International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 7, ISSN No. 2455-2143, Pages 169-171
Published Online November 2022 in IJEAST (http://www.ijeast.com)

OPTIMIZING REAL-TIME INTERNET OF THINGS DATA USING BIG DATA COMPUTING PLATFORM

Dharmita. A
III yr, Department of CSE
SRM Ist, Chennai, Tamil Nadu, India

Abstract— **IOT offers the capability to connect and integrate both digital and physical entities of an appliance. A fundamental challenge center's around managing big IOT data that these appliances produce, which is not only extremely large in scale and volume but also noisy and continuous. In our research we have explored Big Data based IOT driven technologies and the importance of pre-processing, meta data, data storage formats, data management and how big data is closely associated with IOT technologies. To achieve the objective of managing big data we are going to use the two most widely used framework namely Apache Hadoop and Apache Spark explained in detail in the research. Now some of the challenges faced by the IOT based appliances and the users are privacy, data storage and analysis, scalability, pre-processing. However, using the concepts of bigdata we can overcome these challenges. This paper finally helps in the implementation of combined efforts of IOT and bigdata altogether in the field of education, healthcare, urban planning, agriculture sector, and industries and also illustrates how these two technologies working simultaneously can not only improve the quality of the appliances but also enhances the experience of the users. This proposed methodology as per research will be easy and accurate compared to the existing methodology.**

Keywords— **Big Data, IOT, Apache Hadoop, Apache Spark, Volume.**

I. INTRODUCTION

IOT AND BIG DATA

Internet of things is a group of machines that are interrelated and works without the intervention of humans. They are connected through sensors and checked upon using internet. It's a machine-to-machine communication process. While big data is collection of structured, semi-structured and unstructured data that are collected from various organizations. The amount of data being collected is being increased and has crossed over Terabyte and now is the range of Petabyte. Because of this big data we are able to store unlimited amount of data in a secured and structured manner.

Big Data has been characterized by three different parameters which are:

1. **Volume** - It is one of the important characteristics that describes the big data. Since the amount of data is increasing the traditional techniques and the methods used cannot be used and there is a serious requirement of enhanced techniques to process and analyze the data.
2. **Velocity** - It's the speed in which data is being processed. The rate of incoming data from various devices. Large amount of data is being generated through social media like WhatsApp, Instagram post, Twitter.
3. **Variety** – The amount of heterogenous data processed through various organizations and accounts. In other words, it's the number of forms of data. [6]

INTERNET OF THINGS DATA

The data generated using IOT is large in volume and random in nature. To access these data, we need to use the big data tools for analyzing and processing them. The data is generated in multiple ways and variety of devices and is processed in different ways and is transmitted to different locations.

II. EXISTING METHOD AND TOOLS USED

THE BIG DATA TOOLS USED TO ANALYSE IOT DATA

To overcome the ever-increasing data volume in the range of many areas we use a process called cluster computing. It is the process of sharing the computation task among multiple computers and these computers form a cluster. It works on distributed system with the networks.

The two most widely used Big Data framework is Apache Hadoop and Apache Spark.

Apache Hadoop – This framework is an open-source platform that is well established platforms that supports the distribution and parallel processing of massive data. Used for storing large volume of raw data and also provides a general partitioning mechanism across different machines. It's also called a multi-purpose engine but not a real time engine and a high-performance engine. Hadoop contains libraries that use a simple programming language. Hadoop is based upon two nodes: Master node and Slave node. Master node is the one that helps in dividing the problem into sub-problems. Each

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 7, ISSN No. 2455-2143, Pages 169-171
Published Online November 2022 in IJEAST (http://www.ijeast.com)

sub-problem is distributed into different slave node. All the output from the slave node is collected at the end and given to the master node. The Hadoop platform contains the Hadoop kernel, Hadoop MapReduce framework provides highly efficient and reliable programming environment for large volume of data. [9]

Apache Spark – It's also an open-source platform but it is mostly used to get over the restrictions of MapReduce like fault tolerance, linear scalability and for processing large scale data processing framework. Few of its advantages are that it is easy to use, provides high speed, and has sophisticated analysis. Apache spark is not a modified version of Hadoop and it is also not dependent on Hadoop. Hadoop is just one way of implementing spark. Spark uses Hadoop in two ways. They are: Storage and Processing. [10]

Apache Storm – It is used for extensive data processing. It works on real-time data, which should be distributed and fault-tolerant. It forms a cluster of data that is similar to Hadoop clusters. It also works as a Master node and worker node. [3]

Apache Splunk – It's a web interface that allows the user to analyse, search and monitor the data. It helps to register the structure and unstructured data that are produced by machines. This is a smart real-time support system for exploring data. [9]

III. APPLICATIONS OF BIG DATA ANALYTICS AND IOT DATA

EDUCATION

The use of data in education began around eight years ago, due to the emergence of the field of learning analytics, which are today commonly named as Learning Management Systems (LMS). IT focuses on the measurement and analysis of student data to improve learning and learning environments. Big Data is making education more and more interesting, affordable and available. [7]

HEALTHCARE

The AI powered machine learning is continuing to gain foothold in healthcare sector. From various data collected from medical records of patients we can diagnose the person with the medical condition. With this data an early warning can be given in case if a person is suffering from a certain medical condition. [5]

URBAN PLANNING

The complexity of modern and smart cities has risen in the recent years. The traditional urban planning, design and management methods have reached their limits. [5]

AGRICULTURAL SECTOR

In this sector the farmers are undergoing a digital revolution. Big data is being used to deliver predictive awareness in farming operations and drive real-time operational decisions. In many developing countries the changes in the weather pattern, plantation, topography is less informed so, data is needed to note all these changes and inform them that might have a change. [7]

INDUSTRY

In manufacturing industries, the ability to successfully analyse big data can process flaws, update customer service, enhance production quality, increase efficiency that saves time and money. [5]

PUBLIC SERVICES DELIVERY

To improve productivity, performance, innovation and policy making processes, the government has an opportunity to tackle big data solutions. [5]

IV. CHALLENGES FACED

PRIVACY

The data being captured can be vital and privacy of it has to be considered. This happens when we try restoring particular or private data using the big data tools. Currently the privacy issues are in the data mining domain of big data. Presently there is no way these challenges and the way they manage the privacy and security. [8]

DATA STORAGE AND ANALYSIS

The volume of data over the years have increased drastically through mobile devices, remote sensing, radio sensing, etc. The diversity of the datasets have also increased over the years significantly. [2]

DATA MINING

The process of extracting data from IoT data. It provides an excellent solution to obtain graphical or analytical solution for the new data. [3]

SCALABILITY

It is the biggest challenge for big data analytics. The amount of data generated by IoT devices results in the natural shift of processor technology with increased number of cores. The hierarchical storage is critical here. Scalability consists of 2 dimensions: Horizontal and Vertical. [1]

Horizontal scalability: When more servers with less RAM and processors are added.

Vertical scalability: When new resources are added to the existing system to meet the requirements.

VISUALISATION

Due to the large amount of data the visualisation of this data is very difficult. So, for this big data and visualisation should work effortlessly. Graphical representation of the data provides the link between the data and gives proper interpretation. Challenges also arise due to the heterogenous and diverse nature of data. Different types of methods are needed for different kinds of data and a single method is unsuitable for this. Real-time analytics is another challenge. [4]

PREPROCESSING

The IoT data that is produced has a lot of unnecessary data that needs to be removed so a few steps have to be done before they are processed and analysed.

Data cleaning: Filling the misplaced values and correcting the inconsistencies in the facts.

Data integration: Data with different representation are put together.

Data transformation: The received data is generalized and normalized.

Data Reduction: Represents a reduced data in data warehouse. [3]

INTEGRATION

The data that is generated from IOT devices are in different format. This is the process of giving data in a single format that has data from diverse springs and then collects the assessment of data. [8]

INFORMATION SECURITY

The data security can be ensured by using techniques like authentication, authorization and encryption. Other security measures that big data face is variety of devices, lack of intrusion system, real-time security monitoring. There is a need to improve the security systems. Here the major difficulty would be creating a multi-level security, privacy preserved data model. [2]

V. CONCLUSIONS

IOT has become significant source of Big Data which is useless if it is not analyzed properly. This paper focuses on Big Data context concerning the Internet of Things. It describes the basic concepts of IoT and its architecture. It gives an elaborated structure of Gartner's 3 V's model. This paper enhances the understandability of the reader for the relation between IoT, Big data, and analytics. It familiarizes reader to different Big Data Analytics platforms which can handle various IoT datasets. After reading this paper, a reader will be aware of different platforms and will be able to select one for their particular problems.

VI. REFERENCE

[1] Kundhavai, K. R., & Sridevi, S. (2016). Iot and big data-the current and future technologies: A review. International Journal of Computer Science and Mobile Computing, 5(1), 10-14.

[2] Faisal, M., & Katiyar, V. INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY SECURITY CONCERNS IN IOT BASED SMART MANUFACTURING FOR INDUSTRY 4.0.

[3] Jenifer, J., & Priyadarsini, R. J. (2021). Big IoT data analytics: open research challenges.

[4] Acharjya, D. P., & Kauser, A. P. (2016). A survey on big data analytics: challenges, open research issues and tools. International Journal of Advanced Computer Science and Applications, 7(2).

[5] Rabah, K. (2018). Convergence of AI, IoT, big data and blockchain: a review. The lake institute Journal, 1(1), 1-18.

[6] Yu, H., Mihai, I. C., & Srivastava, A. (2021). Study and Research on IoT and Big Data Analysis for Smart City Development. Scalable Computing: Practice and Experience, 22(2), 215-225.

[7] Ali, A., Qadir, J., Rasool, R., Sathiaseelan, A., Zwitter, A., & Crowcroft, J. (2016). Big data for development: applications and techniques. Big Data Anal 1: 1–24.

[8] Haider, I., Haider, M. A., & Saeed, A. (2020, November). Big Data in Internet of Things: Architecture and Open Research Challenges. In 2020 IEEE 23rd International Multitopic Conference (INMIC) (pp. 1-6). IEEE.

[9] Gulia, P., & Chahal, A. (2020). Big data analytics for IoT. International Journal of Advanced Research in Engineering and Technology (IJARET), 11(6).

[10] Mahmud, M. S., Huang, J. Z., Salloum, S., Emara, T. Z., & Sadatdiynov, K. (2020). A survey of data partitioning and sampling methods to support big data analysis. Big Data Mining and Analytics, 3(2), 85-101.

[11] Bomatpalli, T., & Vemulkar, G. J. (2016). Blending iot and big data analytics. International Journal Of Engineering Sciences & Research Technolog, 192-196.

[12] Hussein, W. N., Kamarudin, L. M., Hussain, H. N., Hamzah, M. R., & Jadaa, K. J. (2019). Technology elements that influence the implementation success for big data analytics and IoT-oriented transportation system. Int. J. Adv. Trends Comput. Sci. Eng, 8(5), 2347-2352.

[13] Yu, W., Dillon, T., Mostafa, F., Rahayu, W., & Liu, Y. (2019). A global manufacturing big data ecosystem for fault detection in predictive maintenance. IEEE Transactions on Industrial Informatics, 16(1), 183-192.

[14] Kumar, S., Tiwari, P., & Zymbler, M. (2019). Internet of Things is a revolutionary approach for future technology enhancement: a review. Journal of Big data, 6(1), 1-21.

[15] Qi, M., & Wang, J. (2021). Using the internet of things E-government platform to optimize the administrative management mode. Wireless Communications and Mobile Computing, 2021.

[16] Wu, M., Lu, T. J., Ling, F. Y., Sun, J., & Du, H. Y. (2010, August). Research on the architecture of Internet of Things. In 2010 3rd international conference on advanced computer theory and engineering (ICACTE) (Vol. 5, pp. V5-484). IEEE.