12

Big IoT Data Analytics in Fog Computing

Manash Kumar Mondal[1]*, Riman Mandal[2] and Utpal Biswas[1]

[1]*Department of Computer Science and Engineering, University of Kalyani, Kalyani, WB, India*
[2]*Department of Computer Science and Engineering, Annasaheb Dange College of Engineering and Technology, Ashta, Sangli, Maharashtra, India*

Abstract
IoT data analysis is a significant role in economic growth, social development and people's life. So, IoT data analysis is the combination of artificial intelligence, machine learning and big data. Fog computing is a way of bringing to market the growing amounts of data gathered for Internet-of-Things (IoT) devices. It works by pooling the computational power from multiple nodes connected through the cloud, each node being able to support the requests generated by its IoT sensors. The more devices you add to the network, the more powerful it becomes, but it has its limits in terms of memory and processing speed. In this chapter, we have discussed a new model by combining these technical methods and it has a strong function in the effective analytics of IoT-generated big data in fog computing environments. This chapter highlights a brief introduction to fog computing, the generation of big data, sources of big data, and how fog computing is used to analyze IoT-generated big data. The chapter also covers the regions of choosing fog computing for big data analytics over the cloud. Additionally, this chapter highlights big data characteristics and various applications of big data. The fog engine is used to analyze data in the fog node.

Keywords: Big data, IoT, fog computing, big data analytics, fog-engine

12.1 Introduction

Fog computing is the future of data analytics and one of the most significant innovations in recent years. This new paradigm shifts the way we think about data storage, processing and analysis away from the traditional infrastructure

**Corresponding author*: manashkumarmondal@gmail.com

Chandan Banerjee, Anupam Ghosh, Rajdeep Chakraborty and Ahmed A. Elngar (eds.) Fog Computing for Intelligent Cloud IoT Systems, (279–308) © 2024 Scrivener Publishing LLC

such as server farms to a much more distributed cloud architecture that can be managed at any scale. Fog computing is a widely used, most efficient and cost-effective approach to deploying big data analytics. With thousands of systems all operating inside the cloud, the application moves from local computations to distributed computation. This makes it feasible for large-scale dimension data mining and analysis. But the problem that remains unsolved is how to do big data analytics fast, to get insights in a short time frame.

Fog computing is a mechanism for processing data in which the information is not contained in any single place. Data is placed in another place and analyzed. This helps to make it cost-efficient as well as more secure than traditional centralized computing where all data processing occurs at a specific place. Fog computing can be used for many applications such as social media, marketing, financial services etc. Furthermore, fog computing has seen some interesting use cases such as smart cities, big data analytics, network management, etc. The third one is a graphical user interface (GUI). The GUI is needed for managing smart devices and applications [1, 2].

A typical Internet of Things system collects and exchanges data in real-time. There are three parts to an IoT system. The first one is a smart device. Smart devices like security cameras, television or other devices experience computational capabilities [3]. These devices collect data from the environments, taking inputs from the users and sharing data via the internet using IoT applications. The second one is the internet of things application. The IoT application uses machine learning or AI technology to analyze the data generated by smart devices to make proper decisions [4–6].

IoT devices produce an enormous amount of data for several reasons:

High Volume of Connected Devices

IoT networks can consist of thousands or even millions of devices that are interconnected and constantly communicating with each other. Each device generates a stream of data, which accumulates over time and leads to massive amounts of data.

High Frequency of Data Collection

IoT devices often collect data at high frequencies, such as every second or even every millisecond. This means that data is generated rapidly, leading to large volumes of data being collected in a short period.

Richness of Data

IoT devices can generate a variety of data types, including sensor data, video data, audio data, location data, and more. This data can be highly detailed and provide valuable insights, but it also requires a lot of storage and processing power.

Continuous Data Generation

IoT devices are always on and always collecting data, generating a continuous stream of data that can quickly add up.

Edge Computing

Numerous IoT devices are furnished with sensors and computing capabilities that allow them to process and analyze data at the network's edge, instead of sending it all to a central location for processing. This guides to better data being produced at the edge and the need for more storage and processing capabilities. Edge computing cloud computing and machine learning are the three main technologies that are widely used in IoT systems. Edge computing corresponds that the technology that makes smart devices more computationally capable. In this edge computing paradigm, smart devices are not only used for sending and receiving data but also used for computing purposes. It reduces the data transfer latency and decreases response time. In IoT systems, cloud computing is used for data repository and managing of IoT devices. It also provides data access to all other smart devices over the network. Machine learning represents the collection of software or algorithm which is used to process the data to make a suitable judgement in real-time [2, 7].

To process and analyze big data, fog computing and cloud computing are two essential concepts that can be used. However, in some scenarios, fog computing may offer advantages over cloud computing. Here are some reasons why:

Lower Latency

The most useful feature that makes the fog an edge over the cloud is latency. Fog computing brings the computing power closer to the edge devices of the grid, which can decrease the latency of data transfer and processing. This is particularly critical for real-time applications, such as autonomous vehicles or industrial IoT, where milliseconds can make a significant difference.

Bandwidth Optimization

It is one of the most useful features of fog computing. By processing data locally, the fog computing can decrease the quantity of data that requires to be transferred to the cloud for processing. This can help optimize network bandwidth and reduce the cost of data transfer, especially for applications that generate large volumes of data.

Security

Fog computing can offer improved security compared to cloud computing, as data can be computed and analyzed locally (near the sources) without transmitting over the internet. This can be particularly important for applications that deal with sensitive data, such as healthcare or financial services.

Resilience

Fog computing can provide greater resilience in the face of network disruptions or failures, as the local processing capability can continue to function even when connectivity to the cloud is lost.

Overall, fog computing is more promising than cloud computing for big data analytics in scenarios where low latency, bandwidth optimization, security, and resilience are critical factors. However, it is essential to mention that fog computing is not a substitute for cloud computing but relatively a complementary technology that can work alongside cloud computing to deliver more efficient and effective big data processing and analysis.

The rest of the sections of the chapter are organized as follows: Section 12.2 contains some related studies regarding big data analytics in fog computing. The main motivation behind the writing of this chapter describes in Section 12.3. Section 12.4 gives a concise idea of fog computing, the characteristics of fog computing, the fog nodes and various deployment and services models. A brief discussion of big data and various sources for the generation of big data discusses in Section 12.5. Section 12.6 illustrates the details of big data analytics in fog computing. Ultimately, the conclusion and future dimensions of research are covered in Section 12.7.

12.2 Literature Review

A significant amount of literature has been published on big IoT data analytics in fog computing. These studies contribute a lot to this literature. The cloud is one of the prominent options for analyzing big data. Most of the investigation regarding big data analytics is carried out in cloud computing environments [8–11]. Big data analytics (BDA) using machine learning (ML) and cloud computing is proposed in the article. Large amounts of data are produced by smart homes, the article proposed a big data analytics system using fog computing [12]. To provide the services of smart cities with a huge number of infrastructure management using hierarchical fog computing is proposed in studies [13, 14]. FogGIS a fog computing-based framework is proposed for geo-positional data mining and analytics [15]. FogData a telehealth monitoring model using fog computing is suggested by Dubey *et al*. [16]. A multiple-tier fog computing is proposed for large-scale data analytics in various smart city applications. The study used the Raspberry Pi system for deploying the fog model for resource management and tackling problems like pollution, traffic management, public budget, and health data. A novel state-of-the-art regarding fog computing is highlighted in article [17]. Liu *et al*. suggested an outline of fog computing and explain several architectures and cover many latency-sensitive applications [18]. This paper provided an outline of fog computing and its application to IoT. The authors discussed various fog computing architectures and the challenges and opportunities of using fog computing for IoT data analytics [19]. The authors discussed various smart city scenarios, such as traffic management, energy management, and public safety, and how fog computing can enable real-time data analytics and decision-making [20]. Ahmed *et al*. proposed a fog computing-based framework for real-time IoT data analysis. The researchers described a case study involving a smart healthcare system and demonstrate how fog computing can provide timely and accurate insights for healthcare professionals [21]. Collectively this literature supports the importance of the study. These papers demonstrate the potential of fog computing for IoT data analytics and provide insights into the challenges and opportunities of using fog computing in various IoT applications. The researchers described a case study involving a power plant and demonstrated how fog computing can reduce maintenance costs and downtime by enabling real-time predictive analytics [22].

12.3 Motivation

The Internet of Things (IoT) is nothing but a device that attaches to networks that allow communication and exchange of information with other intelligent devices via the internet. The Internet of Things (IoT) has been one of the numerous exciting developments in recent years. This technology has allowed companies to reimagine how and what they do, mainly for their business strategy. With IoT and cloud technologies, businesses can create better products, improve customer relationships and increase revenue.

The IoT becomes a part of daily life. IoT provides several applications including agriculture, consumer applications, healthcare, Insurance, Manufacturing, Retail, transportation, Utilities/Energy, traffic monitoring, hospitality, water supply, fleet management, smart pollution control, smart cities etc.

Intelligent Internet of things devices produces an enormous amount of data. Generally, a massive amount of data refers to as big data. Using cloud computing generally analysts such big data for better decision makings. The main motivation behind this chapter that how such big data can be analyzed near the source. Cloud computing encountered some limitations like data loss or theft, data leakage, service hijacking, DoS attack, technology vulnerabilities high latency time etc. Cloud computing generally follows *pay-as-you-go* service model or subscription-based model. Although, deploying a fog environment requires a high establishment cost but in long run, it will be more cost-effective than the cloud [23, 24].

The motivation behind writing a chapter on IoT-generated Big Data analytics using fog computing is driven by the growing importance of these technologies in modern business operations. With the proliferation of IoT devices, organizations are producing massive amounts of data that can be analyzed to gain valuable insights into customer behaviour, product performance, and operational efficiency.

However, processing and analyzing this data can be a significant challenge, as it requires powerful computing resources and high-speed connectivity. Moreover, the conventional approach of processing data in the cloud can lead to latency and network congestion, which can impede real-time insights and hinder decision-making.

Fog computing has emerged as a profitable resolution to these challenges, by bringing computing aids nearer to the boundary of the network, where the data is generated by some intelligent devices. However, in processing data near the edge, fog computing can reduce latency and network congestion, enabling faster and more efficient data analysis. Moreover, fog computing can also help diminish the cost of IoT Big Data analytics by minimizing the need for data transmission and storage.

The potential advantages of IoT-generated big data analytics using fog computing are vast and include enhanced customer experience, improved operational efficiency, and increased innovation. As more organizations adopt IoT technologies and generate massive amounts of data, the need for efficient and cost-effective data processing and analysis will only continue to grow.

Thus, the motivation behind writing a chapter on IoT big data analytics using fog computing is to deliver a comprehensive overview of this important and emerging field and to highlight the potential benefits and challenges associated with its adoption. By doing so, we hope to contribute to a better insight into the role that big IoT data analytics and this fog computing can play in driving innovation and improving business operations.

12.4 Fog Computing

A layered approach called fog computing enables universal access to shared computational resources. The concept of fog computing consists of computational units called fog nodes, placed halfway between intelligent end-edge-devices and the topmost layered cloud services, creates it more effortless to develop dispersed, latency-aware apps and services [25]. The fog nodes support a standard data management and communication framework and it is also context-aware. They can be grouped either vertically (to promote solitariness), horizontally (to facilitate association), or in relation to the latency space between fog nodes and the intelligent end devices. The fog computing reduces the time it takes for subsidized apps to receive requests and respond, and it gives end devices access to regional computing resources and, if necessary, grid connectivity to centralized services [26, 27].

Figure 12.1 illustrates a layered architecture of fog computing in the larger context of a cloud-based environment supporting intelligent edge devices. Both the centralized (cloud) service and fog computing are not thought to be necessary layers for these ecosystems in order for smart end-device capabilities to be supported. Depending on the best method for enabling end device functionality, various use case scenarios could have various architectural designs. The decision to use this model was made with the goal of accurately portraying a complicated architecture that uses fog computing services [28, 29].

Cloud computing can offer significant benefits for IoT data analytics, such as scalability, flexibility, and cost savings. However, there are also limitations to consider, including:

Latency: IoT devices often generate large volumes of data in real time, which requires fast processing to provide timely insights. However, cloud computing can introduce latency due to data transmission and processing delays, which can affect the accuracy and relevance of the analytics results.

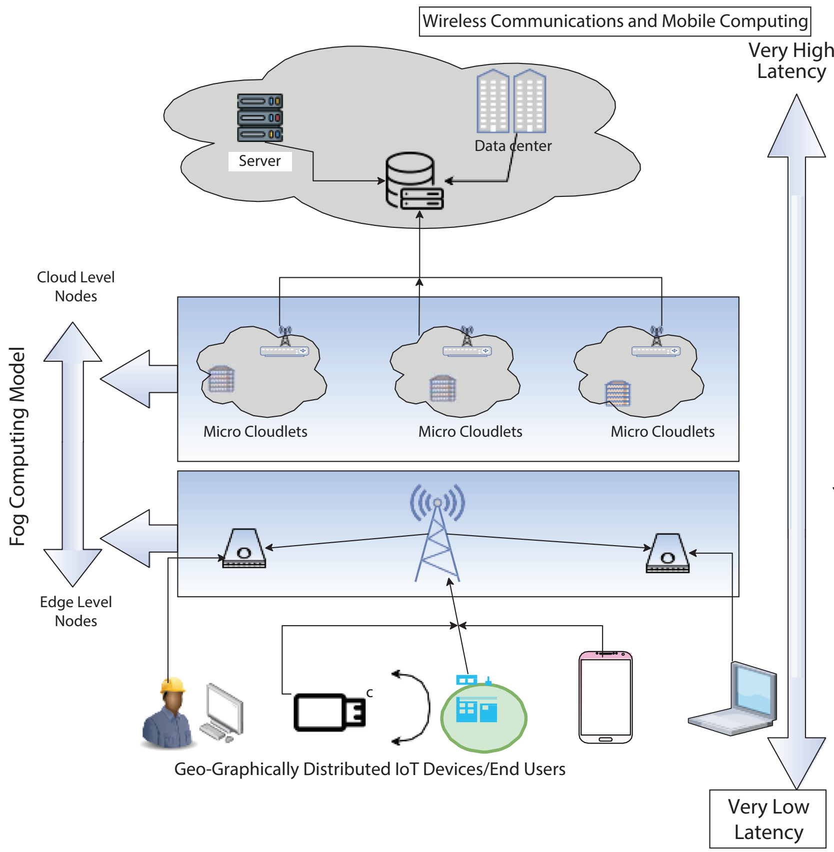

Figure 12.1 Fog computing in larger perspective.

Network bandwidth: IoT devices can produce a huge amount of data, which requires high bandwidth networks to transport data to the cloud for processing. However, many IoT devices operate in remote areas or have limited network connectivity, which can result in data transmission issues.
Security: IoT devices often collect sensitive data, such as private or economic statements, which requires secure storage and processing. However, cloud computing introduces potential security risks, such as data breaches and unauthorized access.
Data privacy: IoT devices may generate personal or sensitive data that is subject to privacy regulations, such as GDPR or CCPA. Cloud computing

introduces the possibility of data exposure or unauthorized sharing, which can result in legal or reputational consequences.

Cost: Cloud computing can be expensive, especially for large-scale IoT deployments that require high-performance computing resources. The cost of keeping and processing enormous volumes of data can quickly add up, which can impact the ROI of IoT analytics projects.

To mitigate these limitations, organizations may need to consider hybrid cloud solutions that combine on-premises and cloud-based computing resources, or edge computing solutions that process data locally on the IoT devices themselves.

12.4.1 Fog Node

The fog node is a crucial component in fog architecture. The fog nodes provide computing resources to smart endpoints or access networks through close connections. Routers, switches, gateways, servers, etc. are examples of physical fog nodes, whereas virtual fog nodes are examples of virtualized switches, virtualization software, cloudlets, etc. A fog node is aware of its cluster's physical location as well as its logical position inside it. Moreover, the fog nodes deliver several types of data governance and communication services as needed between the edge layers of the network, where end smart devices are located, and the centralized (cloud) computing resources. To serve the service, fog nodes can be set up as independent fog nodes that are in contact with one another, or they can be unified to form groups that offer horizontal scalability over-dispersed geo-locations via mirroring or extension mechanisms [30]. To deploy a specific fog computing facility, the fog nodes might work centrally or decentralized.

Fog computing is a fully distributed computational instance that expands cloud computing to the edge or middle layer of the network, closer to where data is generated and consumed. It provides a service model that is created to support the deployment of applications and assistance in distributed circumstances, such as the IoT, where there are large numbers of connected devices generating data that requires to be analyzed in real-time [31].

The fog computing service model is based on the principles of resource sharing, virtualization, and service-oriented architecture. It consists of the following components [32]:

Fog Nodes: These are the physical computing instruments that make up the fog computing infrastructure. They can be any type of computing device, such as routers, networking switches, gateways, or servers that have the sufficient processing power, memory, and storage capacity to support fog computing applications.

Fog Services: These are the software applications and services that are deployed on fog nodes. They can be custom-built applications or off-the-shelf software that provides a specific functionality, such as data analytics, machine learning, or security.

Fog Broker: This is a software component that acts as an intermediary between fog nodes and cloud services. It helps to manage the allocation of fog services across the fog computing infrastructure and provides a unified interface for users and applications to access fog services.

Fog Orchestrator: This is a software component that manages the deployment, configuration, and monitoring of fog services. It ensures that fog services are deployed to the appropriate fog nodes based on their resource necessities and workload characteristics.

The fog computing service model provides several benefits, including lower latency, reduced network bandwidth, improved security, and greater resilience. It enables the deployment of real-time applications and benefits that need fast processing and analysis of huge volumes of data, while also providing a more cost-effective and scalable alternative to traditional cloud computing.

12.4.2 Characteristics of Fog Computing

Fog computing can be differentiated from other computer paradigms by the following six features. An intelligent end-device or Internet of Things user does not necessarily have to utilize every feature when using fog computing benefits.

Low latency and location awareness Generally real-time applications require a low latency system. Fog computing achieves low latency by using the location awareness feature. The logical location of edge devices is known to the fog node, it is associated with the overall system context. The end device may contain some sensor or actuator. Due to the clear understanding of each device's location for communicating with those devices with a low latency path. It makes an optimal path between end devices and the fog node before transferring data. Location awareness is a great feature that helps the system to achieve the low latency feature.

Geographically distributed The services and applications that fog computing is designed to support require widely dispersed installations that may yet be uniquely identified by their geographic location, in sharp contrast to the more centralized cloud. For instance, fog computing will actively participate in supplying high-quality data streaming features to moving automobiles via proxy servers and access points strategically placed along roadways and railroads.

Heterogeneity It enables the collection and analysis of data in various forms obtained through a variety of network communication methods.

Interoperability and federation Being a distributed system the fog system allows federation. The collaboration of several suppliers is necessary for the seamless provision of some services (real-time streaming services are an excellent example). As a result, users must be united across fields and fog computing elements must be able to communicate with one another.

Real-time communication Nowadays all IoT applications become a time-sensitive applications. Time-sensitive applications need to respond in a moment. Applications of fog computing entail real-time interactions as opposed to batch processing methods. In fog computing data is processed at the prime component called fog nodes. The fog node is generally placed near the end devices. Whereas the cloud is not suitable for this type of application. Due to the implementation of hardware near the end devices, it processes faster than the cloud.

Agility and scalability Agility means how quickly and easily data move from the source and the destination. Whereas scalability defines the ability of a computing process to be used or produced in a range of capabilities. Fog computing is adaptable in nature, providing elastic computing, pooling of hardware resources, data-load management, and variations in the network state, to name a few of the supported flexible services, at the cluster level.

Some additional features of fog computing are

Wireless access In general, fog computing is utilized in wired systems, however, the extensive use of wireless sensors in the IoTs necessitates distributed analytics and computing. Fog computing is hence a perfect fit for wireless IoT access networks.

Mobility Various fog computing applications must be capable to interface with moving portable devices directly, thus they must support mobility approaches like the LISP, which separates host identification from location originality along with necessitating a distributed directory approach.

Fog computing provides a platform for deploying IoT applications in a more efficient and cost-effective way than traditional cloud computing.

There are several reasons why fog computing is suitable for IoT applications:

Low Latency IoT applications require low latency because they often involve real-time interactions between devices. Fog computing reduces latency by processing data closer to the location where data is created, at the network's edge.
Reduced Network Congestion With fog computing, data processing is distributed between the edge and the cloud, which reduces network congestion and lowers the amount of data that needs to be sent to the cloud for processing.
Improved Security IoT devices are often vulnerable to cyber attacks, and fog computing can provide an auxiliary layer of security by allowing data to be processed and stored locally, reducing the quantity of data that requires to be transferred over the network.
Cost-Effective With fog computing, data processing and storage can be distributed between the edge and the cloud, which reduces the demand for expensive data centers and infrastructure.
Increased Scalability Fog computing allows IoT applications to scale more easily by distributing data processing and storage across a larger number of devices.

Overall, fog computing is a suitable platform for deploying IoT applications due to its low latency, reduced network congestion, improved security, cost-effectiveness, and increased scalability.

12.4.3 Attributes of Fog Node

Fog nodes must support one or more of the following qualities to make it easier to create a fog computing capability that demonstrates the six key traits listed.

Autonomy The fog nodes can function autonomously at the single node or cluster of nodes level, making local decisions.
Hierarchical cluster The fog nodes follow hierarchical systems, wherein a continuum of interconnected service functions is provided by several service levels.
Heterogeneity The fog nodes are available in a range of form factors and can be set up in a combination of settings.
Manageability Complex systems that are capable of doing the majority of ordinary tasks automatically organize and orchestrate the fog nodes.
Programmability Programmability is a great feature of any system. Customization is an obvious feature for better flexibility. Fog nodes can

naturally be programmed at various levels by a variety of parties, including network administrators, subject matter experts, equipment suppliers, and end users.

12.4.4 Fog Computing Service Model

Fog computing provides a scalable and efficient way to process data and run applications in a decentralized way. Like cloud computing, fog computing also supports service-based models. The fog computing architecture can be implemented in several layers of network topology. The following service model types can be used. Figure 12.2 depicts the service model architecture of fog computing.

Software as a Service (*SaaS*) It is the topmost layer of the service-layered architecture. Software as a service delivers software services to the user. The user can only use the software in this layer. Platform and infrastructure layers are abstract at this layer. Utilizing the apps of the fog service provider operating on a cluster of confederate fog nodes maintained by the provider is the capacity offered to the customer of the fog service. This form of service indicates that the end-device or smart devices access the service as applications of fog node's using a thin client junction or an application program interface (API), akin to cloud computing's software as a service. With the probable exception of a small number of user-specific application configuration parameters, the end users have no grant of access to the network components, operating systems, servers, storage units, or even the other specific operation abilities of the underlying fog node. In this model, the fog computing infrastructure provides software applications to the users over the network. The applications are typically hosted and managed by the fog computing infrastructure, and the users can access them using various devices and platforms.

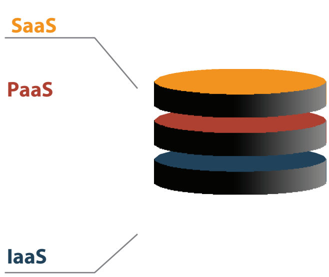

Figure 12.2 Fog node architectural service model.

Platform as a Service (*PaaS*) PaaS is the second layer of the service-layered architecture. The underlying layer of PaaS is infrastructure as a service layer. PaaS provide a platform base related environment to the users. In simple form, the user got key access to the virtual machine (VM). Users can customize the platform as per their needs. These layer users are also capable of deploying the SaaS layer in the platform layer. The features of SaaS can be achieved by installing the required software. The platform as a service capability offered to the fog service consumer is comparable to cloud computing and enables the deployment of customer-created or developed applications utilizing programming languages, services, libraries, and equipment endorsed by the fog service provider onto the platforms of federated fog nodes forming a cluster. The customer of the fog service does not manipulate or have authority over the network, servers, operating systems, or storage of the underlying fog platform(s), but does have control over the distributed applications and perhaps some arrangement options for the app hosting environment. In this model, the fog computing infrastructure provides a platform for the users to develop, deploy, and control their applications without worrying about the underlying infrastructure. The platform typically includes tools and libraries for application development, as well as middleware for managing application deployment and scaling.

Infrastructure as a Service (*IaaS*) SaaS is the bottom layer of the service-layered architecture. The IaaS users have no restrictions on any platform or service. Using the infrastructure of the fog nodes constituting a federated cluster, the fog service customer is given the capacity to require processing units, massive amounts of storage, networks, and other basic computing aids. The customer is capable to deploy and execute any application software, which can contain operating systems and specific application software, such as cloud computing IaaS services. The consumer has limited control over some networking components but does not manage or control the fog nodes cluster's underlying architecture. Rather, the client has full authority over the installed application softeners, storage, and operating systems. In this model, the fog computing infrastructure provides virtualized computing resources, such as processing power, storage, and networking, to the users. The users can then deploy their applications and services on these resources and manage them as per their requirements.

Data as a Service (*DaaS*) In this model, the fog computing infrastructure provides access to data from various sources, such as sensors, IoT devices, and cloud services. The users can then use this data for analysis, machine learning, or other purposes. The fog computing infrastructure typically provides data storage, processing, and analysis capabilities for efficient data management. The following advantages come from using DaaS (Data as a Service):

Agility The DaaS is based on a service-oriented architecture (SOA), and it offers a lot of flexibility when it comes to accessing vital data through cloud or fog services. Data may be accessed quickly thanks to the architecture that supports them. Changes to the data are also simple to implement when the data structure needs to be altered or when there are geographic requirements.
High-quality data The DaaS provider ensures that consumers have access to high-quality data by enforcing a strict method of data control and processing (collection, purification, accumulation, and enrichment).
Simple access The DaaS paradigm enables simple access to data via a variety of devices, including desktops PC, tablets, laptops, and smartphones, wherever and whenever they are needed.
Preventing provider lock-in with DaaS The DaaS paradigm makes it possible to move data quickly from one platform to another.
Function as a Service (FaaS) In this model, the fog computing infrastructure provides a server-less environment for running code snippets or functions. The users can upload their code to the fog computing infrastructure, and the infrastructure automatically handles the scaling and execution of the code as per the incoming requests.

12.4.5 Fog Computing Architecture

Instead of processing data in data centers as in the traditional cloud computing paradigm, this new extent of big data necessitates processing the data close to the sensors at the edge. The fog computing model allows for rapid control loops because the data is not transferred to the cloud it is generally processed at the edge.

Figure 12.3 depicts a four-layer fog computing design. The sensing network is located at Layer 4 of the network, which has a lot of sensory nodes. These sensors are low-cost, extremely dependable, and non-invasive, allowing them to be widely installed in different general infrastructures to track changes in their state from time to time.

Be aware that these sensors provide enormous sensing data streams that are spatially dispersed and must be handled as a unit [13].

The edge nodes transmit the unprocessed data to the third layer, which is made up of several low-power and highly-performance computing nodes or intelligent edge devices. Low-powered smart edge device are various devices, including desktops PC, tablets, laptops, and smartphones, wherever and whenever it covers a small community or area. The output generated by the edge device includes moving data quickly from one platform

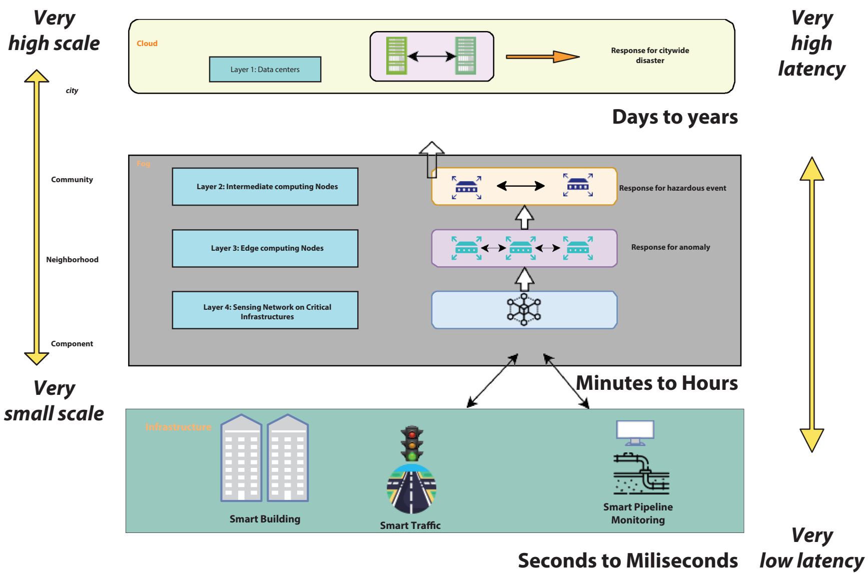

Figure 12.3 Fog computing architecture.

to other possible outcomes to a middle computing node at its subsequent higher layer, and the second is straightforward and speedy response controls to regional infrastructure to react to disjoined and minor hazards to the overseas infrastructure elements.

In order to identify potentially dangerous events, Layer 2 comprises a number of intermediary computational nodes that are individually associated with a pool of end-edge devices at Layer 3. When dangerous events are discovered, the infrastructure can be controlled quickly [33].

Layers 2 and 3 immediate feedback controls serve as localized "reflex" judgements to avert probable harm. For more intricate and extensive behaviour analysis and circumstances observation, the top layer receives reports on all data analysis findings from the bottom two layers.

In conclusion, the four-layered fog computing framework facilitates rapid reaction at neighborhood, community, and metropolis levels, offering high computing enactment and intelligence in the hereafter.

12.4.6 Data Flow and Control Flow in Fog Architecture

Solving the problems of huge data processing, the layered fog computing architecture enables high-performance computing and connectivity. It also delivers quick responses. It is demonstrated that data analysis workloads can be worked parallel across vast edge devices and computing nodes. Since only lightweight computations are carried out by each edge device or middle computer node, their extensive parallel utilization provides high-performance computing capacity for citywide data processing. However, the main benefit of such parallel computing systems is their simplicity in balancing throughput and loading all computer nodes and edge devices to prevent potential computing bottlenecks. The data and control flow enable architecture is depicted in Figure 12.4.

At Layer 4, an extended number of sensors are used to generate an enormous amount of data. The hierarchically dispersed edge devices and intermediary computer nodes at Layers 3 and 2 only upload high-level data depiction while carrying out interrelated computing operations, which can significantly decrease the amount of data transported to the cloud [34]. In order to exchange data with their neighbors, sensor nodes or computing nodes can communicate with other nodes in the same layer. Depending on the exact uses of a smart city, conveyed information between neighboring levels and within a layer may be wireless or wired. Wireless networks have numerous crucial uses in isolated and portable observation, and wired networks can offer stable and dependable connections between various devices. Additionally, when abnormalities and dangerous occurrences are discovered, the computing machines in each layer can send command signals (dashed line) in a convenient fashion to the depicted infrastructures to guarantee their protection (see Figure 12.4). A large number of computers are linked together at the top layer to create clouds, which are used for data storage and city-scale computation [35, 36].

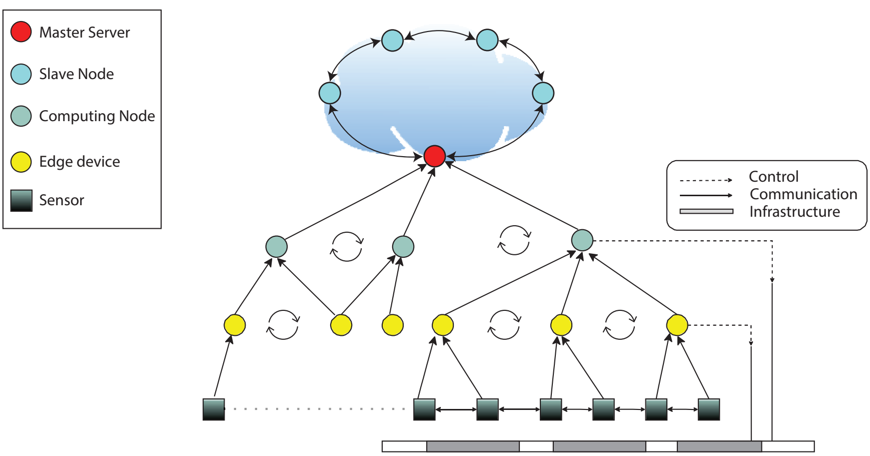

Figure 12.4 Data and control flow for layered fog computing architecture.

12.4.7 Fog Deployment Models

The following deployment strategies are also supported because fog computing is recognized and described as an addition to the conventional cloud-based computing model:

Public fog Users can avail of the services without restrictions. The fog node is made available for open use by everyone. A business, academic, or governmental body, or an assortment of them, may own, administer, and run it. It is present on the grounds of the fog service supplier. The public fog node is nearly similar to the public cloud.

Private fog This fog node is designated only for use by one company with numerous consumers (e.g., business units.) It may establish on or off premises and be owned. Private fog nodes are controlled and handled by the self-organization, third-party organizations, or some collaboration of both (self and third-party). It provides the same functionality and services as the private cloud.

Community fog This fog node is designated for service only by a particular group of clients from companies with similar concerns (such as tasks, safety needs, guidelines, and obedience issues). It may be owned, controlled, and run by one or more community associations, a third party, or a combination. It may also be located on or off-site. Community fog is sometimes identical to community cloud.

Hybrid fog This fog node is made up of two or more different fog nodes (public or private, community), each of which is distinct and continues to exist as a separate entity, but which are connected by standardized or proprietary technology that allows for the portability of data and applications.

12.5 Big Data

Terabytes of data are gathered by businesses, organizations, and research facilities from a variety of origins, including emails, social media, survey answers from various customers, call logs, click-stream data from the Internet, web server logs, and sensors. Massive amounts of continuously flowing unstructured, semi-structured, or organized data are referred to as "big data" in businesses. The idea of big data has been around for a while, and most organizations today are aware that they may use analytics to their data to generate understandings that can be put to use. Advanced analytics for big data includes complicated applications with elements like statistical models, what if analyses, and predictive analytics that are supported by high-performance analytics platforms. Business analytics is used to provide basic insights into business operations and performance. Large volumes of data are analyzed using big data analytics to find undiscovered patterns, correlations, and other insights. Both batch and streamlined methods can be used to process big data. This indicates that for various applications, data analysis and result generation are based on the store-and-process approach. Many time-sensitive applications, like stock market data processing, generate data continually and anticipate the processed result in real time [37].

12.5.1 What is Big Data?

A data set is a collection of data values that are related in meaningful and specific ways. Big Data is an umbrella term for structured, unstructured and semi-structured data sets. Big data is a massive data set which provides valuable information to organizations in order to run businesses efficiently.

Big data can be described as a set of large, complex structured and unstructured data items that are readily available, but too voluminous or diverse to be stored or manipulated with conventional database systems. Thus, the processing of big data requires advanced high-performance computing (HPC) capabilities and technologies [38].

Big data refers to enormous amounts of information that may or may not be organized in a specific way. Typically, in big data, the individuals involved are less concerned with identifying each piece of information as unique and individually important, but instead, seek to identify patterns using these data.

The application of big data has been one of the most significant technological innovations in recent decades. It is expected to be a key factor in the coming expansion of artificial intelligence (AI) and machine learning (ML) techniques, which are also anticipated to guide to further refinements in many other scopes such as health care and natural or social science research [39].

Big data includes a broad range of applications across different industries and domains, including:

Healthcare: Big data is used in healthcare to enhance patient results, decrease expenses, and optimize healthcare delivery. It can help identify high-risk patients, predict disease outbreaks, and personalize treatments based on individual patient data.

Finance: Big data is used in finance to analyze market trends, manage risks, and detect fraudulent activities. It can help financial institutions make informed investment decisions and improve customer experience.
Retail: Big data is used in retail to analyze consumer behavior, predict sales trends, and optimize supply chain management. It can help retailers personalize marketing campaigns, improve inventory management, and increase customer satisfaction.
Manufacturing: Big data is used in manufacturing to optimize production processes, reduce costs, and improve quality control. It can help manufacturers detect defects in real-time, predict maintenance needs, and optimize production schedules.
Transportation: Big data is used in transportation to optimize routes, reduce congestion, and improve safety. It can help transportation companies predict traffic patterns, optimize logistics, and improve fuel efficiency.
Education: Big data is used in education to personalize learning experiences, improve student outcomes, and optimize educational programs. It can help educators identify at-risk students, predict student performance, and improve curriculum design.
Energy and Utilities: Big data is used in energy and utilities to optimize energy consumption, reduce costs, and improve sustainability. It can help utility companies monitor energy usage, predict demand, and optimize energy production.

These are just a few examples of how big data is being used to solve complex problems and drive innovation across different industries. The possible applications of big data are virtually limitless, and as the amount of data persists to increase, its impact will only become more significant.

12.5.2 Source of Big Data

As per McKinsey Global Institute, the potential source of big data includes [40]:

Public sector: The governments and other non-governmental organizations promote digital systems. They move to digital platforms for creating transparency, data accessibility, better data discovery, and performance improvements making decisions for product services-related activity. It is one of the main sources of big data in recent days.
Manufacturing: The manufacturing industry generates and analytics big data for efficient demand prediction, supply chain management, sales support, production operations etc.

Personal location data: Personal location data refers to the data generated by each person's geographical location using GPS. That data is used by Geo tagged advertising, for emergency purposes, smart urban planning etc.
Retail: Retail industry generates a huge amount of data every single second. By using this data the organization analyze store behaviour, labor management, product placement strategies etc.
Healthcare: Healthcare sector is one of the finest sources of big data generation. Nowadays data is everywhere in the medical field like patient profiles, clinical decision support data, and analysis of diseases by pattern recognition, performance-based price options etc.

12.5.3 Characteristic of Big Data

Big data is generally characterized by the following four attributes, commonly referred to as the "4 Vs of Big Data":

Volume: Big data refers to very large data sets, typically in the terabytes or petabytes range. This vast amount of data often requires specialized hardware and software tools to store, manage, and process.
Velocity: Big data is often generated and collected at high speed and in real-time, such as data from sensors, social media feeds, or financial transactions. This requires fast processing and analysis to provide timely insights.
Variety: Big data comes in many forms, including structured, semi-structured, and unstructured data. This data may be generated from a variety of sources, such as text, images, audio, or video. This variety of data requires flexible and adaptable tools to process and analyze it effectively.
Veracity: Big data is often characterized by its quality and accuracy, which can vary depending on the source and collection methods. Big data may contain errors, inconsistencies, or biases that can affect the analysis and interpretation of the data. It is essential to validate and verify the accuracy of the data before making any decisions based on it.

In addition to these four characteristics, big data may also exhibit the following characteristics:

Variability: Big data may exhibit high variability or fluctuation over time, making it challenging to analyze and interpret.
Complexity: Big data may be complex, with many interrelated variables and relationships, making it challenging to model and analyze.
Privacy and Security: Big data may contain sensitive or confidential information, which requires appropriate security measures to protect against

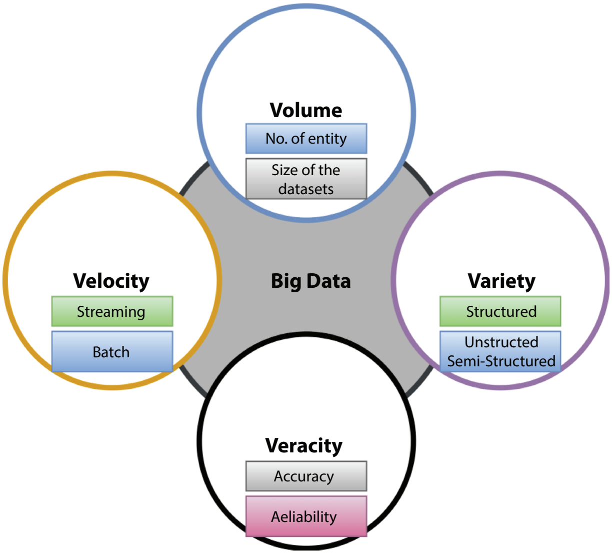

Figure 12.5 Characteristic of big data.

unauthorized access or disclosure. Figure 12.5 depicts the characteristics of big data.

12.6 Big Data Analytics Using Fog Computing

"Big data" is the phrase utilized to define the vast quantities of unstructured, semistructured, or organized data that are constantly flowing through and around companies. Enormous amounts of data are gathered by businesses, organizations, and research institutes from a wide range of sources, such as social networking sites, consumer emails and questionnaire replies, telephone records, Internet user behaviour data, IoT devices, and sensors. The idea of big data has been around for a while; currently, the majority of enterprises are aware that they may use analytics to big data to obtain useful ideas. Large volumes of data are processed using big data analytics to find undiscovered patterns, relationships, and other truths. Both batch and streamlined methods may be used to analyze big data. Many applications use store-and-process schemes to analyze and produce results from data. Some real-time applications regenerate a mass of data continuously and anticipate the possible result in a real-time manner like health data, stock market data etc.

Big data analytics has some advantages. Firstly, quick and efficient decision-making is one of the excellent advantages of big data analytics. Secondly, discovering new opportunities to improve business leads to higher profit. Finally, new services and products can be introduced based on data analytics results.

Typical big data analytics consists of several components like a platform, storage scheme, data management scheme, core analytics and its operations, add-on tools, and presentation. Hadoop is the most popular platform that acts as an underlying foundation of core analytics. It acts like a big data processing engine [41].

Big data analytics is a complicated analytical job on massive data. Figure 12.6 highlights a typical outflow of big data processing [42]. Data collection from numerous sources is done in the first step. The second step is responsible for data cleaning. This step requires a large processing time additionally this step reduces the data size by discarding impurities, missing values and redundant data. It also detects and removes inconsistencies from data. As a result, the next step requires less effort and time for data analytics [43]. Data is stored in a storage database before processing. Generally, raw data is unstructured data, before processing it is mandatory

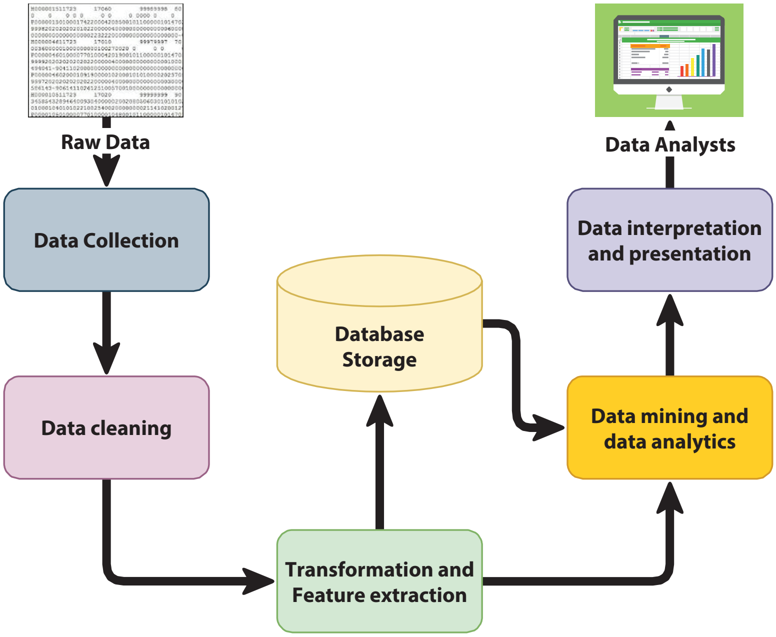

Figure 12.6 A typical big data analytic flow diagram.

to transform the data into semi-structured or structured data. Data processing is one of the most crucial jobs of data mining. How to deal with the data, and what kind of information need to be extracted from the data set, are determined by a rule-based model [44]. To discover knowledge from the data, the job of data analysts comes into account. They run several queries, machine learning algorithms, and other techniques to visualize the data as per requirements and extract information from them.

Fog engine support complete solution data analytics near the data sources device and, exchange information among other devices and the cloud. Figure 12.7 shows data transmission among the fog engines and the cloud. Generally, the data generated from the different sources are gathered, cleaned and processed at the fog-engine on-premise of data source devices. But when data become significantly large and the fog engine is unable to process the data, it transfers the data to the cloud.

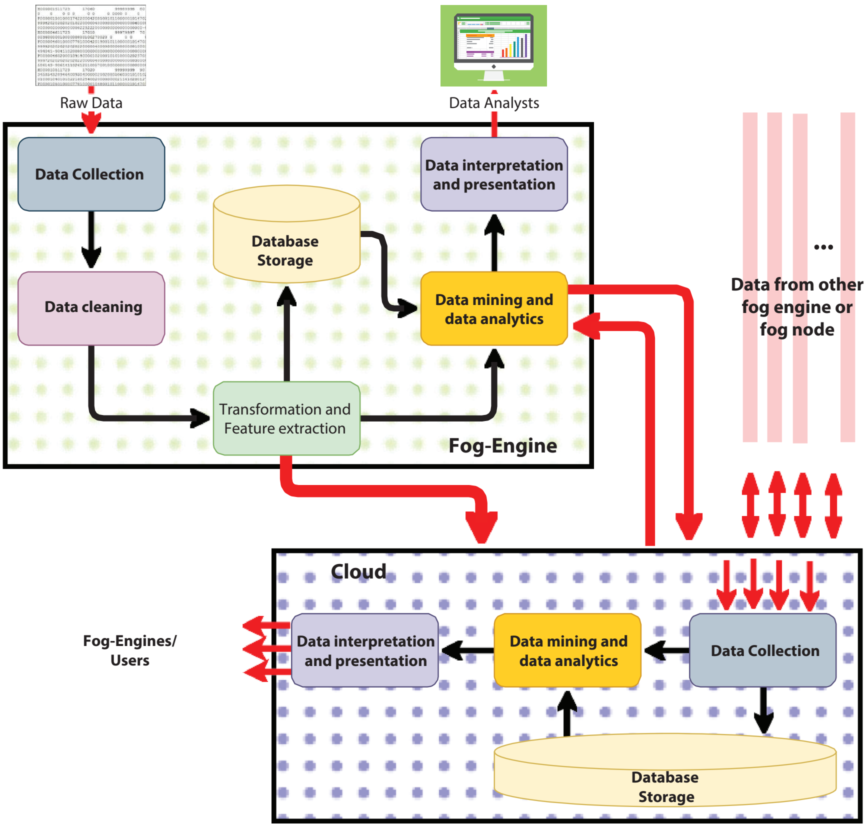

Figure 12.7 Big data analytic using fog-engine.

In Figure 12.7, the data collection step of the cloud collects and integrates the data from all the fog-engine. Data can be stored in the database for future use. After mining the data knowledge can be exchanged between the fog engine and the cloud. Real-time applications like continuous monitoring require real-time processing. So, that kind of analytical model can be deployed in the fog engine. Although the fog-engine limited computing capabilities and storage capacity than cloud. Fog-engine may consist of different types of hardware such as FPGA, and graphics processing units, the multi-core processor for real-time processing. Cloud is not capable of real-time processing due to several dependencies like location, bandwidth, latency etc. Finally, process information sent back to fog engines or users from the cloud through proper channels.

12.7 Conclusion

In conclusion, IoT big data analytics is a critical component of modern business operations, providing valuable insights into customer behaviour, product performance, and operational efficiency. However, processing and analyzing large amounts of IoT data can be a challenging task that requires powerful computing resources and high-speed connectivity.

Fog computing has appeared as an optimal solution to address the challenges of IoT big data analytics, by bringing computing resources nearer to the edge of the network, where the IoT data is generated. By processing data closer to the origin, fog computing reduces latency and network congestion, enabling faster and more efficient data analysis.

Moreover, fog computing can also help decrease the cost of IoT big data analytics by minimizing the need for data transmission and storage. By processing data at the edge, fog computing can filter out irrelevant data and only send relevant insights back to the cloud, reducing the amount of data that needs to be stored and transmitted.

Overall, IoT big data analytics using fog computing has the potential to revolutionize business operations, enabling real-time insights that can drive innovation, improve efficiency, and enhance customer experience. As more organizations adopt IoT technologies and generate massive amounts of data, fog computing is poised to become an increasingly essential tool for unlocking the full possibility of IoT big data analytics.

References

1. Dastjerdi, A.V. and Buyya, R., Fog computing: Helping the internet of things realize its potential. *Computer*, 49, 8, 112–116, 2016.
2. Dastjerdi, A.V., Gupta, H., Calheiros, R.N., Ghosh, S.K., Buyya, R., Fog computing: Principles, architectures, and applications, in: *Internet of things*, pp. 61–75, Elsevier, USA, 2016.
3. Banerjee, A., Dutta, B., Mandal, T., Chakraborty, R., Mondal, R., Blockchain in IoT and beyond: Case studies on interoperability and privacy, in: *Blockchain based Internet of Things*, pp. 113–138, Springer, Singapore, 2022.
4. Atlam, H.F., Walters, R.J., Wills, G.B., Fog computing and the internet of things: A review. *Big Data Cogn. Comput.*, 2, 2, 10, 2018.
5. Saharan, K. and Kumar, A., Fog in comparison to cloud: A survey. *Int. J. Comput. Appl.*, 122, 3, 10–12, 2015.
6. Alrawais, A., Alhothaily, A., Hu, C., Cheng, X., Fog computing for the internet of things: Security and privacy issues. *IEEE Internet Comput.*, 21, 2, 34–42, 2017.
7. Mandal, R., Mondal, M.K., Banerjee, S., Chakraborty, C., Biswas, U., A survey and critical analysis on energy generation from datacenter, in: *Data Deduplication Approaches*, pp. 203–230, Elsevier, USA, 2021.
8. Buyya, R., Ramamohanarao, K., Leckie, C., Calheiros, R.N., Dastjerdi, A.V., Versteeg, S., Big data analytics-enhanced cloud computing: Challenges, architectural elements, and future directions, in: *2015 IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 75–84, IEEE, 2015.
9. Trovati, M., Hill, R., Anjum, A., Zhu, S.Y., Liu, L., *Big-Data Analytics and Cloud Computing*, Springer, Derby, UK, 2015.
10. Gupta, R., Gupta, H., Mohania, M., Cloud computing and big data analytics: what is new from databases perspective?, in: *Big Data Analytics: First International Conference, BDA 2012, New Delhi, India, December 24-26, 2012. Proceedings 1*, pp. 42–61, Springer, Berlin, Heidelberg, 2012.
11. Yetis, Y., Sara, R.G., Erol, B.A., Kaplan, H., Akuzum, A., Jamshidi, M., Application of big data analytics via cloud computing, in: *2016 World Automation Congress (WAC)*, pp. 1–5, IEEE, 2016.
12. Singh, S. and Yassine, A., IoT big data analytics with fog computing for household energy management in smart grids, in: *Smart Grid and Internet of Things: Second EAI International Conference, SGIoT 2018, Niagara Falls, ON, Canada, July 11, 2018, Proceedings 2*, pp. 13–22, Springer, Rio Grande, PR, USA, 2019.
13. Tang, B., Chen, Z., Hefferman, G., Pei, S., Wei, T., He, H., Yang, Q., Incorporating intelligence in fog computing for big data analysis in smart cities. *IEEE Trans. Ind. Inf.*, 13, 5, 2140–2150, 2017.

14. Tang, B., Chen, Z., Hefferman, G., Wei, T., He, H., Yang, Q., A hierarchical distributed fog computing architecture for big data analysis in smart cities, in: *Proceedings of the ASE BigData & SocialInformatics 2015*, pp. 1–6, 2015.
15. Barik, R.K., Dubey, H., Samaddar, A.B., Gupta, R.D., Ray, P.K., Foggis: Fog computing for geospatial big data analytics, in: *2016 IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics Engineering (UPCON)*, pp. 613–618, IEEE, 2016.
16. Dubey, H., Yang, J., Constant, N., Amiri, A.M., Yang, Q., Makodiya, K., Fog data: Enhancing telehealth big data through fog computing, in: *Proceedings of the ASE BigData & SocialInformatics 2015*, pp. 1–6, 2015.
17. Mouradian, C., Naboulsi, D., Yangui, S., Glitho, R.H., Morrow, M.J., Polakos, P.A., A comprehensive survey on fog computing: State-of-the-art and research challenges. *IEEE Commun. Surv. Tutor.*, 20, 1, 416–464, 2017.
18. Liu, Y., Fieldsend, J.E., Min, G., A framework of fog computing: Architecture, challenges, and optimization. *IEEE Access*, 5, 25445–25454, 2017.
19. Puliafito, C., Mingozzi, E., Longo, F., Puliafito, A., Rana, O., Fog computing for the internet of things: A survey. *ACM Trans. Internet Technol. (TOIT)*, 19, 2, 1–41, 2019.
20. Vimal, S., Suresh, A., Subbulakshmi, P., Pradeepa, S., Kaliappan, M., Edge computing-based intrusion detection system for smart cities development using IoT in urban areas, in: *Internet of Things in Smart Technologies for Sustainable Urban Development*, pp. 219–237, 2020.
21. Ahmed, A., Arkian, H., Battulga, D., Fahs, A.J., Farhadi, M., Giouroukis, D., Gougeon, A., Gutierrez, F.O., Pierre, G., Souza Jr., P.R. *et al.*, Fog computing applications: Taxonomy and requirements, Cornell University, preprint arXiv:, 1, 1907, 2019, arXiv preprint arXiv:1907.11621.
22. Teoh, Y.K., Gill, S.S., Parlikad, A.K., Iot and fog computing based predictive maintenance model for effective asset management in industry 4.0 using machine learning. *IEEE Internet Things J.*, 10, 2087–2094, 2021.
23. Mandal, R., Mondal, M.K., Banerjee, S., Srivastava, G., Alnumay, W., Ghosh, U., Biswas, U., MECpVmS: An SLA aware energy-efficient virtual machine selection policy for green cloud computing. *Cluster Comput.*, 26, 651–665, 1–15, 2022.
24. Mandal, R., Mondal, M.K., Banerjee, S., Chatterjee, P., Mansoor, W., Biswas, U., PbV mSp: A priority-based Vm selection policy for Vm consolidation in green cloud computing, in: *2022 5th International Conference on Signal Processing and Information Security (ICSPIS)*, pp. 32–37, IEEE, 2022.
25. Luan, T.H., Gao, L., Li, Z., Xiang, Y., Wei, G., Sun, L., Fog computing: Focusing on mobile users at the edge, Cornell University, 1, 2015, arXiv preprint arXiv:1502.01815.
26. Mondal, M.K. and Bandyopadhyay, M., A comparative study between cloud computing and fog computing. *Brainwave: Multidiscip. J.*, 2, 1, 36–42, 2021.

27. Mondal, M.K., Mandal, R., Banerjee, S., Biswas, U., Lin, J.C.-W., Alfarraj, O., Tolba, A., Design and development of a fog-assisted elephant corridor over a railway track. *Sustainability*, 15, 7, 5944, 2023.
28. Gupta, H. and Vahid Dastjerdi, A., Ifogsim: A toolkit for modeling and simulation of resource management techniques in the internet of things, edge and fog computing environments. *Softw.: Pract. Exper.*, 47, 9, 1275–1296, 2017.
29. Mondal, M.K., Mandal, R., Banerjee, S., Sanyal, M., Ghosh, U., Biswas, U., Fog assisted tiger alarming framework for saving endangered wild life, in: *SoutheastCon 2023*, IEEE, pp. 798–803, 2023.
30. Mondal, M.K., Mandal, R., Banerjee, S., Biswas, U., Chatterjee, P., Alnumay, W., A cps based social distancing measuring model using edge and fog computing. *Comput. Commun.*, 194, 378–386, 2022.
31. Chang, Y.-C.P., Chen, S., Wang, T.-J., Lee, Y., Fog computing node system software architecture and potential applications for NB-IoT industry, in: *2016 International Computer Symposium (ICS)*, pp. 727–730, IEEE, 2016.
32. Mukherjee, M., Shu, L., Wang, D., Survey of fog computing: Fundamental, network applications, and research challenges. *IEEE Commun. Surv. Tutorials*, 20, 3, 1826–1857, 2018.
33. Bonomi, F., Milito, R., Natarajan, P., Zhu, J., Fog computing: A platform for internet of things and analytics, in: *Big Data and Internet of Things: A Roadmap for Smart Environments*, pp. 169–186, 2014.
34. Yi, S., Li, C., Li, Q., A survey of fog computing: concepts, applications and issues, in: *Proceedings of the 2015 Workshop on Mobile Big Data*, pp. 37–42, 2015.
35. Mandal, R., Mondal, M.K., Banerjee, S., Chatterjee, P., Mansoor, W., Biswas, U., Design and implementation of an SLA and energy-aware VM placement policy in green cloud computing, in: *2022 IEEE Globecom Workshops (GC Wkshps)*, IEEE, pp. 777–782, 2022.
36. Chakraborty, R., Ghosh, A., Mandal, J.K., Balamurugan, S., *Convergence of Deep Learning in Cyber-IoT Systems and Security*, John Wiley & Sons, USA, 2022.
37. Davenport, T.H., Barth, P., Bean, R. *et al.*, How'big data'is different. *MIT Sloan Manage. Rev.*, 54, 1, 22–24, 2012.
38. Daas, P.J., Puts, M.J., Buelens, B., van den Hurk, P.A., Big data as a source for official statistics. *J. Off. Stat.*, 31, 2, 249–262, 2015.
39. Fan, J. and Liu, H., Challenges of big data analysis. *Natl. Sci. Rev.*, 1, 2, 293–314, 2014.
40. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A. *et al.*, *Big data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, USA, 2011.
41. White, T., *Hadoop: The Definitive Guide*, M. Loukides (Ed.), O' Reilly Media, Inc., USA, 2009.
42. Mehdipour, F., Noori, H., Javadi, B., Chapter two energy-efficient big data analytics in datacenters, in: *Advances in Computers Energy Efficiency in Data*

Centers and Clouds, vol. 100, A.R. Hurson and H. Sarbazi-Azad (Eds.), pp. 59–101, Elsevier, USA, 2016.
43. Rahm, E. and Do, H.H., Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.*, 23, 4, 3–13, 2000.
44. Analytics, B.D., Extract, transform, and load big data with apache hadoop, Intel White paper, Big Data, 1, 1–8, 2013.