applications", *International Journal of Science Engineering and Technology Research (IJSETR)*, Vol. 5, No. 2, February 2016.
[7]   Zeinab Kamal, Aldein Mohammeda, and Elmustafa Sayed Ali Ahmedb, "Internet of Things Applications Challenges and Related Future Technologies", *World Scientific News*, Vol. 67, No. 2, pp. 126-148, 2017.
[8]   Mohsen Marjani, Fariza Nasaruddin, and Abdullah Gani, "Big IoT Data Analytics: Architecture Opportunities and Open Research Challenges", *IEEE Access*, March 2017.
[9]   Amir Taherkordi, Frank Eliassen, and Geir Horn, "From loT Big Data to IoT Big Services", *IEEE, 2017.*
[10] Anup S. Polgavande, "Internet of Things (IoT): A Literature Review", *International Journal of Research in Advent Tech*, pp. 2321-9637, 2017.
[11] Mehdi Mohammadi, "Deep Learning for IoT Big Data and Streaming Analytics: A Survey", *IEEE Communications Surveys & Tutorials*, 2018.

[12] Pradnya, A. Hukeri, and P. B. Ghewari, "Review paper on IoT Based Technology", *International Research Journal of Engineering and Technology*, Vol. 04, No. 01, Jan 2017.

[13] Tanmaya Mahapatra, Ilias Gerostathopoulos, and Christian Prehofer, "Towards Integration of Big Data Analytics in Internet of Things Mashup Tools", *ACM, ISBN 978-1-4503-2138-9*.

[14] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A Survey", *Computer networks*, Vol. 54, pp. 2787-2805, 2010.

[15] V.C. Patil, K.A. A1-Gaadi, D.P. Biradar, and M. Rangaswamy, "Internet Of Things (IoT) and Cloud Computing for Agriculture: An Overview", *Proceedings of Aipa 2012*.

[16] A. Botta *et al.*, "On the integration of cloud computing and internet of things", *IEEE International Conference of Future Internet of Things and Clouds*, August 2014.

[17] "Big data: Preliminary report 2014", *ISO/IEC, 2015*.

[18] The industrial internet of things volume g1: reference architecture, 2017.

[19] B. Leuckert *et al*., "IoT 2020: smart and secure IoT", *White Paper IEC*, 2016, [online] Available: http://www.iec.ch/whitepaper/pdf /iecWP-IoT2020-Lr.pdf.

[20] M. Chen, S. Mao, Y. Zhang, and V. C. M. Leung, "Big Data: Related Technologies Challenges and Future Prospects", *Springer,* 2014.