ISSN 2319 – 1953
International Journal of Scientific Research in Computer Science Applications and Management Studies

Internet of Things with BIG DATA Analytics – A Survey

[1]A.Pavithra, [2]C.Anandhakumar, [3]V.Nithin Meenashisundharam

[1]*Assistant Professor,* [2]*Student,* [3]*Student,*
Department of Computer Science, Sree Saraswathi Thyagaraja College, Pollachi.
[1]pavithra@stc.ac.in, [2]kumaranand16498@gmail.com, [3]nithinmsv@gmail.com

Abstract- **The INTERNET OF THINGS (IOT) is an internet among things through advanced communication protocols without human's operation. The main idea of IOT is to reach and improve the common goal by their intelligent networking. The IOT is an integrated technology of several sub technologies, such as wireless sensor or semantic and IOT is the way of connecting devices using sensors and monitored by Internet. But the data produced by the IOT is growing rapidly because of the large scale development of various applications. As the data is turned and crossed over terabytes and leading to petabytes, there should be a solution to manage the overwhelming increase in data. Big data is data that are too big (volume), too fast (velocity) and too diverse (variety). Big data is a collection of large data sets that include different types such as structured, unstructured and semi-structured data. Big data is the solution for the data problem and it is considered as the future's data dream. The upcoming IoT will be greatly presented by the enormous quantity of heterogeneous networked embedded devices that generate intensively "Big data". In this article we discuss about Big data on IoT and how it is interrelated to each other along with necessity of implementing Big data with IoT and its benefits, job market.**

Keywords: **Internet of Things, Big Data, Streaming, Spatial, Time Series, Prescriptive Analysis**

I. INTRODUCTION

1.1 Internet of Things

The Internet Of Things (IoT) is a computing concept that describes the idea of everyday physical objects being connected to the internet and being able to identify themselves to other devices. The term is closely identified with RFID as the method of communication, although it also may include other sensor technologies, wireless technologies or QR codes.

The IoT is significant because an object that can represent itself digitally becomes something greater than the object by itself. No longer does the object relate just to its user, but it is now connected to surrounding objects and database data. When many objects act in unison, they are known as having "ambient intelligence."

The internet of things is a difficult concept to define precisely. In fact, there are many different groups that have defined the term, although its initial use has been attributed to Kevin Ashton, an expert on digital innovation. Each definition shares the idea that the first version of the internet was about data created by people, while the next version is about data created by things. In 1999, Ashton said it best in this quote from an article in the RFID Journal:

"If we had computers that knew everything there was to know about things – using data they gathered without any help from us – we would be able to track and count everything, and greatly reduce waste, loss and cost. We would know when things needed replacing, repairing or recalling, and whether they were fresh or past their best."

Most people think about being connected in terms of computers, tablets and smartphones. IoT describes a world where just about anything can be connected and communicate in an intelligent fashion. In other words, with the internet of things, the physical world is becoming one big information system.

1.2 Big Data

Big data refers to a process that is used when traditional *data* mining and handling techniques cannot uncover the insights and *meaning* of the underlying *data*. *Data* that is unstructured or time sensitive or simply very *large* cannot be processed by relational database engines.

Quite simply, big data reflects the changing world we live in. The more things change, the more the changes are captured and recorded as data. Take weather as an example. For a weather forecaster, the amount of data collected around the world about local conditions is substantial. Logically, it would make sense that local environments dictate regional effects and regional effects dictate global effects, but it could well be the other way around. One way or another, this weather data reflects the attributes of big data, where real-time processing is needed for a massive amount of data, and where the large number of inputs can be machine generated, personal observations or outside forces like sun spots.

Processing information like this illustrates why big data has become so important:

- Most data collected now is unstructured and requires different storage and processing than that found in traditional relational databases.
- Available computational power is sky-rocketing, meaning there are more opportunities to process big data.
- The Internet has democratized data, steadily increasing the data available while also producing more and more raw data.

Data in its raw form has no value. Data needs to be processed in order to be of valuable. However, herein lies the inherent problem of big data. Is processing data from native object format to a usable insight worth the massive capital cost of doing so? Or is there just too much data with unknown values to justify the gamble of processing it with big data tools? Most of us would agree that being able to predict the weather would have value, the question is whether that value could outweigh the costs of crunching all the real-time data into a weather report that could be counted on.

The Volume: The Volume challenge refers to storing, processing and quickly accessing large amounts of data. While it is hard to quantify the boundary for a volume challenge, common data sets in the order of hundreds of Terabyte or more are considered to be big. In contrast to traditional storage technologies such as relational data base management systems (RDBMS), new Big Data technologies such as Hadoop are designed to easily scale with the amount of data to be stored and processed. In its most basic form, the Hadoop system uses its Hadoop Distributed File System (HDFS), to store raw data. Parallel processing is facilitated by means of its Map Reduce framework that is highly suitable for solving any embarrassingly parallel processing problems. With Hadoop it is possible to scale by simply adding more processing nodes to the Hadoop cluster without the need to do any reprogramming as the framework takes care of using additional resources as they become available. Summarizing the trends in the volume challenge one can observe a paradigm shift with respect to the way the data is handled. In traditional database management systems the database design is optimized for the specific usage requirements, i.e., data is pre-processed and only the information that is considered relevant is kept. In contrast, in a truly data-driven enterprise that builds on Big Data technologies, there is awareness that data may contain value beyond the current use. Thus a master data set of the raw data is kept that allows data scientists to discover further relationships in the data, relationships that may reside beyond the requirements of today. As a side effect it also reduce the costs of human error such as erroneous data extraction or transformation.

The Velocity: Velocity refers to the fact that data is streaming into the data infrastructure of the enterprise at a high rate and must be processed with minimal latency. To this end, different technologies are applicable, depending on the amount of state and complexity of analysis. In cases where only little state is required (e.g., maintaining a time window of incoming values), but complex calculations need to be performed over a temporally scoped subset of the data, Complex Event Processing (CEP) engines offer efficient solutions for processing incoming data in a stream manner. In contrast, when each new incoming data set needs to be related to a large number of previous records, but only simple aggregations and value comparisons are required, noSQL databases offer the necessary write performance. The required processing performance can then be achieved by using streaming infrastructures such as Storm.

The Variety: In a data-driven economy the objective is to maximize the business value by considering all available data. From a technical perspective one could formulate that problem by evaluating a function over all accessible data sets. In practice, however, this approach must confront the challenge of heterogeneous data sources ranging from unstructured textual sources (e.g., social media data) to the wide disparity in the formats of sensor data. Traditionally this challenge is addressed by various forms of data integration. In the context of Big Data there is however a new dimensions to the integration challenge which is the amount of different data sources that need to be integrated. Social media, open (governmental) data sources, and data platforms and markets, result in a data ecosystem of significant variation. As the integration of new data sources requires manual work to understand the source schema, to define the proper transformations and to develop data adapters, existing approaches do not scale effectively.

The Veracity: Apart from the original 3V's described above, an almost inexhaustible list of Big Data V's are discussed. For instance veracity relates to the trust and truthfulness of the data. Data may not fully be trusted, because of the way it has been acquired, for instance by unreliable sensors or imperfect natural language extraction

The Value: When we talk about value, we're referring to the worth of the data being extracted. Having endless amounts of data is one thing, but unless it can be turned into value it is useless. While there is a clear link between data and insights, this does not always mean there is value in Big Data. The most important part of embarking on a big data initiative is to understand the costs and benefits of collecting and analyzing the data to ensure that ultimately the data that is reaped can be monetized.

II. Why Big data with IoT

The first thing that comes to mind when talking about Big Data and IoT is the increase in the volume of data that will hit the data storage framework of companies. Data centers will have to be set up to handle this entire additional data load.

Taking into consideration the enormous impact IoT will on data storage infrastructure, organizations have begun to move towards the Platform-as-a-Service model, a cloud-based solution, as opposed to maintaining their own storage infrastructure. Unlike, in-house data systems that need to be constantly updated as the data load increases, PaaS provides flexibility, scalability, compliance, and a sophisticated architecture to store all valuable IoT data. Cloud storage options include public, private, as well as hybrid models. If a company has sensitive data that is subject to any regulatory compliance requirements that require heightened security, using a private cloud would be the best course of action. For other companies, a public or hybrid cloud can be used for the storage of IoT data.

III. Necessity of Implementing Big Data with IoT

IoT and data remain intrinsically linked together. Data consumed and produced keeps growing at an ever expanding rate. This influx of data is fueling widespread IoT adoption as there will be nearly 30.73 billion IoT connected devices by 2020. The Internet of Things (IoT) is an interconnection of several devices, networks, technologies, and human resources to achieve a common goal. There are a variety of IoT-based applications being used in different sectors and have succeeded in providing huge benefits to the users.

The data generated from IoT devices turns out to be of value only if it gets subjected to analysis, which brings data analytics into the picture. Data Analytics (DA) is defined as a process, which is used to examine big and small data sets with varying data properties to extract meaningful conclusions and actionable insights. These conclusions are usually in the form of trends, patterns, and statistics that aid business organizations in proactively engaging with data to implement effective decision-making processes.

IV. Merging Data Analytics and IoT will Positively Impact Businesses

Data Analytics has a significant role to play in the growth and success of IoT applications and investments. Analytics tools will allow the business units to make effective use of their datasets as explained in the points listed below.

- **Volume:** There are huge clusters of data sets that IoT applications make use of. The business organizations need to manage these large volumes of data and need to analyze the same for extracting relevant patterns. These datasets along with real-time data can be analyzed easily and efficiently with data analytics software.
- **Structure:** IoT applications involve data sets that may have a varied structure as unstructured, semi-structured and structured data sets. There may also be a significant difference in the data formats and types. Data analytics will allow the business executive to analyze all of these varying sets of data using automated tools and software.
- **Driving Revenue:** The use of data analytics in IoT investments will allow the business units to gain an insight into customer preferences and choices. This would lead to the development of services and offers as per the customer demands and expectations. This, in turn, will improve the revenues and profits earned by the organizations.

V. Types of Analytics can be performed

- **Streaming Analytics:** This form of data analytics is also referred as event stream processing and it analyzes huge in-motion data sets. Real-time data streams are analyzed in this process to detect urgent situations and immediate actions. IoT applications based on financial transactions, air fleet tracking, traffic analysis etc. can benefit from this method.
- **Spatial Analytics:** This is the data analytics method that is used to analyze geographic patterns to determine the spatial relationship between the physical objects. Location-based IoT applications, such as smart parking applications can benefit from this form of data analytics.
- **Time Series Analytics:** As the name suggests, this form of data analytics is based upon the time-based data which is analyzed to reveal associated trends and patterns. IoT applications, such as weather forecasting applications and health monitoring systems can benefit from this form of data analytics method.
- **Prescriptive Analysis:** This form of data analytics is the combination of descriptive and predictive analysis. It is applied to understand the best steps of action that can be taken in a particular situation. Commercial IoT applications can make use of this form of data analytics to gain better conclusions.

VI. The IoT and Big Data job market

Big Data and the Internet of Things are the two most-talked-about technology topics of the last few years. This is one of the chief reasons why they occupy prominent places on analyst firm Gartner’s most recent Hype Cycle for Emerging Technologies. These two technologies are set to transform all areas in business as well as everyday life. In the 2015 Internet of Things predictions, IDC notes that over 50% of IoT activity is centered in manufacturing, transportation, smart city, and consumer applications, but that within five years every industry will have rolled out IoT initiatives.

Data Science Central conducted a survey that showed how widespread IoT jobs are, today. This is a list of the top companies that are hiring for IoT related jobs:

1. PTC – The Product Development Company
2. Amazon
3. Continental
4. Savi Group
5. Intel
6. Ayla Networks
7. HP
8. LogMeln.Inc
9. Red Hat. Inc
10. Honeywell
11. IBM
12. Renesas
13. Cisco Systems. Inc
14. Dell
15. InterDigital

The IoT and Data related positions that companies are hoping to fill with qualified people are:

1. Big Data Lead (IoT)
2. Data Scientist - IoT
3. Data Engineer - Sensors and IoT
4. Data Engineer Sensors and IoT Applications

Given these developments, the opportunities available to certified Big Data professionals in the rapidly growing ‘Internet of Things’ domain are endless. Simplilearn’s Big Data Hadoop Architect Masters Program was designed with the IoT-driven world of the future in mind. With over 200 hours of high-quality e-learning content, access to CloudLab – a cloud based Hadoop environment lab–on-demand support by Hadoop experts, simulation exams, and a certification to validate your skills, the Big Data Hadoop Architect Masters program will see you ready to take on the challenges and opportunities of a world where the Internet of Things is commonplace.

VII. Benefits of IoT based on Big Data

In literature, various 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 presented structure consists of three main components which are big data management, IoT sensor, and data analysis. These analyzes use are in the realtime management of oxygen level, dangerous gases/soot and the amount of ambient light in smart buildings. In addition to 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 analysis of data which obtained from different sensors such as water sensors, transportation network sensors, monitoring devices, smart home sensors and smart car park sensors. These data are generated and processed in a multi-stage model and ultimately reached a decision-making stage. These steps are data production, data collection, data integration, data categorization, data processing and decision making. 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 devised, and a conceptual framework has been proposed consisting of 5 layers: • Data Collection layer - collected data from various sources, the input layer is the proposed framework. • Extract-transform-load (ETL) layer - provides the ability to change the format of information received from different types of sensors into a defined format. • The semantic reasoning rules layer - an inference engine that acts on the information received from the ETL layer • Learning layer - From the data tailored to the existing extraction data, extract the various specifications and attributes, and finally, Machine learning-based models provided. • 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, clustering mechanisms, health, privacy security, performance evaluation of monitoring algorithms, manufacturing systems, and energy development [17]

VIII. Conclusion

The development of IoT devices, smartphones, and social media provides decision makers with opportunities to extract valuable data about users, anticipate future trends and fraud detection. With the creation of transparent and usable data, big data can create the organizations' values, make the changes clear and expand their performance. The use of data generated from the IoT and the analytical tools creates many opportunities for organizations. These tools use predictive modeling technologies, clustering, classification to provide data mining solutions. IoT improves the decision-making habits of decision-makers. The emergence of IoT and related technologies, such as cloud computing, provides the ability to remove data sources in different domains. Typically, 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 key 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 agility. It is anticipated 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 efforts to update our lives to the Internet of Things technology regarding hardware and software.

References

[1] Networking and Information Technology Research and Development (NITRD) Program. (2015). Smart and Connected Communities Framework. [Online]. Available: https://www.nitrd.gov/sccc/materials/ scccframework.pdf.

[2] Y. Sun, Y. Xia, H. Song, and R. Bie, ''Internet of things services for small towns,'' in Proc. Int. Conf. Identificat., Inf. Knowl. Internet Things (IIKI), Oct. 2014, pp. 92–95.

[3] A. J. Jara, Y. Sun, H. Song, R. Bie, D. Genooud, and Y. Bocchi, ''Internet of Things for cultural heritage of smart cities and smart regions,'' in Proc. IEEE 29th Int. Conf. Adv. Inf. Netw. Appl. Workshops (WAINA), Mar. 2015, pp. 668–675.

[4] H. Song, R. Srinivasan, T. Sookoor, and S. Jeschke, Smart Cities: Foundations and Principles. Hoboken, NJ, USA: Wiley, 2016.

[5] U.S. Department of Transportation. (2013). Livability 101. [Online]. Available: http://www.dot.gov/livability/101.

[6] "A survey on Chromecasat digital device" Journal of Emerging Technologies and Innovative Research (ISSN : 2349-5162) Published in Volume 5 Issue 10 , October- 2018.

[7] "Multimedia And Its Applications" in International Journal for Research and Development in Technology (ISSN: 2349-3585) Published in Volume 10 Issue 5 on November 2018.

[8] "TensorFlow in Deep Learning" in International Journal of Innovative Research in Technology" (ISSN: 2349-6002) Published in Volume 5 Issue 9 on February 2019.