

# Design of Distributed Network Massive Data Processing System Based on Cloud Computing Technology

Gong Fanghai

Department of Information Engineering, Lingnan Institute of Technology, Guangzhou, Guangdong, 510663, China

email: 7732694@qq.com

**Keywords:** Mobile Internet, Traffic Data, Data Processing

**Abstract:** In recent years, with the rapid development of cloud computing industry, big data processing technology is also constantly mature. At the same time, the scale of domestic mobile Internet market continues to expand, with more than 500 million users, bringing huge mobile Internet business data. In this context, the method of large-scale network data processing based on cloud computing big data processing technology is a very valuable research topic. From the perspective of the current situation of mobile Internet, on the one hand, mobile data services have risen sharply. And that puts a lot of pressure on the operator. Moreover, they need to invest more money in network construction and upgrading. Traditional voice and other SMS services are down. The operators then reached a point where they did not increase their earnings. Therefore, it is very important for communication operators to study the usage of large-scale mobile Internet data service resources obtained through service channels.

## 1. Introduction

With the development of Internet technology, human beings can enter the information age. With the popularization of mobile communication equipment, all kinds of mobile communication equipment can access the Internet at any time and anywhere. This is the most important function to distinguish traditional Internet from mobile Internet [1]. There is also the integration and innovation of mobile Internet and mobile technology. Mobile Internet has the characteristics and mobile attributes of Internet, and has many advantages such as high-speed, real-time, open, portable and so on. Based on the mobile Internet, people can access the Internet through mobile communication network in various mobile terminals such as mobile phones and tablets, and various network services such as language, video, social network and e-commerce can be used. And, more. With the popularization of 3G mobile communication network and the opening of 4G era, China's mobile Internet has entered a stage of rapid development.

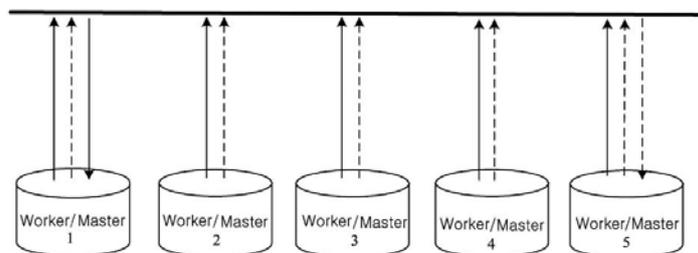


Figure 1 Load balancing processing block diagram

## 2. Network Data Processing Technology Based on Cloud Computing

The development of mobile Internet brings massive data to the massive network and enriches the source of massive data analysis. At present, there is no unified definition of big data [2]. Generally speaking, big data, usually processed data scale is too large, ordinary computer or server collection, including reasonable time will be processed data collection can not be completed. The network data

of big mobile Internet is generally connected to the Internet by mobile according to the mobile terminal devices. By the way, the large-scale data generated by users on the basis of a series of network activities such as information transmission, e-commerce transactions, etc. This kind of data has the common characteristics of big data. According to the definition of IBM, big data has three characteristics. As shown in Figure 1-2, the volume of data (volume), the diversity of data and the low density of data (veracity). At present, the data volume is increasing, and the data scale is expanded from the original GB to TB and Pb level. In the future, ZB level data can also be processed at the same time. According to IDC's forecast, by 2020, there will be more than 35zb data in the world [3]. At the same time, big data has a variety of data types and information sources. The object of initial data processing is all kinds of structural data. However, with the advent of the era of big data, data types show a variety of characteristics, semi-structured and unstructured data types are various, growing rapidly. Now, on the Internet, unstructured data far exceeds the volume of structured data. Meanwhile, Internet, Internet and other channels, Internet in various industries, portable devices, RFID are important information sources. In addition, big data also has the problem of low density. Because of the diverse nature of big data, the types of data it stores are also diverse. For example, for video files, the amount of valuable data can only account for one percent of the file size. Moreover, most video content is worthless data.

### **3. The Massive Data in Mobile Internet has its Complexity and Uncertainty**

Mobile Internet data shows a trend of multiple types and complex structures [4]. The application of mobile network is in a state of continuous growth, various types of network applications are increasing, and data types are also increasing. For example, the popularity of mobile game industry, wechat mobile e-commerce and other mobile Internet applications, independent data type and data flow transmission characteristics. Structured and unstructured data, including images, text, video, and many. All of these improve the technical difficulty of big data storage and processing in mobile Internet. On the other hand, the large-scale data collected through mobile Internet has great uncertainty. The data collected in mobile network, the incompleteness of data and the granularity of data collection, the quality of collected data is uncertain, and the original data processing method is very difficult. Good response to mobile Internet big data. Therefore, new theories and technologies are needed to collect, accumulate, process and analyze mobile Internet data. Technology, distributed computing, parallel processing technology, virtual technology, merge and grid technology, large-scale server cluster action integration, large-scale data storage, processing analysis and expansion can be applied software package. Real time scale cluster, increase computing power based on demand [5]. Cloud computing technology has the following characteristics.

### **4. Mass Data Collection and Distributed Transmission Technology**

Data acquisition layer is the lowest layer of the whole large-scale network data processing platform architecture. It collects network traffic through traffic monitoring equipment deployed on the network and stores it in the form of flow records. At present, the key data of the main nodes, such as the network link of the capital circle, the exit of the link intelligent network, the link between the export backbone network and so on, are collected. Generate data flow record list and binary architecture, large-scale network data processing + sending to Taiwan[6]. The platform needs to first receive the collected data and complete the decentralized storage of a large number of network data. This requires that these data be allocated to different data storage nodes for storage according to specific rules. Now, there are two main types of raw data. One is the original traffic data collected. This type of data is large and often has the size of several terabytes in an hour. A data center with a large-scale network data processing platform, and then upload and store it to the cloud platform locally. The other is the main business indicators collected by real-time analysis. This type of data is mainly based on the business needs of operators, real-time monitoring and mobile Internet related service quality. In order to protect the consistency and reliability of data, independent modules must carry out large-scale data transmission. At the same time, the changes of

external environment, such as the increase and decrease of the composition of memory data nodes, are the changes that may be caused by data transmission rules. Therefore, the topic of data transmission layer and compatibility system is the premise. In addition, the data transmission layer also provides several control management interfaces to control the whole distribution process [7]. From the perspective of function implementation, the task of data transmission layer is to aggregate the function modules of distributed forwarding. In order to realize the communication and distribution of control information and original data, TCP connection is used.

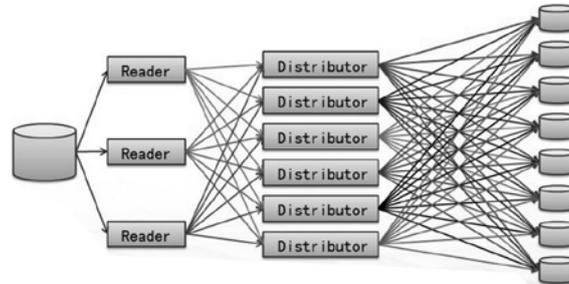


Figure 2 Flow chart of parallel data segmentation

## 5. The Problem of Mass Data Processing

### 5.1 Reliability Transmission of Massive Data Collection

Large scale network data collection is the foundation of large-scale network data processing platform based on Hadoop. As long as the integrity and accuracy of the collected data are ensured, the analysis results based on the data have its value and importance. Large scale network data collection needs to collect data from key nodes of mobile Internet. This data is the mobile Internet service collected from the backbone nodes of domestic operators [8]. After data collection, its data transmission, distribution, and intensive operation. In order to implement the level of complex number of transmission nodes, it is necessary to analyze the final internal network service quality operation. In the field of data memory of the sub type network data processing platform of data analysis service, it is necessary to analyze the user's behavior, user's business process control, etc. in the hydraulic mobile network.

### 5.2 Efficient Storage of Massive Network Data

The large-scale network data processing platform based on Hadoop needs to receive a large number of data sent from the data transmission layer and perform effective data storage [9]. The processing has the characteristics of high-speed transmission, huge data volume and half structure. HDFS technology realizes efficient and fast distributed storage of large amount of network data, and provides subsequent data processing services based on MapReduce technology.

### 5.3 Performance Problems of Massive Network Data Processing

The amount of semi-structured business data collected from the backbone nodes of operators is huge. Even with frequent data preprocessing, the amount of data that needs to be stored each week can reach tens of TBS. How to read, write and analyze these data quickly needs to consider the large-scale network processing platform.

### 5.4 Network Security of Massive Network Data Processing Platform

Security is an important problem that limits the development and popularization of cloud computing technology. Researchers have done a lot of research on the problem, but after the introduction of large-scale network data processing project of mobile Internet, the existing Hadoop standard is still the security detection of data flow, which needs to be strengthened. The large-scale network data processing platform based on Hadoop is in the network environment[10]. It is necessary to receive large-scale data streams collected in real time. At the same time, the application layer of the platform needs to provide multiple simultaneous data access interfaces to

various data users. The network security of this platform, the requirement of high speed and high precision of anomaly detection, the original network security technology with data encryption and security authentication, can not solve the security problem in this large-scale network data processing platform. Therefore, it is necessary to study the architecture of large-scale network data processing platform, the traditional network security devices including firewall, network and intrusion detection device, and the application of large-scale network data processing platform with a large number of data flow detection technology. Considering the improvement of detection accuracy and processing speed, the relevant research is carried out.

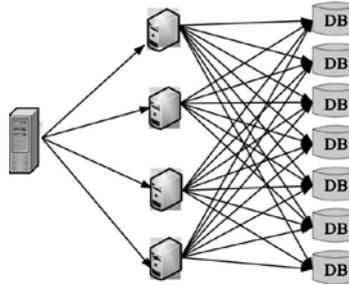


Figure 3 Overall system architecture

## 6. Conclusion

With the development of information technology, people 's demand for information will continue to increase. The contradiction between the technical requirements of rapid processing of massive data and the lag of current technical conditions will become more and more intense, which has become a hot issue in the world.

## References

- [1] Jiachen Yanga, Yurong Hana, Yafang Wanga,. Optimization of Real-Time Traffic Network Assignment Based on IoT Data Using DBN and Clustering Model in Smart City. *Future Generation Computer Systems*, 2017.
- [2] Markus AUER. Traffic services for vehicles: the process from receiving raw probe data to space-time diagrams and the resulting traffic service. *Frontiers of Engineering Management*, vol. 4, no. 4, pp. 490-497, 2017.
- [3] Song Xueya, Desheng Li. Development of Wireless Mobile Video Surveillance on Windows Mobile Using DirectShow Technology. *Journal of Computational & Theoretical Nanoscience*, vol. 14, no. 7, pp. 3163-3169, 2017.
- [4] Jennifer K Carroll, Anne Moorhead, Raymond Bond,. Who Uses Mobile Phone Health Apps and Does Use Matter? A Secondary Data Analytics Approach. *Journal of Medical Internet Research*, vol. 19, no. 4, pp. e125, 2017.
- [5] Muhammad Babar, Fahim Arif. Real-time data processing scheme using big data analytics in internet of things based smart transportation environment. *Journal of Ambient Intelligence & Humanized Computing*, no. 1, 2018.
- [6] Tasneem S J Darwish, Kamalrulnizam Abu Bakar. Fog Based Intelligent Transportation Big Data Analytics in The Internet of Vehicles Environment: Motivations, Architecture, Challenges and Critical Issues. *IEEE Access*, no. 99, pp. 1-1, 2018.
- [7] Fei Ding, Ruoyu Su, En Tong,. Toward a M2M-based Internet of Vehicles Framework for Wireless Monitoring Applications. *IEEE Access*, vol. 6, pp. 67699-67708, 2018.
- [8] Sanhita Das, Anuj Budhkar, Akhilesh Kumar Maurya,. Multivariate Analysis on Dynamic Car-Following Data of Non-lane-Based Traffic Environments. *Transportation in Developing Economies*,

vol. 5, no. 2, 2019.

[9] Syed Attique Shah, Dursun Zafer Seker, M. Mazhar Rathore,. Towards Disaster Resilient Smart Cities: Can Internet of Things and Big Data Analytics be the Game Changers?. *IEEE Access*, vol. 7, no. 1, pp. 91885-91903, 2019.

[10] Joahannes B. D. da Costa, Allan Mariano de Souza, Denis Rosário,. Efficient Data Dissemination Protocol based on Complex Networks' Metrics for Urban Vehicular Networks. *Journal of Internet Services and Applications*, vol. 10, 2019.