Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 10, No. 3, 2023

The Relationship Between Big Data and IoT

Tran Anh Tu

Institute of Information Technology, Hanoi Pedagogical University 2, Vinh Phuc 15910, Viet Nam

Abstract: The evolution of Internet of Things (IoT) makes the life of human being very easier. IoT makes it easy to control any device in a bit. Several IoT applications like smart homes, manufacturing, transportation, and consumer goods like wearables, smartphones are available. IoT is actually the network of devices that contain sensors, electronics, software's, actuators, and connectivity which allows these things to connect, interact and exchange data. IoT made easy way to control the devices, collect the information from such devices, disseminate such information and transfer this information to analyze the same and to predict or prescribe the solution for the problems might exist. The large quantity of data is collected from IoT devices through the sensors where the Big Data came into picture. The paper reviews the relationship between IoT and Big Data.

Keywords: IoT; Big Data.

1. Introduction

IoT makes it possible to connect number of devices with each other and these devices are now linked to the internet, which transmits the data through sensors for the purpose of analysis. IoT devices are built to make a positive impact on our lifestyle, energy conservation, smart agriculture, transportation, and health. The data collected from these devices are used to learn more about trends and patterns that can be utilized. The data and IoT are closely interlinked with each other. These are closely intertwined and althoughthey are not the same thing, it is very hard to talk about one without the other. This paper tries to provide a basic understanding of the relationship between IoT and Big data.

2. Iot And Big Data

2.1. IOT

The Internet of things (IoT) describes the network of physical objects "things"—that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the Internet.

The Internet of Things (IoT), firstly coined by Kevin Ashton as the title of a presentation in 1999 [1], is a technological revolution that is bringing us into a new ubiquitous connectivity, computing, and communication era. The development of IoT depends on dynamic technical innovations in a number of fields, from wireless sensors to nanotechnology [2]. For these ground-breaking innovations to grow from ideas to specific products or applications, in the past decade, we have witnessed worldwide efforts from academic community, service providers, network operators, and standard development organizations [3]–[5]. Although IoT has created unprecedented opportunities that can help increase revenue, reduce costs, and ameliorate efficiencies, collecting a huge amount of data alone is insufficient. To generate benefits from IoT, enterprises must create a platform where they can collect, manage, and analyze a massive volume of sensor data in a scalable and cost-effective manner [6].

2.2. BIG DATA

Big data comes and is composed through electronics operations from multiple sources. It requires proper processing power and high capabilities for analysis [7]. The importance of big data lies in the analytical use which can help generate an informed decision to provide better and faster services [8].

The term big data is called on the huge amount of high-speed big data of different types; this data cannot be processed andstored in regular computers. The main characteristics of bigdata, called V's 5 As in Figure 1 , can be summed up in the fact that the issue is not only about the volume of data, other dimensions of big data, known as 'five Vs', are as follows:

Volume: It represents the amount of data produced frommultiple sources which show the huge data in numbersby zeta bytes. The volume is most evident dimension in what concerns to big data.

Variety: It represents data types, with, increasing the number of Internet users everywhere, smart phones and social networks users, the familiar form of data has changed from structured data in databases to unstructured data that includes a large number of formats such as images, audio and video clips, SMS, and GPS data [9].

Velocity: It represents the speed of data frequency from different sources, that is, the speed of data production such as Twitter and Facebook. The huge increase in data volume and their frequency dictates the need for a system that ensures super-speed data analysis.

Veracity: It represents the quality of the data, it shows the accuracy of the data and the confidence in the data content. The quality of the data captured can vary greatly, which affects the accuracy of analysis. Although there is wide agreement on the potential value of big data, the data is almost worthless if it is not accurate [10].

Value: It represents the value of big data, i.e. it shows the importance of data after analysis. This is due to the fact that the data on its own is almost worthless. The value lies in careful analysis of the exact data, the information and ideas it provides. The value is the final stage that comes after processing volume, velocity, variety, contrast, validity and visualization [11].

The type and nature of the data

Data in general is a set of values that are in the form of numbers, letters, symbols and other forms where they are concerned with a particular idea and subject. The data does not make sense without analysis, and is, therefore, compiled

for use. It represents input, while information is output after processing, i.e. data is entered into the system first, then processed until it comes out in the form of useful information that has a clear meaning and against which decisions are made.



Figure 1. Characteristics of Big Data

Big data comes from multiple sources including sensors and free texts such as social media, unstructured data, metadata and other geospatial data collected from web logs, GPS,medical devices, etc. [12]. The big data is gathered from different sources, so it is in several forms, including:

Structured data: It is the organized data in the form oftables or databases to be processed.

Unstructured data: It represents the biggest proportion of data; it is the data that people generate daily as texts, images, videos, messages, log records, click-streams,etc.

Semi-structured data: or multi-structured, it is regarded a kind of structured data but not designed in tables or databases, for example XML documents or JSON [13].

Difference between traditional data and big data

In general, the data in the world of technology is a set of letters, words, numbers, symbols or images, but with the evolution of multitasking technology tools the data has become different in content and source [14]. In light of this,big data emerged which differs from traditional data. Differences between traditional data and big data are shown in Table1:

Table 1. Comparison between traditional and big data [18]

	Traditional Data	**Big Data**
Volume	MB and GB	PBs And ZBs
Data Generation Rate	Long periods of time	More rapid
Data Type	Structure	Sim-Structure , Unstructured
Data sources	Centralized	multiple sources, and distributed
Data Store	RDBMS	HDFS, No SQL

2.3. Necessity of Iot And Big Data Implementation

IoT will enable big data, big data needs analytics, and analytics will improve processes for more IoT devices. IoT and big data can be used to improve various functions and operations in diverse sectors. Both have extended their capabilities to wide range of areas. The figure below shows the areas of big data produced. Some or the other way, data is produced through connected devices.

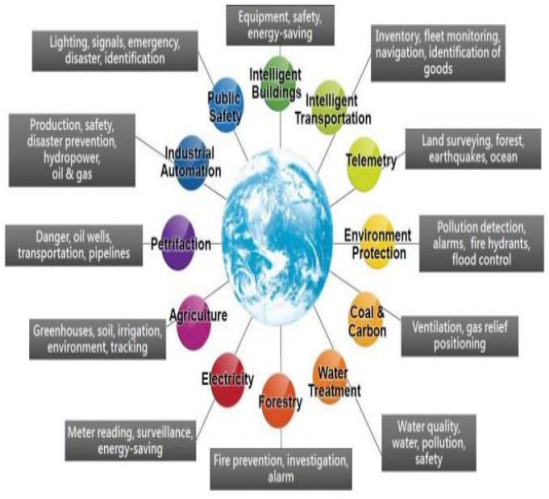


Figure 2. Big data produced across various areas

The important basis behind why to implement IoT and big data are:

a. *Analytical monitoring*

b. *More Uptime*

c. *Lower reject rates*

d. *Higher throughput*

e. *Enhanced safety*

f. *Efficient use of labor*

g. *Enable mass customization*

h. *Analyze the activities for real – time marketing*

i. *Improved situational alertness*

j. *Improved quality*

k. *Sensor – driven decision analytics*

l. *Process optimization*

m. *Optimized resource utilization*

n. *Instant control and response in complex independent systems.*

The above are some possible reasons to implement IoT and Big data. As the requirements of both the technologies go hand in hand, a proper improved system is needed to

overcome the challenges they pose. Many companies strive to meet the challenges and take possible steps to overcome them.

2.4. Impacts of Iot On Big Data

The main factors that big data is impacted by IoT are:

2.4.1. Big Data storage

At basis, the key necessities of big data storage are that it can handle very huge amounts of data and continuous balancing to keep up with expansion and that it can provide the input/output operations per second (IOPS) necessary to deliver data to analytics tools. The data is of different form and format and thus, a datacenter for storing such data must be able to handle the load in changeable forms. ObviouslyIoT has a direct impact on the storage infrastructure of big data. Collection of IoT Big Data is a challenging task because filtering redundant data is mandatorily required. After Collection, the data has to transfer over a network to a data center and maintained. Many companies started to use Platform as a Service (PaaS) to handle their infrastructure based on IT. It helps in developing and running web applications. By this way, Big data can be managed efficiently without the need of expanding their infrastructural facilities to some extent. IoT Big Data Storage is certainly a challenging task as the data grows in a faster rate than expected.

2.4.2. Data Security Issues

The IoT has given new security challenges that cannot be controlled by traditional security methods. Facing IoT security issues require a shift. For instance, how do youdeal with a situation when the television and security camera at your home are fitted with unknown Wi-Fi access.Few security problems are:

a. *Secure computations in distributed environment*

b. *Secure data centers*

c. *Secure transactions*

d. *Secure filtering of redundant data*

e. *Scalable and secure data mining and analytics*

f. *Access control*

A multi-layered security system and proper network system will help avoid attacks and keep them from scattering to other parts of the network. An IoT system should follow rigorous network access control policies and then allowedto connect. Software-defined networking (SDN) technologies should be used for point-to-point and point-tomultipoint encryption in combination with network identity and access policies.

2.4.3. Big Data analytics

Data analytics is the science of examining raw data with the idea of coming to conclusions about that information. Data analytics is used in many industries to allow them to make better business decisions and in the sciences to verify or disprove existing models or theories. IoT Big data analytics is very much needed to end up in a optimized decision. Big data analytics will help you understand the business value it brings and how different industries are applying it to deal with their sole business necessities. According to the Gartner IT dictionary, Big Data is variety of information assets, high-volume, and high-velocity and, innovative forms of information processing for enhanced approach and decision making.

Volume refers to the size of data. Data sources can be social media, sensor and machine-generated data, structured and unstructured networks, and much more. Enterprises are flooded with terabytes of big data.

Variety refers to the number of forms of data. Big data deals with numbers, 3D data and log files, dates, strings, text, video, audio, click streams.

Velocity refers to the speed of data processing. The rate at which data streams in from sources such as mobiledevices, click streams, machine-to-machine processes is massive and continuously fast moving. Big data mining and analytics helps to reveal hidden patterns, unidentified correlations, and other business information.

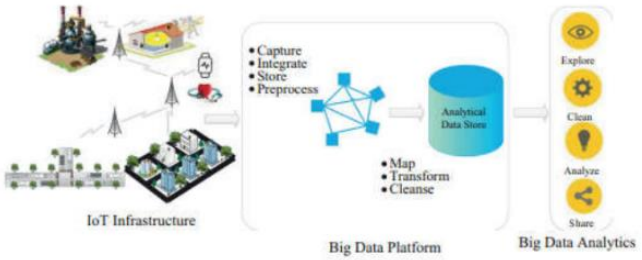


Figure 3. Big Data Flow and IOT

2.5. IOT APPLICATION WITH BIG DATA

2.5.1. *Smart* Agriculture*/ Smart Farm*

Smart agriculture is just getting underway and involves exploiting data frommany sources – sensors on farm equipment and plants, satellite images and weather tracking. The use of water and fertilizers will be measured and monitored in detail sometimes on plant by plant basis.



Figure 4. IoT Smart Farm issued by IoT.nxt

The IoT is set to push the future of farming to the next level. Smart agriculture is already becoming more commonplace among farmers, and high-tech farming is quickly becoming the standard thanks to agricultural drones and sensors. IoT sensors report weather conditions and monitor soil moisture and acidity while animal farmers track the movement and behavior of livestock remotely via embedded devices. Industrial IoT applications are also useful for monitoring indoor agricultural facilities such as silos, dairies and stables. IoT agriculture application areas include farm vehicle tracking, livestock monitoring, large and small field farming, and storage monitoring. Drones have become an invaluable tool for farmers to survey their lands and generate crop data. Farmers can use their smartphones to remotely monitor their

equipment, crops,and livestock, as well as obtain stats on their livestock feeding and produce. They can even use this technology to run statistical predictions for their crops and livestock.

2.5.2. Smart metering

Smart metering is one of the IoT application use cases that generates a large amount of data from different sources, such as smart grids, tank levels, and water flows, and silos stock calculation, in which processing takes a long time even on a dedicated and powerful machine [16]. A smart meter is a device that electronically records consumption of electric energy data between the meter and the control system. Collecting and analyzing smart meter data in IoT environment assist the decision maker in predicting electricity consumption. Furthermore, the analytics of a smart meter can also be used to forecast demands to prevent crises and satisfy strategic objectives through specific pricing plans. Thus, utility companies must be capable of high-volume data management and advanced analytics designed to transform data into actionable insights

2.5.3. Smart Transportation

Transportation Transportation today allows us to access public transit, shipping, ride sharing, and an unquantifiable amount of convenience. Whether by air, ground or sea, transportation and logistics are essential components to many enterprises' productivity, and access to real-time data iscritical. Many businesses have already discovered the advantages of using mobile technologies; however, the unpredictable nature of fuel costs, rising labor rates, increased traffic and a changing regulatory environment, continue to make operations challenging. With the advent of today's mobile technologies and the Internet of Things (IoT), enterprises can accelerate productivity, profitability and operations with solutions designed specifically for their processes. With the right IoT solution in place, enterprises can connect all devices across a centralized cloud network, and capture and share their mission-critical data, allowing them to gain real-time visibility of their operations.

2.5.4. Smart Supply Chains

Embedded sensor technologies can communicate bidirectionally and provide remote accessibility to over 1 million elevators worldwide [17]. The captured data are used by on and off-site technicians to run diagnostics and repair options to make appropriate decisions, which result in increased machine uptime and enhanced customer service. Ultimately, big IoT data analytics allows a supply chain to execute decisions and control the external environment. IoT-enabled factory equipment will be able to communicate within data parameters (i.e., machine utilization, temperature) and optimize performance by changing equipment settings or process workflow [18]. In-transit visibility is another use case that will play a vital role in future supply chains in the presence of IoT infrastructure. Key technologies used by intransit visibility are RFIDs and cloud-based Global Positioning System (GPS), which provide location, identity, and other tracking information. These data will be the backbone of supply chains supported by IoT technologies. The information gathered by equipment will provide detailed visibility of an item shipped from a manufacturer to a retailer. Data collected via RFID and GPS technologies will allow supply chain managers to enhance automated shipment and accurate delivery information by predicting time of arrival. Similarly, managers will be able to monitor other information, such as temperature control, which can affect the quality of in-transit
products.

2.5.5. Smart Grid

The smart grid is a new generation of power grid in which managing and distributing electricity between suppliers and consumers is upgraded using two-way communication technologies and computing capabilities to improve reliability, safety, efficiency with real-time control, and monitoring [19], [20]. One of the major challenges in a power system is integrating renewable and decentralized energy. Electricity systems require a smart grid to manage the volatile behavior of distributed energy resources (DERs) [21]. However, most energy systems have to follow governmental laws and regulations, as well as consider business analysis and potential legal constraints [22]. Grid sensors and devices continuously and rapidly generate data related to control loops and protection and require real-time processing and analytics along with machine-to-machine (M2M) or human-tomachine (HMI) interactions to issue control commands to the system. However, the system must fulfill visualization and reporting requirements.

3. Conclusion

Many of the conversations taking place around the Internet of Things (IoT) are incomplete without a mention of big data. Connected devices, sensors, and algorithms all operate in ways that involve massive amounts of data. As organizations step into IoT, they must understand the symbiotic relationship between IoT and big data. For IoT deployments to really make an impact, they must provide some sort of useful tool or service, while also collecting relevant data. Just like with any big-data play, merely collecting the data isn't enough. The data mustbe processed and analyzed to glean insights, and those insights must drive actionable steps that can improve the business.

References

[1] K. Ashton, "That 'internet of things' thing in the real world, things matter more than ideas," RFID Journal, June 2009, http://www.rfidjournal. com/article/print/4986 [Accessed on: 2013-10-25].

[2] International Telecommunication Union (ITU), "ITU internet reports 2005: The internet of things," ITU,Executive Summary, Nov. 2005, http://www.itu.int/pub/S-POL-IR.IT-2005 [Accessedon:2013-10-25].

[3] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," Computer Networks, vol. 54, no.15, pp. 2787-2805, Oct. 2010.

[4] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context aware computing for the internet of things: A survey," IEEE Communications Surveys & Tutorials, Accepted for Publication.

[5] M. R. Palattella, N. Accettura, X. Vilajosana, et al, "Standardized protocol stack for the internet of (important) things," IEEE Communications Surveys & Tutorials, vol. 15,no.3,pp.1389-1406,ThirdQuarter2013.

[6] F.J. Riggins, S.F. Wamba, Research directions on the adoption, usage, and impact of the internet of things through the use of big data analytics, in: Proceedings of 48th Hawaii International Conference on System Sciences (HICSS'15), IEEE, 2015, pp. 1531–1540.

[7] Boyd, D., & Crawford, K. (2011, September). Six provocations for big data. In A decade in internet time: Symposium on the dynamics of the internet and society (Vol. 21). Oxford: Oxford Internet Institute.

[8] SHAN, Y. C., Chao, L. V., ZHANG, Q. Y., & TIAN,X. Y. (2017). Research on Mechanism of Early Warning of Health Management Based on Cloud Computing and Big Data. In Proceedings of the 23rd International Conference on Industrial Engineeringand Engineering Management 2016 (pp.291-294). Atlantis Press, Paris.

[9] Parvin Ahmadi Doval Amiri and Mina Rahbari Gavgani, 2016. A Review on Relationship and Challenges of Cloud Computing And Big Data: Methods of Analysis and Data Transfer. Asian Journal of Information Technology, 15: 2516-2525.

[10] Chen, Min, et al. Big data: related technologies, challenges and future prospects. Heidelberg: Springer, 2014.

[11] Demchenko, Yuri, et al. "Big security for big data: Addressing security challenges for the big data infrastructure." Workshop on Secure Data Management. Springer, Cham, 2013.

[12] Liebowitz, J. (Ed.). (2014). Bursting the big data bubble: The case for intuition-based decision making. CRC Press.

[13] Sremack, Joe. Big Data Forensics–Learning Hadoop Investigations. Packt Publishing Ltd, 2015.

[14] Franks, Bill. Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics. Vol. 49. John Wiley & Sons, 2012.

[15] Furht, Borko, and Flavio Villanustre. Big Data Technologies and Applications, Chapter 1, Springer, 2016.

[16] S. Darby, ‘‘Smart metering: What potential for householder engagement?’’ Building Res. Inf., vol. 38, no. 5, pp. 442–457, 2010.

[17] N. Tohamy, ‘‘What you need to know about the Internet of Things,’’ MHD Supply Chain Solutions, vol. 45, no. 3, p. 32, 2015.

[18] C. Pettey. (2015). Five Ways the Internet of Things Will Benefit the Supply Chain. [Online]. Available: http://www.gartner.com/smarterwithgartner/five-ways-the-internet-of-things-will-benefitthe-supply-chain-2/.

[19] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, ‘‘A survey on smart grid communication infrastructures: Motivations, requirements and challenges,’’ IEEE Commun. Surveys Tuts., vol. 15, no. 1, pp. 5–20, Feb. 2013.

[20] S. Bera, S. Misra, and J. J. P. C. Rodrigues, ‘‘Cloud computing applications for smart grid: A survey,’’ IEEE Trans. Parallel Distrib. Syst., vol. 26, no. 5, pp. 1477–1494, May 2015.

[21] T. Dethlefs, Energy Service Description for Capabilities of Distributed Energy Resources. Berlin, Germany: Springer, 2015.

[22] C. Neureiter, ‘‘A standards-based approach for domain specific modelling of smart grid system architectures,’’ in Proc. 11th Int. Conf. Syst. Syst. Eng. (SoSE), Kongsberg, Norway, Jun. 2016, pp. 1–6.