Chapter 9
Big Data Analytics and Its Applications in IoT

Shaila S. G.
Dayananda Sagar University, India

Bhuvana D. S.
Dayananda Sagar University, India

Monish L.
Dayananda Sagar University, India

ABSTRACT

Big data and the internet of things (IoT) are two major ruling domains in today's world. It is observed that there are 2.5 quintillion bytes of data created each day. Big data defines a very huge amount of data in terms of both structured and unstructured formats. Business intelligence and other application domains that have high information density use big data analytics to make predictions and better decisions to improve the business. Big data analytics is used to analyze a high range of data at a time. In general, big data and IoT were built on different technologies; however, over a period of time, both of them are interlinked to build a better world. Companies are not able to achieve maximum benefit, just because the data produced by the applications are not utilized and analyzed effectively as there is a shortage of big data analysts. For real-time IoT applications, synchronization among hardware, programming, and interfacing is needed to the greater extent. The chapter discusses about IoT and big data, relation between them, importance of big data analytics in IoT applications.

OVERVIEW OF IOT:

Internet of Things (IoT) deals with global network of physical objects or things interconnected through the internet. These are embedded with electronics, software, sensors, and network connectivity. The major advantage here is the objects can be sensed and controlled remotely across the network. IoT is helping us in building the physical world based smart environment. In IoT applications, things or ob-

jects refers to a wide variety of devices. IoT concepts have made many contributions to society right from an automated lighting system until automated driving cars. The IoT is a new revolution that has low costs and high technology implementations. According to expert's analysis by 2020, there will be at least 50 billion devices/objects connected via the internet. IoT is just not concerned with connecting things, it also helps in information exchange which leads to knowledge development. IoT deals with a series of steps such as filtering, processing, categorizing, condensing and contextualizing the data into information. Further, this information is organized and structured to infer the knowledge. Thus knowledge is defined as information obtained from organizing and structuring and put into action to achieve the specific business objective.

IoT helps in handling the resources efficiently, support in minimizing the human effort, time, offers improved security, helps in automation using Artificial Intelligence. IoT devices connecting the real world use certain set of standards and protocols like Zigbee, Bluetooth, MQTT (Message Queue Telemetry Transport), CoAP (Constrained Application Protocol), NFC (Near Field Communication), AMQP (Advanced Message Queueing Protocol) to access the internet. The data generated by the IoT devices is huge in terms of Velocity, Volume and Variety and this huge data set is termed as Big Data.

OVERVIEW OF BIG DATA:

In general, the Big Data is defined using 3Vs such as Velocity, Volume and Variety. Velocity is the rate at which the data grows and how fast the data is gathered for analysis. Volume refers to the enormous data being generated, whereas Variety refers to the different types of data being generated like structured, semi- structured and unstructured data. There is a fourth V to describe the big data, it is referred as veracity, which includes availability and accountability. In general, data generated by sensors in IoT applications will be considered as raw data. This raw data need to be fine-tuned before it is used by the decision makers. In general, during data collection, it is noticed that data missing, data redundant, data in the wrong format, etc. Hence, preprocessing is needed to get the relevant data in required format to avoid erroneous and misleading outputs that reduce the efficiency. According to the B2B report of data quality index test, it was shown that every data repository has got 40% bad data in which 15% are duplicate, 8% are missing, 11% are invalid and 6% comes from malicious or unauthorized users. These data weaken the organization's marketing and financial automation, increase the resource consumption and cost and leads to lower customer satisfaction and invalid reports. Thus, there is a great need of data preprocessing. Apache Hadoop, Apache Spark, etc are used for this purpose. Hence, we use techniques such as data mining, analytical tools and machine learning to extract the useful information from the big data.

RELATION BETWEEN BIG DATA AND IOT

The real value of an IoT application is determined by creating smarter products, delivering new business outcomes, and also delivering the intelligent insights. As lots and lots of devices are interconnected in an IoT network massive amount of data inflow is encountered. Researchers have predicted that by 2020 1.7MB of data will be produced each second by a single person. Researchers have found that connected cars alone have the capability to send 25 gigabytes of data to the cloud every hour. In today's world, it is not possible to imagine a world without data and processing the data manually or using a traditional

Related Content

A Generic, Multi-Period and Multi-Partner Cost Optimizing Model for Cloud-Based Supply Chain
Goknur Arzu Akyuzand Mohammad Rehan (2016). *International Journal of Cloud Applications and Computing (pp. 55-63).*
www.irma-international.org/article/a-generic-multi-period-and-multi-partner-cost-optimizing-model-for-cloud-based-supply-chain/159852

Resource Allocation in Cloud Computing Environment using AHP Technique
Anil Singh, Kamlesh Duttaand Avtar Singh (2014). *International Journal of Cloud Applications and Computing (pp. 33-44).*
www.irma-international.org/article/resource-allocation-in-cloud-computing-environment-using-ahp-technique/111146

On the Exploration of Equal Length Cellular Automata Rules Targeting a MapReduce Design in Cloud
Arnab Mitra, Anirban Kundu, Matangini Chattopadhyayand Samiran Chattopadhyay (2018). *International Journal of Cloud Applications and Computing (pp. 1-26).*
www.irma-international.org/article/on-the-exploration-of-equal-length-cellular-automata-rules-targeting-a-mapreduce-design-in-cloud/202387

Hard Clues in Soft Environments: The Cloud's Influence on Digital Forensics
Andrea Atzeni, Paolo Smiragliaand Andrea Siringo (2015). *Handbook of Research on Security Considerations in Cloud Computing (pp. 258-284).*
www.irma-international.org/chapter/hard-clues-in-soft-environments/134295

Cloud in Digital Government: Problems and Perspectives in the Case of Azerbaijan
Farhad Yusifov (2023). *Convergence of Cloud Computing, AI, and Agricultural Science (pp. 33-66).*
www.irma-international.org/chapter/cloud-in-digital-government/329128