Computer Engineering and Intelligent Systems
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol.8, No.1, 2017

The Internet of Things: A Main Source of Big Data Analytics

Parth Goel
Babaria Institute of Technology, Varnama, Vadodara, Gujarat, India

Dweepna Garg
Parul University, Limda, Vadodara, Gujarat, India

Abstract
IoT is regarded as a platform or a framework for the objects/devices to interact with one another in an electronically manner with the world around. Not only the humans can communicate with one another using system, but rather IoT had made possible for the system to communicate with each other too. It is being enabled by the presence of other independent technologies which make fundamental components of IoT. One of such technology is Big Data Analytics. Big data has emerged as a connecting point between the objects on the internet. Massive data is generated in day to day life and data mining has played a vital role in converting the data to information useful for the end users. But as the amount of data kept on increasing, it became difficult to extract useful information from it. It is when Big Data Analytics came into picture. In today's world, various sensors interact over a wireless network by exchanging huge amount of data with one another. It is with the help of IoT, designing of a good infrastructure for storing and managing such huge amount of sensor data became possible. It resulted in an easy search and utilization of sensor data by the users. This paper deals with the so called relationship between IOT and Big data. In particular, the focus of the paper will be in how the things in IOT generate massive data (Big Data) and if both combined then how it can lead to wonders in real world.
Keywords: Internet of things, Big Data Analytics, Hadoop, Big Data, Unstructured Data

1. Introduction

The term big data[17] refers to the large and complex datasets which are difficult to be processed by the traditional data processing techniques like batch processing. Analyzing, capturing, searching, storing, transferring, querying and updating from such large amount of data is difficult. To overcome such challenges of high volume, high velocity and high variety of data, Big Data came into picture. The 3V namely velocity, volume and variety became the sole of Big Data and it expanded to other complimentary characteristics of Big Data. Variability and veracity are the characteristics of Big Data besides velocity, volume and variety.

The term Internet of things[9] was first coined by Kevin Ashton in the year 1999 as one of the title of his presentation. Internet of Things (IoT) is the network of interconnected physical objects or "things" embedded with software, electronics, sensors, and network connectivity. It enables the objects or "things" to first collect and then exchange the data. The technology is growing itself at a faster rate from wireless sensor to nanotechnology and the development of IoT solely depends on such growing technologies. IoT mainly focuses on how to enable general objects or things to hear, smell, see the physical world around by themselves and connect them to share the data. The objects here include mobile phones, refrigerators, air conditioners, sensors, actuators, RFID tags and many more which, through unique addressing schemes, are able to interact with each other and cooperate with their neighbors to reach common goals. The core part of IOT is embedded system. It is an electro mechanical system designed to perform a specific function and is a combination of both hardware and software.

Various applications of IoT include smart home, mood lighting, smart waste management, smart shopping, smart eating, smart egg tray, smart propane tank, smart care of plants, smart walking, smart cities, smart water management, smart logistics, smart emergency control, smart agriculture, smart health and many more.

IoT is mainly driven by analytics[7], artificial intelligence and data lakes. Data lakes include sensors, actuators, communication, people and process. Practically, all such above mentioned IoT applications which generate the massive data and they are collected from vehicles, cameras and all such heterogeneous sensors which provides the data like video, voice, text and so on. Processing of such types of heterogeneous and unstructured data imposes a challenge. All the applications of IoT uses various devices like Web Hooks, mobile devices, cameras, and other devices from where we can operate the household things or others. Besides, being imposing the challenge, it also offers several ideas about improving the system.

2. IoT ARCHITECTURE

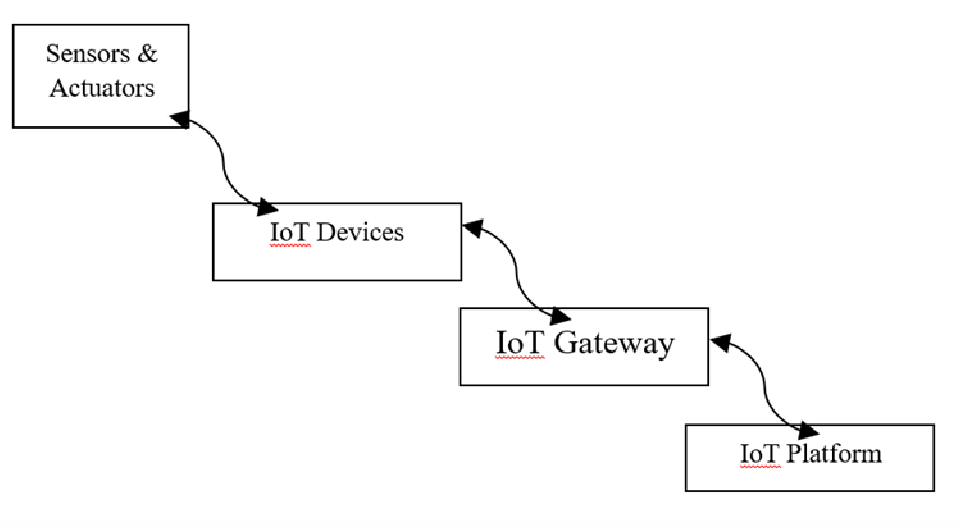

Figure 1: Architecture of IoT

Sensors[13] sense the changes i.e it accepts the input. It is found in a variety of applications like healthcare, smart mobile devices, automotive systems, climate monitoring and many more. Actuators get the output. Human body is the best example of sensors and actuators. It has a lot of sensors and actuators. The brain of the human body acts as a microcontroller and fuses the data collected from multiple sensors in order to get a reliable and accurate answer. Sensors are almost everywhere used. IoT gateway[14] performs various functions like data processing and filtering, security, device connectivity, updating and many more. Not only it performs the above mentioned critical functions but also operates as platforms for application code which processes the data and becomes a device intelligent system. Each of the devices uses different protocols such as MQTT[4], Zigbee, 6 low Pam, COAPP and others. Each of these protocols is connected to different control environment and has different models for security and management. MQTT is the highly used protocol in IoT. It is light weight and is publish/subscribe messaging protocol. It is an open standard and was developed by IBM. It is a client/server model where client is a sensor which is connected to the server known as broker, over TCP. The function of IoT gateway can be reframed as to aggregate the sensor data, translate between sensor protocols, processes the sensor data before sending it onward and more.

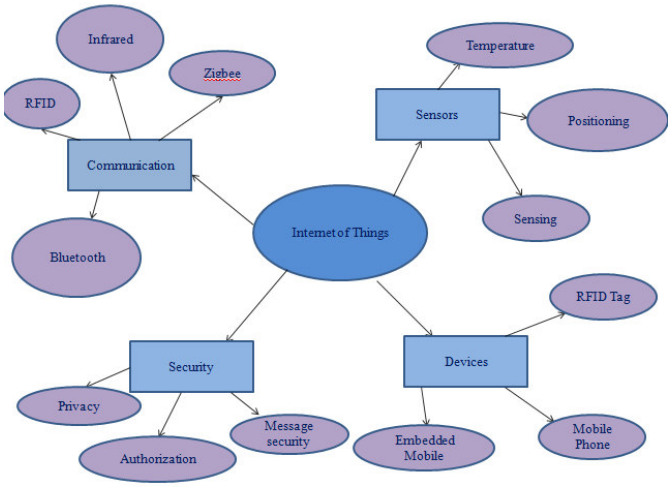

Figure 2: Basics of IoT

IoT platform[12] fills the gap between the data networks and device sensors. In recent times, lots of IoT platforms with various enterprise offerings are increasing the connectivity of devices in both private and public cloud. For successful implementation of IoT the interoperability among the devices should be strong and robust IoT platforms and frameworks should be built on top of it.

In earlier times, the people did not have the facility of using smart phones and take the advantage of smart devices. Due to the advancement in latest technologies like Android, cloud computing[11], big data analytics and many more. Today, there are more people connected through devices and a tremendous increment is shown in connected devices rather total population of the world.

3. Bond between Big Data and IoT

The question arises in how the conversion can be carried out in converting the data captured by IoT into fruitful knowledge. This is where big data analytics comes into play. Research is mainly focused in combining Big Data with IoT[2]. The massive amount of unstructured data generated by IoT devices can be converted to useful information with the help of Big Data. Unstructured data are everywhere. They include images, videos, sensor data, GPS, mobile phones, social media platforms, etc. Sparse data arise in situations where the sensor data does not seem to be initially event-driven. Say for example, many sensors measure the data with a fixed frequency but they should be able to output the data when the value has changed as compared to the last measurements. This is a one of the challenges one has to deal with. It is evident that the available tools of data analysis are not powerful enough to handle and analyze big data of IoT. The methods like divide and conquer, random sampling, incremental learning are not successful in dealing with big data of IoT as it leads to an increased in processing time and increased CPU utilization. One way to handle the big data of IoT is to make the sensors capture only the interested data. This can be done using the dimensionality reduction method like principle component analysis (PCA)[1] or by the reduction in the number of patterns by using the pattern reduction method. Another way is to make the data store on cloud platform. Cloud computing is one of the leading technologies today. Cloud computing is an interconnected remote machines which are hosted on the internet to collect, manage and process data rather than a local machine i.e. personal desktop, tablets, mobile phone, etc. Then from this data centre, the data can be analyzed using the Hadoop eco system. The Hadoop ecosystem consists of HBase, Zookeeper, Hive, Mahout etc. Apache Hadoop[18] has intended to deliver better performance and cost-effectiveness for processing of large clusters. One of the vital feature of Hadoop eco system is the high throughput which is able to handle the large scale data analysis and processing related problems. The two main big data infrastructure technologies include Spark and MapReduce.

Big data and IoT can be regarded as the two sides of the same coin. The data collected from the connected objects is high in quality, quantity and sensitivity i.e in terms of big data it can be stated that the data is large in variety, volume and veracity. Hadoop has the capability of handling such types of data. In short, it can be stated that Big Data is all about data and IoT is about data, devices/objects and connectivity. Industries like that of advertising, biotechnology, consulting have started believing that IoT is the most important for them. Mobile device support, in-memory analysis, location intelligence are the major areas where IoT has contributed with great efficiency and has integrated various systems in an effective manner. The ability of the organization to manage the big data is critically important to their success or failure with IoT. Even in predictive analysis, IoT has advocated significantly.

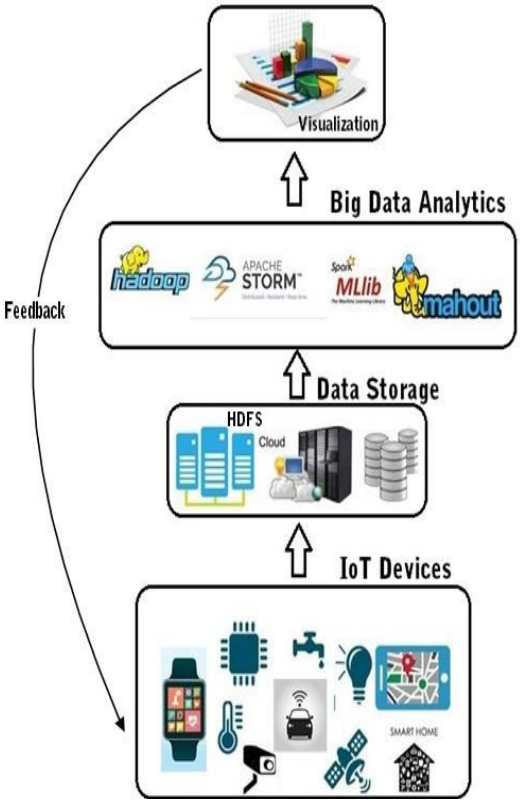

Figure 3: Process of IoT Data generation to analysis/visualization

4. Challenges in IoT

The key challenge is to analyse the big data and maintain the coordination between the database systems and the tools of data mining[3]. The data warehouse is generally used to manage a large amount of data. Extracting the useful information from such massive data is the challenging task.In recent years, Big data Analytics has been accumulated in several areas like surfing internet, social computing (which includes recommender systems, prediction market), retailers, health care, and other interdisciplinary scientific researches. Considering this advantages of big data it provides a new opportunities in the knowledge processing tasks for the upcoming researchers. IoT aims to integrate the information from the above mentioned areas. Even very simple things have RFID tags within them which help in tracking them without human intervention. Say, for example a person may have a smart watch worn which keeps the count of his/her heartbeat and other records like the steps walked, calorie burnt and other parameters. It could help the person to know what kind of food he/she should have and what type of routine he/she should follow to stay fit. The sensors and actuators are used in transforming the real world things to virtual object. In this manner, the things interact and collaborate over the Internet.

Other challenges[8] include heterogeneous hardware, Slow CPU, Little memory (of the order of 100KB), limited energy resources, robustness and self-organization and real time requirements like scheduling support. Big Data can be integrated with IoT to solve in solving various challenges. Say for limited energy resources, big data can be helpful by integrating itself to cloud. The data can be sent to cloud and using Amazon web services, we can have as many numbers of virtual machines as desired. Such virtual machine overcomes the challenge of IoT of heterogeneous hardware. Most of the data generated via this heterogeneous hardware is unstructured and are difficult to analyze. It also increases the processing time. So, appropriate strategies are developed to manage such data. Google is processing around 20,000TB of data daily. More than 6 billion people over the world text, call, browse and tweet on cellular devices. Thus, efficient management techniques and tools are required for the same. The new Big Data technology has improved the performance and has facilitated the innovations in the services and products in an efficient manner and had also helped in decision making. Big data aims to minimize the hardware and reduce the processing cost by effectively utilizing the resources. In order to manage data in an effective manner, various tools and techniques are used like Not only SQL (NoSQL), Google big table, Voldemort, MemcacheDB. For Big Data, the data management tools and techniques include Hadoop, MapReduce and BigTable.

The critical challenge is using the data when it is still in motion and extracting the valuable information from it[5]. The infrastructure of IoT is very complex due to the huge amount of data generated and consumed by billions of simultaneous occurring events which are geographically distributed all over the world. Difficulty in representing, interpreting, processing and predicting augments the infrastructural complexity of IoT. Parameters to be kept in mind are reliability, safety, fault tolerance and security. Apache Strom can handle all such complexities in an efficient manner.

5. Conclusion

In this paper, we reviewed the study on how big data analytics and IoT are depended on each other. The analysis depicted that how the unstructured data generated from various IoT devices resulted in useful information using Big Data analytics. There is no doubt that in the coming years, big data analytics will be the future trend as all the devices will upload the data to the internet. Most of the apps today are handling the handheld devices to control the household appliances. For the big data generated from IoT, cloud computing is also providing a powerful resource to do the computation. The area of research is mainly to focus in improving the accuracy for extracting the useful knowledge using big data analytics for IoT. Actual challenges faced in IoT are also discussed and relevant work can be carried out in that area. Using this knowledge, an efficient way for extracting and presenting the useful data can be carried out using IoT platforms.

References

[1] Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, and Laurence T. Yang: Data Mining for Internet of Things: A Survey, IEEE Communications Surveys & Tutorials, VOL. 16, NO. 1, FIRST QUARTER 2014.
[2] Daniel e. O'leary: "Big data", the "Internet of Things" and "The Internet of Signs": Intelligent systems in accounting, finance and management Intell. Sys. Acc. Fin. Mgmt. 20, 53–65 (2013), DOI: 10.1002/isaf.1336.
[3] Debi Prasanna Acharjya, Kauser Ahmed P: A Survey on Big Data Analytics: Challenges Open Research Issues and Tools, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016
[4] Debasis Bandyopadhyay , Jaydip Sen: Internet of Things - Applications and Challenges in Technology and Standardization, Nature: Wireless Personal Communications May 2011, Volume 58, Issue 1, pp 49–69
[5] Dr. Léon Bottou: How Big Data changes Statistical Machine Learning, 2015 IEEE International Conference on Big Data (Big Data)
[6] Feng Chen,1 , 2 Pan Deng,1 JiafuWan,3 Daqiang Zhang,4 Athanasios V. Vasilakos,5 and Xiaohui Rong6:

Data Mining for the Internet of Things: Literature Review and Challenges, International Journal of Distributed Sensor Networks Volume 2015, Article ID 431047, 14 pages, http://dx.doi.org/10.1155/2015/431047
[7] Guoru Ding, Long Wang, Qihui Wu: Big Data Analytics in Future Internet of Things: Science Direct article, November 2013, research gate
[8] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic,Marimuthu Palaniswami : Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions, Elsevier, http://dx.doi.org/10.1016/j.future.2013.01.010
[9] Luigi Atzori, Antonio Iera, Giacomo Morabito: The Internet of Things: A Survey, Elsevier, Computer Networks Volume 54, Issue 15, 28 October 2010, Pages 2787–2805, http://dx.doi.org/10.1016/j.comnet.2010.05.010
[10] Lu Tan, Neng Wang: Future Internet: The Internet of Things, IEEE, 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), DOI: 10.1109/ICACTE.2010.5579543
[11] Lizhe Wang, Rajiv Ranjan: Processing Distributed Internet of Things Data in Clouds, IEEE Cloud Computing (Volume: 2, Issue: 1, Jan.-Feb. 2015) DOI: 10.1109/MCC.2015.14
[12] Radu-Ioan Ciobanu, Valentin Cristea, Ciprian Dobre and Florin Pop: Big Data Platforms for the Internet of Things, A Roadmap for Smart Environments, Studies in Computational Intelligence 546, DOI: 10.1007/978-3-319-05029-4_1, © Springer International Publishing Switzerland 2014
[13] Sulayman K. Sowe, Takashi Kimata, Mianxiong Dong, Koji Zettsu: Managing Heterogeneous Sensor Data on a Big Data Platform: IoT Services for Data-intensive Science, 2014 IEEE 38th Annual International Computers, Software and Applications Conference Workshops.
[14] Internet of Things: Springer, Technology, Communications and Computing Series Editors: **Fortino**, Giancarlo, **Liotta**, Antonio, ISSN: 2199-1073
[15] Big Data and Internet of Things: A Roadmap for Smart Environments: Springer, Studies in Computational Intelligence, Editors: Bessis, Nik, Dobre, Ciprian (Eds.)
[16] Internet of things: Blog: https://blogs.microsoft.com/iot/, Accessed on December 2106.
[17] Big data blog: http://dataconomy.com/big-data-blogs/, Accessed on December 2106.
[18] Big data, Hadoop: http://www.michael-noll.com/, Accessed on December 2106.