Research and Application of Key Technologies for Fog Computing Based on Big Data

Research and Application of Key Technologies for Fog Computing Based on Big Data

Baoling Qin

College of Electronic Information Engineering, Foshan University Guangdong Foshan 52800, China

QingBL0083@126.com

Abstract. In big data computing services, fog computing can be said to be an enhanced version of cloud computing. In recent years, the application of intelligent Internet of things has spread all over the industry and has expanded rapidly, and its integration with big data technology particularly stands out. Cloud computing seems to have more serious problems in big data storage, data calculation and data processing, including the lack of storage space, redundant data, and network interruptions and delays. In order to solve these problems, we use the edge technology functions of Intelligent Internet of things (IoT) according to the models and characteristics of fog computing, putting data storage, data calculation, and data processing into edge devices. In this way, we can solve problems that occurs in cloud computing when processing big data and effectively improve the storage and calculation of big data.

1. Introduction

At present, the development and application of intelligent Internet of things is in full swing. A year after cloud computing was introduced in 2006, the IT industry started to research and implement cloud computing technology. In terms of current development and practice, cloud computing is too network centralized. In the intelligent Internet of things system, data from each terminal device generates a huge amount of data and these data differs from their sources, types and constructions, each with its own characteristics. Also, data coherence and dispersion coexist. During the real-time transport, cloud computing often causes network delays or even interruptions in big data cleaning, transmission, processing, and applications. To sum up, in the big data environment, cloud computing has the following shortcomings: (1) the data is too centralized; (2) the technique is too complex; (3) the data center is overloaded; (4) the protection of users' privacy lacks; (5) the operation is difficult. Therefore, in the big data environment, the disadvantages of cloud computing are very prominent and have not been widely popularized so far. On the contrary, these problems have been effectively solved, with the concept of fog computing proposed and quickly implemented. From the current research and practice, it is relatively easy to implement such as big data storage and processing because existing devices and technology have been made full use of. It provides a great foundation and motivation for IoT application technologies and it will make a difference in human social development. As a result, research of the key technology and its application is of great value and great significance in the era of integration of intelligent IoT and big data.

2. The concept and characteristics of fog computing

With the rapid development of new information technologies and the continuous expansion of the application fields, the Internet of Things is heading ahead from simply objects connections to the intelligent Internet of Things. The intelligent Internet of things provides distinctive services and applications to various industries, changing people's life styles and promoting progress of the society. Such applications are intelligent mobile, intelligent home, intelligent factory, environmental monitoring, energy development, drone control, and driverless cars. According to Gartner's forecast report, the number of global intelligent connections will reach 100 billion by 2020, and the market scale will be over US$300 billion. By then, the intelligent Internet of Things will be already mature [1].

2.1. Concept

In 2011, Cisco Corporation Bonomi first proposed the concept of fog computing. In 2012, he pioneered the concept of fog computing [2]. It is a fundamental device in contributed computing of IoT and extends computing capability and data analysis applications to the edge of the network. Thus customers are able to analyze and manage data locally so that they can get immediate insights through links. It is an extension and enhancement of cloud computing and it can also be considered as an edge computing service. In the big data processing and management system, cloud computing's shortcomings have become increasingly prominent while fog computing nearly solve all of these problems. Intelligent IoT terminals can easily access the local cloud through networking [3]. Intelligent IoT terminal is an enhanced version of the cloud computing model. The core is to decentralize the network and to adopt a distributed computing model, which involves many terminal nodes and makes full use of each node's capability. The great number contributes to the accomplishment of service though each nod is not strong enough, and this provides the network system with a distinctive and diverse application service concept. In the application service model system, access to the local cloud is based on any IoT devices. As for the management and control of the total data, its permissions are all controlled by users, in which user's privacy obtains protections.

2.2. Features and Benefits

The fog computing can be considered as a distributed computing service system. Its greatest operating characteristic is the edge equipment of the network. It is not a substitute. Instead, fog computing inherits could computing's advantages, with high-performance prices and enhanced functions. Meanwhile, it has the advantages of a distributed computing system. They are fault tolerance, heterogeneous, security, programming model, etc. [4]. In the current big data computing service system, a lot of data must be processed near the edge, such as health care systems, traffic commanding systems, intelligent parking systems, etc. The biggest feature is the maximization of edge terminal device computing functions and localization of data processing, perfectly solving the problem of rapid response of delay-sensitive applications in computing systems. In contrast, cloud computing is rather weak in the part. In the automated intelligent application system, fog computing has obvious advantages for data storage and output, decentralization and the distributed computing model can effectively make up for the deficiency of cloud computing. The advantages and advantages of fog computing are as follows [5]:

1) The large number of edge nodes

Fog computing consists of many edge devices that are weak but spread across the network. Consequently, the outstanding feature is that there are many nodes, and each node device can play its advantages. The computing construction is a distributed system, with the advantages of being close to the data source and supporting IoT's edge nodes.

2) Localization of data processing

In the construction of the Internet of Things, edge devices can store data and there are large amounts of data and related programs stored in the edge devices. However, in cloud computing, all data is nearly stored in the center of the cloud. While in big data applications, data storage and processing are more dependent on the local device than on the central server. Therefore, in the big data

environment, fog computing is in line with the "decentralized" characteristics of a new generation of distributed IoT systems, especially where data processing is easier to implement.

3) Little possibility in data transmission

The device closest to the data source is the edge device for fog computing. As a result, the data transmission is very timely and the delay rate is nearly zero, which largely overcomes the bottleneck problem of cloud computing data transmission and accelerates the network information patency. In short, fog computing is suitable for intelligence Internet of Things applications.

4) A broad range of distribution

In the intelligent internet of things, especially the mobile internet, edge devices are distributed in each terminal of the entire network where nodes are widely distributed, suitable for Intelligent IoT applications.

5) Strong mobility

In intelligent communication networks, edge computing is of great mobility. Information is directly communicated between different mobile devices where signals do not need to circle. Thus, high mobility is supported, and information flow is improved.

In the field of Internet of Things and big data fusion applications, fog computing's advantages are even more outstanding.

3. Key technologies for fog computing based on big data

In the past few years, smart Internet of Things has been widely deployed and applied in various fields. Sensor terminal devices of the Internet of Things generate a large amount of big data. To store and process these data, we adopt the model of "decentralization" and "everything connected" according to the features of fog computing. With massive devices, we make full use of the edge device storage and calculation functions, take advantages from the characteristics of little delay in fog computing, and improve the data's ability of analyzing, layering and refining so that the key technology of fog nodes can be widely used to solve related problems.

3.1. Fog Computing's Service Model and Functions

Fog computing is the enhancement of cloud computing, and it is also the edge device closest to the data sources. Fog computing's service model and function play a significant role in technology research.

1) Fog computing's Service Model

Fog computing consists of three layers: the cloud, the fog, and the terminal [6]. Fog computing's system is made up of multiple independent fog nodes. Moreover, fog computing is the terminal device itself or the device connecting the network and the terminal device, with task storing and processing the Internet of Things to generate the data stream [7]. Such examples are routers, set-top boxes, proxy servers, and base station (all of them are network middleware). These components provide different computing, storage, and network functions, and they also support the execution of service applications. The fog computing service model diagram is shown in Figure 1.

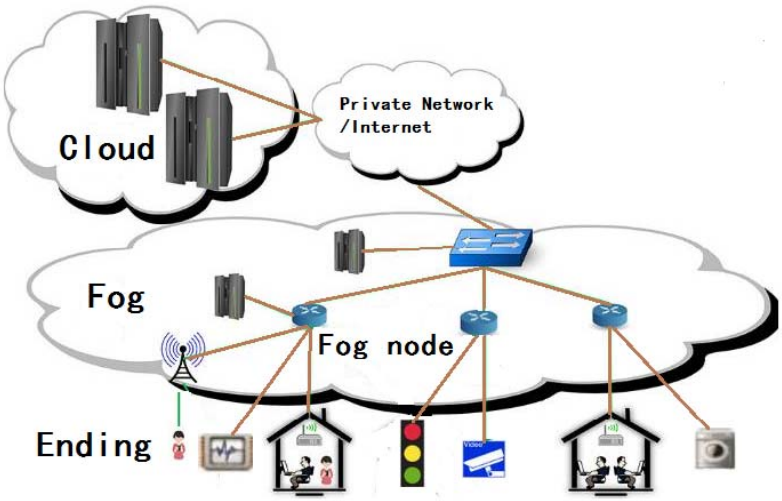

Figure 1. Fog Computing Service Model

2) Fog Computing Function

(1) Edge computing function

Fog computing is the calculation on the device side, which is edge computing and it can help cloud computing meet some of the big data calculation needs. Due to the complexity and diversity of big data types, the timeliness of data processing is more demanding and the amount of data exceeds the ZB level. Cloud computing process big data in a centralized way while fog computing process in a edge-distributed way. Under this comparison, fog computing's advantages stand out. For one thing, it can improve data transmission performance, ensuring real-time data processing. Meanwhile, the pressure on cloud computing center services are reduced. For instance, the sensor needs to collect data 100 times per second, and it can't transfer data to the cloud in real time before processing it. However, it is more appropriate that edge devices perform necessary data calculation, before sending the processing results to the cloud, or that the module converts directly the data analysis to the devices' action so that real-time performance can be improved.

(2) Edge storage function

Edge storage node is a small device for localization services. Such devices like small servers integrate dozens of technology including network access technology, big data low latency channel transmission technology, computing technology, and local storage technology. From the edge of the network to ultimate users, these devices connect the data center of the cloud platform and provide users with data storage services according to the user's requirements in data update. This kind of mobile data storage service is widely used, such as buses, subways, shopping malls and hospitals.

(3) Data preprocessing function

Through the preprocessing algorithm of the fiber optic gyro node, data can be preprocessed,which implements real-time data processing as well as real-time push results, improves the real-time data's efficiency, lightens the pressure of data push in the cloud, makes full use of all resources, and avoids wasting resources.

3.2. Edge Storage Technology

In this model, the edge storage technology is the effective technology to improve data processing efficiency, and its core is the storage buffer replacement technology.

1) Edge Storage Concept

The edge storage technology utilizes the caching function of the localization device, and storage data users' terminal may access the edge fog nodes through the system scheduling algorithm. When accessing the required data, users can quickly obtain data through edge fog nodes, which not only can improve the users' response speed, but also can quickly obtain data through edge fog nodes. Moreover,

edge storage can also reduce the internal bandwidth occupied by data transmission in cloud computing and can reduce cloud computing's workload on the central server at the same time. Because of the limited storage capacity of the cache, it is necessary to use alternative algorithmic techniques to replace data that may have been recently accessed and that are unlikely to be accessed immediately to free up storage space.

2) Edge storage buffer replacement technology

In computer systems, replacement algorithms include FIFO, LFU, and LRU, ant colony, bee colony and other algorithms. The first three algorithms belong to the traditional replacement algorithm, which is simple and easy to implement. Due to the complexity of large data processing and its application requirements, the traditional replacement algorithm barely meet the requirements of big data access. Therefore, a more optimized and more complex algorithm has emerged. For example, ant colony algorithm and bee colony algorithm have a broad application in the intelligent cluster scheduling, multi-variable function computing optimization and distributed combination of calculation optimization. In its edge data scheduling, the algorithm utilizes ant colony's characteristics to determine the optimal transmission path through the amount of information on the network path nodes. In fact, the dynamic optimization arrangement technique is realized by judging the scheduling frequency of certain data.

In these intelligent compounding algorithms, the biggest feature is that it supports distributed concurrent execution of calculations. Each object is an independent data computing task and interacts mainly with the amount of information loaded on the data. In the distributed fog node computing space, each computing task can independently solve a problem. Through the interaction of information quantity, complex combinatorial optimization calculation can be realized, which improves the reliability and global computing ability of the algorithm.

3) Edge storage buffer replacement technology optimization

In blurry edge storage cache replacement technology, ant colony algorithm achieves the optimal selection and replacement of data objects, and judges and obtains results through the amount of information loaded by the data object and the storage space occupied. If the total number of traversals is fixed, the greater the amount of information each object releases within a specified time interval, the greater the importance the selected object is of; the smaller the storage space the object occupies, the higher the expected value the selected object is of. Therefore, the best solution can be chosen by calculating the amount of information and storage space and comparing the results. Information volume replacement (update) model formula:

$$G_i(t + \Delta t) = L \times \Delta G_i \tag{1}$$

Among them,

$$\Delta G_i = \begin{cases} (1 - L) \times G_i(t) + \Delta G_i \\ - \dfrac{d_i}{V} \end{cases} \tag{2}$$

Here, a) $G_i(t)$ represents the amount of information released by the ant on the data object at time t; b) ΔG_i represents the increment of information released by the ant on the data object during storage conversion; c) Δt represents the time interval of information conversion (update); d) L represents the degree of loss of resource object information, $L \in (0, 1)$; e) d_i represents the amount of data objects occupying storage space at a certain time; f) V represents the total amount of data objects occupying storage space.

In the Intelligent Internet of Things, as the location of the mobile terminal is uncertain, the time of the object and the request is random without guaranteed synchronization. Therefore, the time of data replacement occurs randomly. When the buffer storage space is fully utilized, the replacement technology increases the target of updating the information to be loaded periodically. The reason is

that the amount of access of the object is proportional to the amount of load information. That is to say, when the amount of access increases, the amount of load information increases accordingly. Regular updates, can effectively improve the replacement efficiency, reduce storage space load, and shorten network latency.

3.3. Edge Computing Technology

Edge computing is of great significance in its system. Although 2 computing concepts are different, they still have a close relationship. Edge computing promotes the development of the fog computing concept. Its computational model is in each node device in the network instead of the cloud center, enabling its processing capability closer to data source.

1) Edge Computing Concept

Edge computing is the core of networking, computing, storage, and application to end devices or data sources. It provides edge intelligent services to meet the critical needs of industrial digital fast connection, real-time services, data optimization, application intelligence, security and privacy protection.

2) Edge Computing Model

The Edge Computing Alliance currently promotes the convergence of three technologies (operations, information, and communications). Also, it defines four areas of computing objects: (1) device domain; (2) network domain; (3) data domain; (4) Application domain. The platform provides four major facilities: (1) networks; (2) calculations; (3) storage; (4) applications. The model is shown in Figure 2.

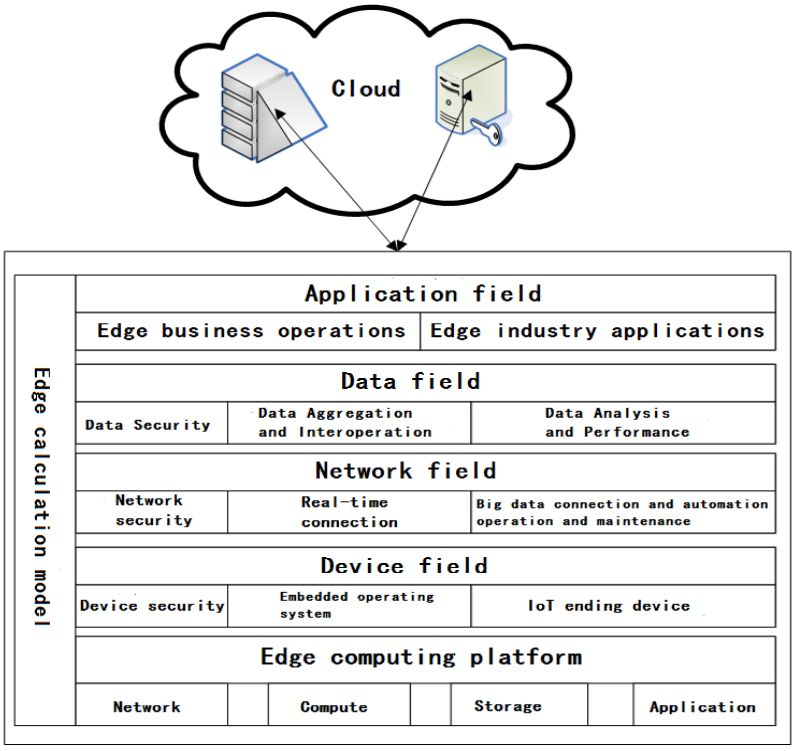

Figure 2. Edge Computing Model

3) Analysis of Edge Computing Technology

Edge computing is developed under an environment of high bandwidth, time sensitivity, and Internet of things integration. Its essence is the control of information which is a decentralized edge data processing. Edge computing technology is a key technology in fog computing systems.

For Intelligent Internet of Things, breakthroughs in technology mean that many of the controls will be implemented by local terminal devices instead of the cloud and the processing will be finished by the local edge computing layer. Meeting various requirements by the edge will undoubtedly improve the the efficiency of big data processing, reduce the load on the cloud and provide users with faster responses, smoothing flow of information throughout the Internet. At the same time, as the data is processed locally, the security of user privacy will also be enhanced.

4. Examples of Applications

In the corporation cases, given that a large manufacturing company produces detergent, a machine is seen as an overall system, an integrated mixer. It takes in different raw materials from different places, stirs them together, and produces a synthetic mixture. In the production process, the operating principle of the agitator is to rotate in a preset and timed manner, and the agitator barrel takes in different raw materials, which costs energy during the operation. What happens if we use the Intelligent IoT ecosystem to make this device a smart mixer? First, a large number of sensors are installed on the mixer to collect all of the parameter data, and then the terminal device sends the data back to the cloud server for data processing such as precision washing, calculation, comparison, and analysis.

Considering the construction of the Internet of Things, we use cloud storage and data analysis to analyze and set up data. However, in order to make machines "intelligent devices", we need to build a fog computing system and need to increase the ability to calculate data flow locally in real time, reducing the load pressure on the cloud server. It is a good vision that machines make better decisions with the equipment embedded by machine learning. In this way, we can build up a intelligent edge device system by machine learning to optimize machines' consumption. In a word, in the industrial application, the edge computing intelligent structure is to calculate machines' ability of intelligently processing data to the perception limit so that we can reduce the reduction of business loss and can increase the product qualification rate.

Based on these autonomous learning rules, the machine runs in the optimal energy consumption mode by adding or subtracting edge device parameter settings. Nevertheless, the machine can autonomously adjust its operating parameters. When data is transmitted to the cloud, the cloud uses new data set to update machine learning models so that it can update data rules and learn autonomously. Once updated, it will be pushed back to the edge, and the edge node updates the rules by updating the model to further improve the results, making the device autonomous and intelligent in the Internet.

Now, many edge devices are more realistic, including laptops, smart phones, smart watches, tablets and so on. For example, the Windows 10 restart manager can find out the best restart system and installation update time by calculating the system and learning users' modes after the system automatically downloads updates.

In commercial cases, reducing energy consumption is only one of the possible improvements. Edge computing can also be used for various real-time optimization processes, for example, to reduce asset failure or to improve output quality; through learning rules, the machine is automatically set up and changes its own operation setting, in order to avoid failure and to improve the outcome.In simple terms, by advancing marginalization, we can also push the intelligence to the edge and improve our equipment and assets by autonomous decisions.

Under the environment of Intelligent Internet of Things and big data, the future fog computing will be combined with cloud computing to provide a more complete hardware and software support platform for information processing of the Internet of Things.

5. Ends

This article focuses on the rapid development and application of intelligent Internet of things. It analyzes several shortcomings of cloud computing, including the more serious load for big data formed by terminals, the extension of data delay, the slow performance of information processing and

so on. For fog computing's features and advantages. The article also analyses 2 key technology of storage and calculation. Fog computing optimizes the storage replacement and refines the calculation model in its structure, and it is of great importance in the Intelligent Internet of Things + big data. In terms of application cases, the article gives examples in smart manufacturing industries and commerce which have been common to highlight the importance and value of key technologies for fog computing, and to emphasize the breadth and practicality of its application.

References

[1] Xue Yuhong. Cloud computing drives the development of fog computing. ZTE Technologies, 2017 Vol.23 No.1, P51-52.
[2] Bonomi F, Milito R, Zhu Jiamg, et al. Fog computing and its role in the internet of things [C] // Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing. New York, NY, USA: ACM, 2012: 13-16.
[3] After cloud computing, fog calculations begin [EB/OL]. http//www.cnblogs.com/dreign /p/3873176. Html, 2014.
[4] F. Bonomi. The internet of things and fog computing [C] The Eighth ACM International Workshop on Vehicular Internet working(VANET 2011). Las Vegas, USA: ACM, 2011.
[5] Zao Jk,Gan TT, You CK, et al. pervasive brain monitoring and data sharing based on multi-tier distributed computing and linked data technology [J]. Frontiers in human neuyoscience, 2014, 8.
[6] Dastjerdi A V, Gupta H, Calheiros R N, et al. Fog computing: principles, Architectures, and Applications [M] // Internet of Things. 2016.
[7] Yang Zhihe. Edge Computing Model of Internet of Things: Fog Computing. Internet of Things Technology, 2014. 12, p65-70.