A REVIEW ON INTERNET OF THINGS WITH BIG DATA

Kunjan P. Adeshara
Electronics & Communication Engg. Dept.
Parul Institute of Engineering &Technology
Vadodara, India
kunjanadeshara@gmail.com

Prof. Umang H Soni
Electronics & Communication Engg. Dept.
Parul Institute of Engineering &Technology
Vadodara, India
umangbhai.soni@paruluniversity.ac.in

Prof. Aslam Durvesh
Electronics & Communication Engg. Dept.
Parul Inst. of Engg. &Tech.
Vadodara, India
aslam.durvesh@parunluniversity.ac.in

Prof. Ravi Parikh
Electronics & Communication Engg. Dept.
Parul Inst. of Engg. &Tech.
Vadodara, India
ravi.parikh@paruluniversity.ac.in

Prof. Utkarsh Shah
Electronics & Communication Engg. Dept.
Parul Inst. of Engg. &Tech.
Vadodara, India
utkarsh.shah@paruluniversity.ac.in

Abstract : Iota (Internet of Things) is the new mojo in the present date when it comes to connecting several devices to the internet. With the rapid increase in the development of Iota (Internet of Things) an immense amount of data has been generated in every single second. In the past few years, most of the data has been generated due to the increase in miniaturization of Iota devices. However, in this entire process of using Iota devices and the immense number of data which is been generated has become a major challenging task to manage the entire data. However, such a big amount of data is just laying waste without analytic tools. Big Data technologies have evolved in recent years and have appeared as an essential data analysis tool to bring the knowledge within Iota infrastructures to better fulfil the purpose of the Iota system and support critical decision making [1]. Even though Big Data Analytics itself is a very large, researched topic but if we see Iota domains somehow it isolated the evolution of Big Data approaches in Iota domains [2]. Thus, more research in the Iota domain can possibly bring more advancement into Big Data research in Iota. In this paper, we are reviewing the role of big data in the Iota domain how analytics have played a major role in not only medical but rest other domains of Iota and what all challenges are faced.

Keywords – IOT (Internet Of Things) and Big Data

I. INTRODUCTION

IoT became very convenient for everyone to access their required information instantly but with an increase in users comes an increase in data generation. Same goes with Iota (Internet of Things) the greater number of devices the more amount of data is being generated. According to IBM, 90% of the data has been created in the last two years [3] out of which hardly 1% per cent of data is been analyzed. Apart from social media, cloud, web, etcetera Iota stands on 4th position in the world's biggest source of big data [4]. More information comes with more responsibility with the rapid increase in the development of Iota and Big Data it had started affecting all the areas of technologies and businesses by increasing the benefit for organizations and individuals [5]. Big Data Analytics can be defined as a process of evaluating such huge amount of data to discover various patterns, market trends, and etcetera. Thus, with the power of analytics, we can manipulate any type of information to represent characteristics of an object or person, market trends, frauds, etcetera. Big data can be classified mainly according to three aspects or in another word the 3 V's of Big Data: (a) volume (b) variety (c) velocity. These categories were first introduced by Gartner to describe the elements of big data challenges [6], where volume stands for the huge quantity of data which can be in terabytes, zettabytes, etcetera, variety is for what kind of data will be gathered for analytics it may be structured, semi-structured or unstructured. Lastly comes to the velocity means at what speed the data is been generated. For example, Google alone processes on average more than "40,000 search queries every second," which roughly translates to more than 3.5 billion searches per day [7]. Hence, big data analytics is all about extracting knowledgeable information using various data mining techniques that help in making predictions, identifying recent trends, hidden information and decision making, etcetera. Data mining techniques are widely deployed for both problem-specific methods and generalized data analytics. Accordingly, statistical and machine learning methods are utilized [8]. With the current research in Iota and Big Data, the fusion of both has created tremendous opportunities for the development of several domains like health care, smart cities, automobile, etcetera.

The main aim of this paper is to create a general level for the mutual understanding of similarities and differences in different Iota domains and Big Data research. The rest of the paper is divided into following sections as an overview of Iota and Big Data, Importance of Big Data in Iota followed by how Big Data has approached in different Iota domains.

Overview of Iota and Big Data

Iota: In recent years, many traditional domains such as manufacture industry, healthcare and etcetera have started adopting Iota technologies for better performance and advancement of communication between machine and humans. As a result, less need for human interference and more efficient output has been seen. In simple terms, it provides a platform where a large number of sensors and devices can come together and create a smart environment to communicate and share information across various

other platforms in a convenient manner. With the increase in wireless technologies, Iota has created new trends where electronics equipment such as smart wristwatches, emergency alarms, garage doors, refrigerators, microwave ovens, air conditioners all are connected to an Iota network which can be remotely controlled. Ciufo [9] stated that these devices "talk" to one another and to central controlling devices. Such devices deployed indifferent areas may collect various kinds of data, such as geographical, astronomical, environmental, and logistical data.

Iota consists of an exceptionally large amount of communication devices to associate with each of them and to transfer data, where a mode of communication must be established. The communication between the sensors and devices are mostly interconnected through a variety of communication technologies such as Bluetooth, WI-FI, ZigBee, GSM etc. Through these communication devices, it helps us in remotely controlling our Iota devices. More than trillion devices ranging from smartphones, laptops, sensors and game consoles are anticipated to be connected to the Internet through several heterogeneous access networks enabled by technologies, such as radiofrequency identification (RFID) and wireless sensor networks [10]. According to various references, mentioned that Iota could be recognized majorly in three parts: Internet-oriented, Sensors and Knowledge [1].

With such huge amount of communication devices comes an gigantic amount of data, therefore, existing architecture of the Internet needed modifications to generate an efficient communication protocol. Multiple proposed layer architectures are discussed in various research papers. Chen [2] discusses the three-layer architecture for Iota while Shuo et al. divide the Iota architecture into four layers [3]. Five and six layers of architectures are presented in [4] and [5], respectively. The division of architecture into different layers makes it possible to process data at each layer which minimizes big data processing at the final layer. Figure 1 shows seven layers of architecture in Iota.

The Iota Reference Model starts with the initial layer (Layer-1): where the multiple devices might controlled by the Physical devices and controllers. These are the “things” in the Iota, and they involves a wide range of endpoint devices that send and receive information. Today, the class of devices are becoming huge. It will become almost unlimited as more equipment is added to the Iota over time. Devices are diverse, and there are no rules about size, location, form factor, or origin. Some devices will be the size of a silicon chip. Some will be as large as vehicles. The Iota must support the entire range. Dozens or hundreds of equipment manufacturers will produce Iota devices. To simplify compatibility and support manufacturability, the Iota Reference Model generally describes the layer of processing needed from layer 1 devices. 2nd Layer describes basic capabilities for a device. The “devices” which are lying in Iota are capable in conversion of Analog to digital as required, generating and Being queried / Controlled the data over-the-net.

Connectivity and communications are concentrated in same layer—layer 2. The most important function of 2nd layer is reliability and timely in the sense of information transmission. Traditional data communication networks have multiple functions, as evidenced by the International Organization for Standardization (ISO) 7-layer reference model architecture. However, a complete Iota system contains many layers in addition to the communications network.

Objective of the Iota Reference Model is for processing and communications to be accomplished by existing networks. The Iota Reference Model architecture does not require or indicate creation of a different network—it relies on existing networks. However, some legacy devices aren’t IP-enabled, which will require introducing communication gateways. Other devices will require proprietary controllers to serve the communication function. However, over time, standardization will increase. As layer 1 devices proliferate, the ways in which they interact with layer 2 connectivity equipment may change. Regardless of the details, layer 1 devices communicate through the Iota system by interacting with layer 2 connectivity equipment, as shown in layer 3. Connective includes: Communicating with and between the layer 1 devices, Reliable delivery across the network(s), Implementation of various protocols, switching and routing, Translation between protocols, Security at the network layer, (Self Learning) Networking Analytics.

The functions of layer 3 are driven by the need to convert network data flows into information that is suitable for storage and higher layer processing at layer 4 (data accumulation). This means that layer 3 activities focus on huge volume data analysis and transformation. For example, a layer 1 sensor device might generate data samples multiple times per second, 24 hours a day, and 365 days a year. A basic precept of the Iota Reference Model is that the most intelligent system initiates information processing as early and as close to the edge of the network as possible. This is sometimes referred to as fog computing. Layer 3 is where this occurs.

Given that data is usually submitted to the connectivity layer (layer 2) networking equipment by devices in small units, layer 3 processing is performed on a packet-by-packet basis. This processing is limited, because there is only awareness of data units—not “sessions” or “transactions”. This includes Data filtering, cleanup, aggregation, Packet content inspection, Combination of network and data layer analytics, Thresholding, Event generation. Layer 4 Edge Fog Computing is a paradigm that extends Cloud computing and services to the edge of the network. Similar to Cloud, Fog provides data, compute, storage, and application service to the end-users. The motivation of Edge Fog computing lies in a series of real scenarios, such as Smart Grid, Smart Traffic Lights in vehicular networks and Software defined networks.

Networking systems are built to reliably move data. The data is “in motion.” Prior to Layer 4, data is moving through the network at the rate and organization determined by the devices generating the data. The model is event driven. As defined earlier, Layer 1 devices do not include computing capabilities themselves. However, some computational activities could occur at Layer 2, such as protocol translation or application of network security policy. Additional compute tasks can be performed at Layer 3, such as packet inspection. Driving computational tasks as close to the edge of the Iota as possible, with heterogeneous systems distributed across multiple management domains represents an example of fog computing. Fog computing and fog services will be a distinguishing characteristic of the Iota. As Layer 4 captures data and puts it at rest, it is now usable by applications on a non-real-time basis. Applications access the data when necessary. In short, Layer 4 converts event-based data to query-based processing. This is a crucial step in bridging the differences between the real-time networking world and the non-real-time application world. Figure 6 summarizes the activities that occur at Layer 4. This includes Event filtering / Sampling, Event comparison, Event joining for CEP, Event based rule evaluation, Event aggregation, Northbound / Southbound alerting and Event persistence in storage.

Iota systems will need to scale to a corporate—or even global—layer and will require multiple storage systems to accommodate Iota device data and data from traditional enterprise ERP, HRMS, CRM, and other systems. The data abstraction functions of Layer 5 are focused on rendering data and its storage in ways that enable developing simpler, performance-enhanced applications. This includes Information Integration.
Layer 6 is the application layer, where information interpretation occurs. Software at this layer interacts with Layer 5 and data at rest, so it does not have to operate at network speeds.

The Iota Reference Model does not strictly define an application. Applications vary based on vertical markets, the nature of device data, and business needs. For example, some applications will focus on monitoring device data. Some will focus on controlling devices. Some will combine device and non-device data. Monitoring and control applications represent many different application models, programming patterns, and software stacks, leading to discussions of operating systems, mobility, application servers, hypervisors, multi-threading, multi-tenancy, etc. These topics are beyond the scope of the Iota Reference Model discussion. Suffice it to say that application complexity will vary widely. This includes Reporting, Analytics and control.
One of the main distinctions between the Internet of Things (Iota) and Iota is that Iota includes people and processes. This difference becomes particularly clear at Layer 7: Collaboration and Processes. The Iota system, and the information it creates, is of little value unless it yields action, which often requires people and processes. Applications execute business logic to empower people. People use applications and associated data for their specific needs. Often, multiple people use the same application for a range of different purposes. So the objective is not the application—it is to empower people to do their work better. Applications (Layer 6) give business people the right data, at the right time, so they can do the right thing.

But frequently, the action needed requires more than one person. People must be able to communicate and collaborate, sometimes using the traditional Internet, to make the Iota useful. Communication and collaboration often requires multiple steps. And it usually transcends multiple applications. This is why layer 7, as shown in Figure 9, represents a higher layer than a single application. This includes Collaboration and Process in that Involving people and business process.

Figure 1: Iota Architecture Layers

Big Data: From past few years the amount of data generated by various sensors, devices, health care, digital devices and etcetera keeps on increasing with huge amount of structured, semi-structured and unstructured data. This huge amount of data results in "Big Data". With rapid growing of data, traditional database systems were not capable of performing numerous activities with the old techniques. Hence, big data come up with new solutions for this problems. According to the recent studies big data is classified into 3 major categories: (a) data sources [6], (b) data analytics, and (c) the presentation of the results of the analytics. The above categories are basically the 3 V's of big data (Volume, Variety & Velocity). With the 3 V's we can easily came up with a solution for such huge volume of data, type of data which we are dealing with and the speed of data in real time, hence we can state that with the help of these 3 V's we can define big data. According to some researchers there are few more characteristics of big data such as veracity, value, variability, visualization and etcetera but the most common characteristics are volume, variety and velocity.

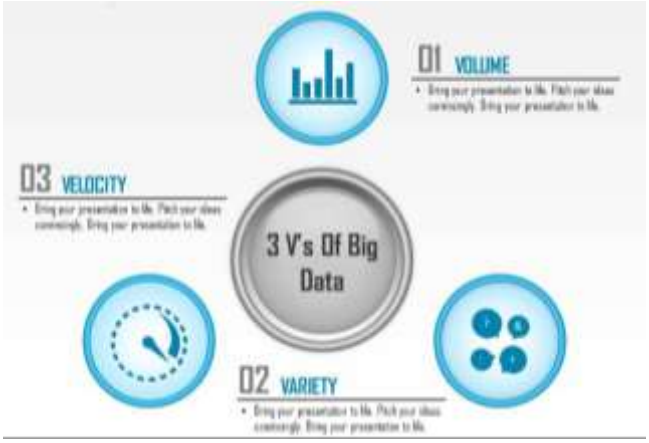

Figure 2 V's of Big Data

Importance of Big Data in Iota

As discussed earlier Big Data is basically characterized into 3 main V's: Volume, Variety & Velocity. With respect to the above characteristics big data is been categorized into various categories as data processing, extensive data storage, speed and etcetera which creates multiple amount of big data in various Iota domains but the most challenging part is yet arrive to come. Numerous amount of devices and sensors not only increases the data volume but also create a large amount of data traffic. For proper functioning of Iota services it's very important for the big data to be analyzed in a precise way as most of the data is in unstructured form which makes it a more challenging task.

Analysis mainly consists of three major part storage, processing and precise result. Earlier ELT (Extract, load and transform) technique was used for data storage. Contrary to ELT, there are other techniques like Massive Parallel Processing (MPP), Non-Relational and In-Memory databases [9] which are not only flexible for new and mobile devices but also have high processing power [7]. Big data not only consists of unstructured data but also has structured and semi-structured data in the domains of Iota so it's really very important to take care of the management of the data. There are end number of tools designed by many organization to develop big data management system, out of many organizations yahoo was the one to come up with Apache Hadoop which has the capability of processing and storing big data from multiple devices through implementing Map Reduce Paradigm. Hadoop basically divides the data into multiple clusters or sets so that its processing can be performed on various devices [8] using parallel processing algorithm. Thus, with such numerous amount of sensors and devices which are continuously generating end number of data it can be said that big about is all about analysis and without analysis it's next to impossible to manage big data in Iota.

Big Data approach In Iota

In the previous section we have seen the importance of big data in Iota and how analysis plays a major role in big data.

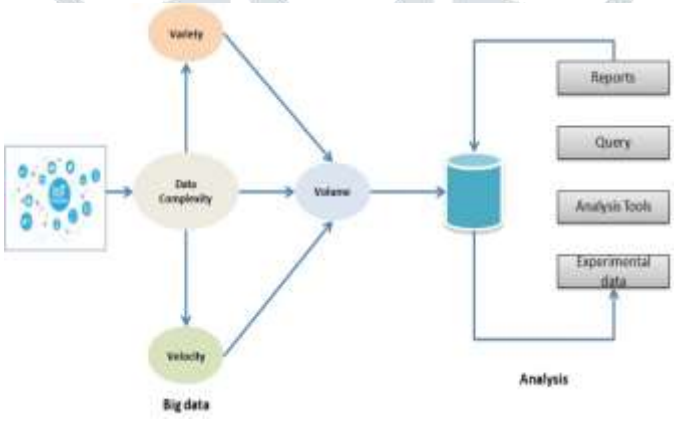

Figure 3 Role of Analytics

Big data analytics is emerging as a key to analyzing Iota generated data from "connected devices" which helps to take the initiative to improve decision making.

The role of big data in Iota is to process a large amount of data on a real-time basis and storing them accordingly using various storage technologies.

Big data processing in Iota mainly follows four sequential steps –

- A large amount of data (mainly unstructured) is generated by Iota devices which are collected in the various big data storage system. This devices generates a large amount of big data which mainly depends on the 3 main factors that are volume, velocity, and variety also known as the 3V's.
- Big data system is basically a shared distributed database which has huge amount of data is stored in big data files.
- Analyzing the stored Iota data using analytic tools like Hadoop MapReduce or Spark.

- Generating the reports of analyzed data in form of dashboards.

Since in Iota most of the data is in unstructured form which is collected via the internet, hence, big data for the internet of things needs lightning-fast analysis with large queries to extract rapid insights from data to make the full of use of that information.
Hence, from the big data perspective, we can say data is the new fuel which drives the Iota.

How do Iota and Big Data Influence Each Other?

It's not just the only interdependent relation between big data and Iota. As they help each other, in addition to that they hugely influence each other. Fact is the more the Iota technology develops it will place more demand on big data capabilities. For example, as the Iota generated data is increasing at a huge rate, conventional data storage technology is already being pushed to its limits. As a result, it demands more advanced and innovative storage solutions to handle these growing workloads resulting in updating the infrastructure of an organization's data storage.
Similarly, both the combined applications boasts the scope of research in both the fields. So, Iota and big data both the technologies carry inter-dependency and need further development with increasing days gradually.

How are Internet of Things and Big Data Together Beneficial for Companies?

Big data analytics can be useful for a variety of Iota data to –

- Explore
- Reveal tendency
- Find unseen Samples
- Find hidden correlations
- Reveal new information

Hence, companies can benefit from analyzing such large amount of Iota big data and managing them to identify how they affect businesses. As a result, it assists business and other organizations to achieve an improved understanding of data, and thus, making efficient and well-informed use of it which makes it possible to take appropriate decisions.

Benefits of IoT and Big Data for Companies in Different Sectors

- *Helps to gain the ROI for the Businesses*

Iota and big data analytics are transforming how businesses are adding value by extracting maximum information from data to get better business insights. With the increased demand for data storage companies prefers big data cloud storage which ultimately lowers the implementation cost for them.

- *It will change the future of e-health system*

The combined features of the Iota and big data can redesign the next generation of e-health care systems. Big data will lead to hypothesis-driven research to data-driven research transformation. On the other hand, Iota will help to control and analyze the different levels of connections between various sensor signals and existing big data. This will enable new ways of remote diagnosis with a better understanding of the disease which will lead to the development of innovative solutions in the healthcare field as well.

- Advantages in manufacturing companies

If manufacturing companies install Iota sensors within its equipment, they can collect significant operational data on the machines. This helps them to have an in-depth look at how the business is performing and enable them to find out which equipment need repairing before much problems arise. This prevents them from more significant expenses by skipping the downtime or replacement of the equipment. Hence, investment in Iota and big data can cause saving of a large amount money in the business.

- *Internet of things and big data will raise self-service analytics*

With more inventions in the Iota field, most of the IT functions can be handled with data automation and integration. Additionally, big data tools will increasingly become self-sufficient and straightforward to perform basic functions. Hence, analytics as a service will become more of a self-service type.

- *Benefits in the transportation industry*

In the transportation sector, Iota sensors have been installed in the vehicles as a way to track them the go and around the world. This doesn't only help companies to keep a closer eye on the vehicles, but it also provides the data regarding fuel efficiency, how drivers utilize their time, delivery routes and etcetera. This information can be indispensable for optimizing fleets and for the improvement of organizational productivity.

- *Edge-Computing will be in high demand*

Working on real-time data is a high priority today and a necessity as well. As Iota and Big data both enable on-demand and real-time action, the importance of deployment of these technologies is high. In this view, the popularity of edge computing is also becoming very high.
As the Iota and big data are closely linked, there are many examples out there of organizational benefits to put them to good use.

Conclusion

In this paper, we have performed a survey on various papers based on Iota and Big Data communities. Based on these papers we have discussed about various Iota domains and its challenges. By reviewing various big data research papers, we have focused on the overview of Iota and Big Data in which we have discussed it in detail about both the technologies along with its architecture and characteristics. We have also highlighted the 3 V's of the big data and how it effects on analysis. Most importantly why big data is important in Iota and why there was a need of analysis. Followed by how big data approached in Iota and comparison of Iota domains from Big Data perspective vice and versa. Lastly we have mentioned a table which shows how different technologies used in big data are used in different Iota domains. The convergence of Iota and big data can provide new opportunities and applications in all the sectors. Along with that, it has the potential to revolutionize many aspects of our society.

REFERENCE

[1] Big Data for Internet of Things: A Survey Institute of Computer Science, Masaryk University, Brno, Czech Republic. Faculty of Informatics, Masaryk University, Brno, Czech Republic. FSTG, Cadi AyyaD University, Marrakesh, Morocco

[2] https://techjury.net/blog/big-data-statistics/#gref

[3] https://www.allerin.com/blog/top-5-sources-of-big-data

[4] Big IoT Data Analytics: Architecture, Opportunities, and Open MOHSEN MARJANI1, FARIZA NASARUDDIN2, ABDULLAH GANI1, (Senior Member, IEEE), AHMAD KARIM3, IBRAHIM ABAKER TARGIO HASHEM1, AISHA SIDDIQA1, AND IBRAR YAQOOB1 Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia, Department of Information System, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia, Department of Information Technology, Bahauddin Zakariya University, Multan, Punjab, Pakistan, Corresponding authors: Fariza Nasaruddin (fariza@um.edu.my); Abdullah Gani (abdullah@um.edu.my)

[5] https://www.slideteam.net/0115-3-vs-of-big-data-volume-velocity-variety-with-icons-ppt-slide.html

[6] A Review on Big Data Analysis and Internet of Things, Umar Ahsan, Electronic Systems Engineering University of Regina SK, Canada, Email: umarahsan@uregina.ca, Abdul Bais, Electronic Systems Engineering University of Regina SK, Canada, Email: abdul.bais@uregina.ca

[7] F. Bonomi, ``Fog computing and its role in the Internet of Things,"
In *Proc. 1st Ed. MCC Workshop Mobile Cloud Comput.*, 2012,
pp. 1_6

[8] K. Steinklauber. (2014). *Data Protection in the Internet of Things*.
[Online]. Available: https://securityintelligence.com/data- protectioninthe-internet-of-things Hu, H. Chen, L. Huang, and X. Zhu, ``A survey of mass data mining based on cloud-computing," in *Proc. Anti-Counterfeiting, Secur. Identi_cat.*, Aug. 2012, pp. 1_4.

[9] Y. Sun, ``Mining knowledge from interconnected data: A heterogeneous information network analysis approach," *Proc. VLDB Endowment*, vol. 5, no. 12, pp. 2022_2023, 2012. M. Chen, L. T. Yang, T. Kwon, L. Zhou, and M. Jo, ``Itinerary planning for energy-efficient agent communications in wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3290_3299, Sep. 2011.