Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 8, No. 4, August : 2023

BIG IoT DATA ANALYTICS UNLEASHED: GETTING AROUND OBSTACLES, FINDING RESEARCH OPPORTUNITIES AND USING RESOURCES

Dr.M.Deepa , Assistant Professor**,** Department of Computer Science, Pavai Arts & Science College for Women, Namakkal, Tamilnadu, India.
Dr.G.D.Praveenkumar, Assistant Professor, School of Computer Science, VET Institute of Arts and Science (Co-education) college, Erode.

Abstract
Since the beginning of the last decade, as Internet of Things (IoT) device miniaturization has increased, enormous volumes of data have been created. Such data, however, are useless without analytical capability. Many big data, IoT and analytics technologies have made it possible for users to gain insightful knowledge about enormous data produced by IoT devices. The domain lacks a thorough survey, and these solutions are still in their infancy. This study examines cutting-edge research initiatives focused on massive IoT data analytics. Big data analytics and IoT are related, as is mentioned. The novel architecture this article suggests for massive IoT data analytics also offers value. The types, techniques, and technologies for big data mining using large IoT data analytics are also covered. Additionally, other noteworthy application examples are provided. The potential that data analytics has brought to the IoT paradigm are then highlighted. Finally, future study prospects are offered for open research issues including privacy, massive data mining, visualization, and integration.

Keywords: Big Data, Smart City, Internet of Things, Data Analytics, Hadoop.

I. Introduction

Big data and the Internet of Things (IoT) are expanding quickly, advancing all fields of technology and industry while enhancing the advantages for both groups of people and corporations. The explosion of IoT-generated data has had a significant impact on the big data environment. Three characteristics of big data may be grouped together: (a) volume, (b) diversity, and (c) velocity [1]. Gartner initially developed these criteria to categorize the components of big data issues [2]. The capacity to comprehend and use enormous volumes of IoT data opens up a world of possibilities, including uses such as smart cities, smart grid systems, energy smart meters, and remote patient healthcare monitoring devices.Big data analytics are difficult because of the IoT environment's processing and data collecting through various sensors, which has contributed to the IoT's growing adoption. According to a forecast by the Global Data Corporation (IDC), the big data industry will surpass US$125 billion by 2019 [3].

IoT big data analytics are the procedures that are used to analyze various IoT data [4] in order to identify trends, unforeseen patterns, unnoticed connections, and new information [5]. Analyzing and managing vast volumes of data that have an impact on enterprises may be beneficial for both firms and people [6]. IoT big data analytics hence strives to help businesses, associations, and other organizations to better their comprehension of data and, as a result, make effective and informed choices. Huge volumes of unstructured data [22] that can be tapped using conventional technologies can be analyzed by data workers and engineers thanks to big data analytics [5]. Big data analytics also attempts to quickly extract accurate information utilizing data mining techniques that support prediction, trend identification, information discovery, and decision-making [7].

Data mining methodologies are often used for both problem-specific approaches and broad data analyses. In light of this, statistical and machine learning techniques are used. Due to the numerous sensors and items used during data collection, IoT data have unique traits from traditional big data acquired via systems, including heterogeneity, noise, diversity, and rapid expansion. According to statistics [8], there will be 1 trillion more sensors in use by 2030. Big data development will be impacted by this trend. IoT has the potential to provide a good solution for integrating data analytics and IoT into big data, which demands significant resources. IoT services offer the necessary resources and platform-intensive apps for optimal communication between the many deployed applications. Such a procedure is suited for meeting the criteria of IoT applications and can lessen certain difficulties in big data analytics in the future.

This fusion of technologies makes it more likely that IoT will be used in a more beneficial way. In addition, using IoT and big data integration solutions can aid in resolving problems with data processing, storage, analytics, and visualization tools. In a smart city, it can also help to improve cooperation and communication between diverse items [9].This poll, however, concentrated on IoT big data in the context of the analytics of enormous amounts of data. These are the contributions made by this survey. Investigations into cutting-edge big data analytics research projects are made. There is a suggested architecture for massive IoT data analytics. We explore a number of novel opportunities presented by data analytics in the IoT space. Useful examples are provided. Challenges in research that need to be resolved are noted and explored. Sections 2 through 4 of these contributions are provided. Section 5 has the conclusion.

II. BIG DATA ANALYTICS

Big data analytics entails the procedures of information seeking, data mining, and analysis aimed at enhancing business performance [10]. Large data sets with a variety of data kinds are examined using big data analytics [4] to uncover hidden connections, market trends, consumer preferences, and other pertinent business data [5]. An organization can cope with significant information that may have an impact on its operations with the aid of the capacity to evaluate massive volumes of data [6]. Big data analytics primary goal is to help organizations in the business world better comprehend their data so they can make effective decisions. Data scientists and miners can now study a significant volume of data that was previously inaccessible thanks to big data analytics [5].

Large amounts of structured, unstructured, and semi-structured data must be converted into a more comprehensible data and metadata format for analytical procedures in order to use big data analytics. These analytical tools algorithms need to find patterns, trends, and correlations in the data over a range of time spans [11]. These tools analyze the data and provide the results in tables, graphs, and geographical maps so that decisions may be made quickly. Because of the complexity of the data and the scalability of the underlying algorithms that underpin such operations, big data analysis is a significant problem for many applications [12]. Those tools analyze the data and provide the results in tables, graphs, and geographical maps so that decisions may be made quickly. Because of the extensive nature of the data and the scalability of the supporting technologies that enable such operations, big data analysis presents a significant problem for many applications.

A. Existing Analytics System

Depending on the needs of IoT applications, many analytical types are utilized [12]. These analytical categories real-time, offline, memory-level, business intelligence (BI) level, and huge level are covered in this part.

Real time Analytics

Instantaneous analytic are frequently carried out using sensor data. In this scenario, data are continually changing, necessitating the use of quick data analytics methodologies to get an analytical conclusion quickly. As a result, two current designs have been suggested for real-time analysis: memory-based computational infrastructures and concurrent computation clusters employing conventional relational databases. Examples of real-time intelligence technology include Greenplum [11] and Hana [13].

Off-Line Analytics

When a prompt response is not necessary, off-line analytics are employed. For instance, to cut the cost of data format translation, many Internet businesses deploy off-line analytical architecture based on Hadoop. These analytics increase the effectiveness of data collecting. Off-line analytics are conducted using systems like SCRIBE , Kafka, Time Tunnel and Chukwa , which may also meet the requirements of data collecting [13].

Massive Analytics

When the scale of the data exceeds the capabilities of both standard databases and BI analysis products, massive analytics is used. Massive analytics employs map reduce for data processing and the Hadoop distributed file system for data storage. By obtaining valuable information from data, massive analytics aids in building a company's foundation and improves its competitiveness. Massive analytics also gathers precise data that takes into account the risks associated with each business action. Massive analytics also successfully delivers services [13].

B. Big Data Analytics System

Big data analytics seek to instantly extract knowledge that aids in creating forecasts, seeing current patterns, uncovering buried facts, and finally, making judgments . Data mining methods are often used for both problem-specific approaches and broad data analyses. In light of this, statistical and machine learning techniques are used. Big data evolution also alters the needs for analytics. For the purposes of our discussion, big data analytics requires the same or faster processing speed than traditional data analytics with the lowest cost for high-volume, high-velocity, and high-variety data, despite the fact that the requirements for efficient mechanisms lie in all aspects of big data management [12], such as capturing, storage, preprocessing, and analysis. There are several solutions for big data analytics and these solutions are always being developed and improved to make them compatible with emerging big data trends. The majority of strategies are produced using data mining algorithms in accordance with a certain scenario and data mining plays a significant role in data analytics and web mining analytics [25]. When assessing and selecting a suitable technique for decision-making, knowledge of the different big data analytics choices is essential. We discuss numerous strategies that may be used for various big data case studies in this section. Some of these analytics techniques work well with large IoT data sets. More big data insights come from diverse and enormously large data sources [7].

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 8, No. 4, August : 2023

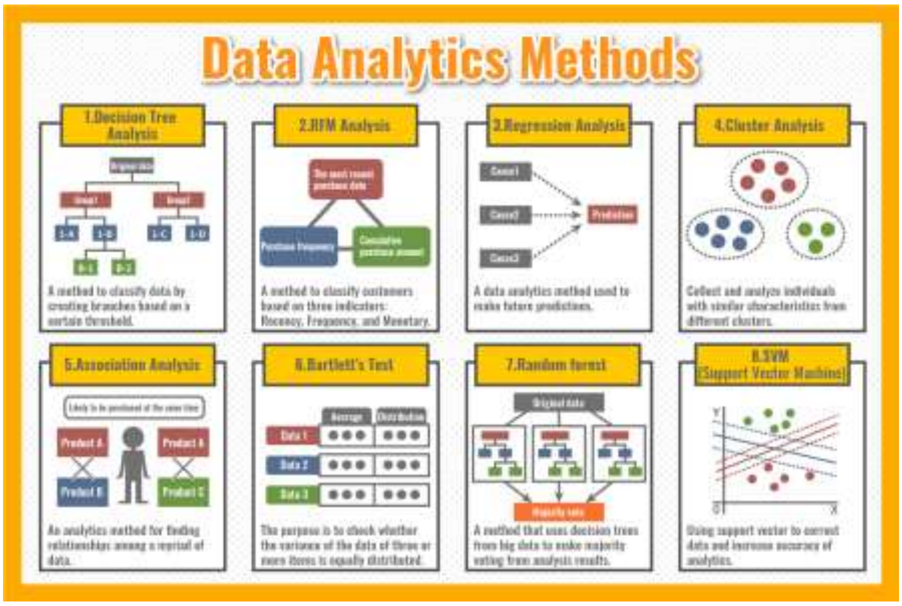

Figure 1: Overview of Big Data Analytics Methods

SVM, a classification strategy for big data analytics, is effective for analyzing data trends and forming groups. SVM analyzes data patterns and categories them using statistical learning theory. Text classification, pattern matching, health diagnostics, and business are a few areas where SVM classification in big data analytics is applied. Comparable to this, KNN is frequently created to offer effective methods for uncovering hidden patterns from large data sets, so that objects are returned are comparable to the preset category. Case-based learning enhances the KNN algorithm for use in scientific studies, high-dimensional data and anomaly detection. [15] While using several artificial intelligence and data mining approaches, classification has further expansions.

Another data mining approach utilized in big data analytics is clustering. Clustering in contrast to classification, employs an unsupervised learning strategy that divides given items into groups based on their unique significant properties [16]. As shown in Figure 1, it is easy to manipulate data when several items are grouped together into clusters. Hierarchical clustering and partitioning are two popular clustering techniques. As tiny clusters of data items are continuously combined, a hierarchical tree is formed, and agglomerative clusters are produced. In contrast, separating a single cluster that comprises all data items into smaller, useful clusters results in the formation of divisive clusters [17].

The most important uses of big data analytics are in market research and corporate decision-making. In order to assess market trends, customer purchasing patterns, and forecasted product demand, association rule mining entails finding intriguing associations between many objects, events, or other entities. In association rule mining [18], rules are found and developed depending on how frequently they appear in both numerical and non-numerical data. In association rules, there are two ways to handle data. In order to find interaction correlations, sequential data processing first employs priori-based algorithms such MSPS [19] and LAPINSPAM [20]. Temporal sequence analysis, which employs algorithms to examine event patterns in continuous data, is an important method of data processing under the association rule.

III. USE CASES

A variety of use cases for massive IoT data analytics are presented in this section. The use cases have been chosen based on which ones are most often utilized in IoT applications and how much data may be generated for analytics, even if they are all important to IoT applications.

A. Smart Metering

One of the IoT application use cases for smart meters generates a lot of data from several sources, including smart grids, tank levels, water flows, and silos stock computation, which requires processing over an extended period of time even on a strong system [21]. The electronic recording of electric energy consumption data between the meter and the control system is done by a smart meter. The decision-makers ability to forecast power use is helped by the collection and analysis of smart meter data in an IoT context. A smart meter's analytics may also be utilized to estimate demand in order to avoid crises and achieve strategic goals through precise pricing schemes.

B. Smart Transportation

An IoT-based use case that tries to promote the idea of smart cities is a smart transportation system. For the operation of smart cities, a smart transportation system aims to implement robust and cutting-edge communication technology. Weather circumstances like torrential rain and dense fog have an impact on traditional transportation systems, which rely on image processing. As a result, the acquired image might not be easily discernible. An effective method for intelligent vehicle monitoring, tracking, and identification is the construction of an e-plate system [22] employing RFID technology. Additionally, integrating IoT with automotive technology would allow traffic congestion management to perform a great deal better than the current infrastructure.

C. Smart Supply Chains

Embedded sensor technologies can communicate bidirectional and provide remote accessibility to over 1 million elevators worldwide [23]. 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 [24]. 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 in transit 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.

D. Smart Agriculture

One advantageous use of massive IoT data analytics is smart agriculture. In the use case for smart agriculture, sensors are the actors. They are positioned in fields to gather information on the soil's moisture content, plant trunk diameter, microclimate, humidity level, and weather forecasting. Sensors use networks and communication tools to communicate the data they collect. Examples of activities taken in response to recommendations from big data analytics include automatic temperature management in accordance with harvesting needs, timely and regulated watering, and humidity control for fungus avoidance.

E. Smart Traffic Light System

The IoT sensors and gadgets used by the smart traffic light system are localized nodes that communicate with them to identify the presence of cars, bicycles, and pedestrians. The aforementioned nodes interact with nearby traffic lights to control green traffic signals and gauge the velocity and length of approaching vehicles [77]. In order to perform necessary tasks, such as adjusting timing cycles in accordance with traffic conditions, sending informative signals to neighboring nodes, and spotting achieving vehicles that use IoT sensors and devices to avoid long lines or accidents, IoT data collected using the system requires real-time analytics processing. Additionally, intelligent traffic light systems may upload their IoT data to cloud storage for additional analyses.

IV. OPPOURTUNITIES

IoT is presently regarded as one of the most significant technological changes. IoT now offers a variety of options for big data analytics.

A. E-Commerce

Huge IoT data analytics provides designed effectively tools for processing huge data in real-time, which deliver findings quickly enough for decision-making. Big IoT data show characteristics of heterogeneity, growing volume, and real-time data processing. Big data and IoT coming together creates new potential and problems for creating a smart environment. There are several uses for big IoT data analytics across almost all sectors of the economy. However, e-commerce, revenue development, expanded customer size, accuracy of sale forecast results, product optimization, risk management, and enhanced client segmentation are the primary success areas of analytics.

B. Smart Cities

Through the use of the right analytics platform and infrastructure, massive IoT data may be analyzed to provide new chances for efficiency benefits in smart cities. In a smart setting, many gadgets connect to the Internet and exchange data. Additionally, the development of cloud computing technologies has greatly decreased the cost of data storage. Analysis skills have advanced dramatically. Therefore, the use of big data in a smart city has the potential to impact every area of a country's economy. Recent developments in big data technology have included Hadoop with the YARN resource management, which can support and handle a variety of workloads, real-time processing, and streaming data intake.

C. Retail & Logistics

IoT is anticipated to be a significant new technology in the logistics and retail sectors. RFID is used in logistics to track crates, pallets, and containers. Additionally, significant improvements in IoT technology may help merchants by offering a number of advantages. IoT devices, however, produce a significant quantity of data every day. As a result, strong data analytics helps businesses to learn from the vast volumes of data generated by IoT technology.Customer shipping experiences may be enhanced by applying data analytics to logistic data sets. Retail businesses may also increase their profits by studying consumer information that forecasts market trends and product demand. Pricing strategies and seasonal promotions may be designed effectively to optimize profits by analyzing client data.

D. HealthCare

Smart health monitoring technology has expanded dramatically in recent years. Huge volumes of data are produced by these gadgets. Healthcare professionals can more effectively analyze patients' physical states by using data analytics on data from fatal monitors, electrocardiograms, temperature monitors, or blood glucose level monitoring, for example. Data analytics also makes it possible for medical personnel to detect dangerous illnesses in their early stages to aid in saving lives. Additionally, data analytics raises the standard of clinical care and guarantees patient safety. Additionally, a physician's profile may be examined by looking at the patients' medical histories, which can enhance client happiness, client acquisition, and client retention.

V. CONCLUSION AND FUTUREWORK

With the spread of smart and sensor devices over the past several years, the growth rate of data output has dramatically accelerated. Currently, the interplay between IoT and big data requires processing, converting, and analyzing massive volumes of data frequently. This survey was carried out within the framework of massive IoT data monitoring. We started by looking through current analytics options. There was also discussion on the connection between IoT and enormous data analytics. Additionally, we put out a design for massive IoT data analytics. A presentation on big data processing kinds, techniques, and technologies was also made. Additionally, several reliable use examples were given. Additionally, we discussed numerous potential created by data analytics in the IoT paradigm to further study the subject. Future research directions were explored for a number of unsolved research problems. Finally, we came to the conclusion that massive IoT data analytics technologies were still in their infancy. Real-time analytics tools with rapid insights will be necessary in the future.

REFERENCES

[1] P. Tiainen, ''New opportunities in electrical engineering as a result of the emergence of the Internet of Things,'' Tech. Rep., AaltoDoc, Aalto Univ., 2016.
[2] M. Beyer, ''Gartner says solving 'Big Data' challenge involves more than just managing volumes of data,'' Tech. Rep., AaltoDoc, Aalto Univ., 2011.
[3] J. Gantz and D. Reinsel, ''Extracting value from chaos,'' IDC Iview, vol. 1142, pp. 1–12, Jun. 2011.
[4] R. Mital, J. Coughlin, and M. Canaday, ''Using big data technologies and analytics to predict sensor anomalies'' in Proc. Adv. Maui Opt. Space Surveill. Technol. Conf., Sep. 2014, p. 84.
[5] N. Golchha, ''Big data-the information revolution,'' Int. J. Adv. Res., vol. 1, no. 12, pp. 791–794, 2015.
[6] P. Russom, Big Data Analytics. TDWI, 4th Quart., 2011.
[7] C.-W. Tsai, ''Big data analytics: A survey,'' J. Big Data, vol. 2, no. 1, pp. 1–32, 2015.
[8] M. Chen, Related Technologies in Big Data. Heidelberg, Germany: Springer, 2014, pp. 11–18.
[9] Z. Khan, A. Anjum, and S. L. Kiani, ''Cloud based big data analytics for smart future cities,'' in Proc. IEEE/ACM 6th Int. Conf. Utility Cloud Comput. (UCC), Dec. 2018, pp. 381–386.
[10] O. Kwon and N. B. L. Shin, ''Data quality management, data usage experience and acquisition intention of big data analytics,'' Int. J. Inf. Manage., vol. 34, no. 3, pp. 387–394, 2014.
[11] S. Oswal and S. Koul, ''Big data analytic and visualization on mobile devices,'' in Proc. Nat. Conf. New Horizons IT-NCNHIT, 2013, p. 223
[12] L. Candela and D. P. C. Pagano, ''Managing big data through hybrid data infrastructures,'' ERCIM News, vol. 89, pp. 37–38, Jun. 2012.
[13] Chen et al., ''Data mining for the Internet of Things: Literature review and challenges,'' Int. J. Distrib. Sensor Netw., vol. 12, Aug. 2015.
[14] R. Luss and A. d'Aspremont, ''Predicting abnormal returns from news using text classification,'' Quant. Finance, vol. 15, no. 6, pp. 999–1012, 2015.
[15] M. W. Pfaffl, ''A new mathematical model for relative quantification in real-time RT-PCR,'' Nucl. Acids Res., vol. 29, no. 9, p. e45, 2010.
[16] M. Waas, ''Beyond conventional data warehousing-massively parallel data processing with Greenplum database,'' in International Workshop on Business Intelligence for the Real-Time Enterprise. Berlin, Germany: Springer, 2018

[17] Rabkin and R. H. Katz, ‘‘Chukwa: A system for reliable large-scale log collection,’’ in Proc. LISA, Nov. 2020, pp. 1–15.
[18] Bifet, G. Holmes, R. Kirkby, and B. Pfahringer, ‘‘MOA: Massive online analysis,’’ J. Mach. Learn. Res., vol. 11, pp. 1601–1604, May 2018.
[19] Mukhopadhyay, U. Maulik, S. Bandyopadhyay, and C. A. C. Coello, ‘‘A survey of multiobjective evolutionary algorithms for data mining: Part I,’’ IEEE Trans. Evol. Comput., vol. 18, no. 1, pp. 4–19, Feb. 2014.
[20] M.-Y. Su, ‘‘Real-time anomaly detection systems for Denial-of-Service attacks by weighted K-nearest-neighbor classifiers,’’ Expert Syst. Appl., vol. 38, no. 4, pp. 3492–3498, 2021.
[21] M. Muja and D. G. Lowe, ‘‘Scalable nearest neighbor algorithms for high dimensional data,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 11, pp. 2227–2240, Nov. 2019.
[22] Hu, ‘‘Data-driven method based on particle swarm optimization and K-nearest neighbor regression for estimating capacity of lithium-ion battery,’’ Appl. Energy, vol. 129, pp. 49–55, Sep. 2020.
[23] N. Tohamy, ‘‘What you need to know about the Internet of Things,’’ MHD Supply Chain Solutions, vol. 45, no. 3, p. 32, 2015.
[24] 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/
[25] GD Praveenkumar, R Gayathri, “A Process of Web Usage Mining and Its Tools” International Journal of Advanced Research in Science, Engineering and Technology, Vol 2, no. 11, 2015.