Research on Cloud Computing Processing and Optimization of Distributed Computer

Ye Sun[1]，Yu Zhang*

China Science Cloud Computing Hong Kong Limited

982809@qq.com

Abstract. With the rapid development of network traffic, video, pictures, information will produce a lot of data, which causes the problem of computer calculation and storage. With the increasing demand for computer processing capacity, the traditional computer computing method has been unable to meet the needs of society, which is also gradually developing in the direction of Cloud Computing (hereinafter referred to as CDC) and distributed computing. Through distributed computing, the computer can decompose a large task into many small tasks, which can be distributed to different computing resources. Therefore, distributed computing has become the main way of CDC processing, which can meet the existing market. At the same time, IPFS is the abbreviation of inter planar file system, which is a global, point-to-point distributed version file system. Through IPFS, we can connect all the computing devices with the same file system together, which can be called the information processing system. Firstly, this paper analyzes the related concepts. Then, this paper analyzes the architecture of CDC. Finally, some suggestions are put forward.

Keywords: Distributed Computing, Cloud Computing Processing, Optimization

1. Introduction

With the development of IT, computer information has become an indispensable part of people's life, which requires us to continuously improve the information computing ability [1]. Therefore, distributed computing has become the main way, which can carry out more efficient computing and processing [2]. IPFS is a point-to-point distributed file system, which aims to replace the traditional HTTP system [3]. Therefore, IPFS has learned many lessons from the past successful systems, which has become the cornerstone of CDC and cloud storage. At the same time, IPFS will become the cornerstone of blockchain. The key technology of CDC is decentralization. However, IPFS is a perfect solution, which can play a significant role [4-6]. The decentralized technology of IPFS has been applied to many fields, which can solve many problems of the existing platform [7].

2. Related concepts

2.1. Distributed computing

Distributed computing is to divide a large task into many small tasks, which can be distributed to different computing resources. A distributed system is a collection of independent computers. Therefore, the distributed computing system is just like a computer, which can effectively solve the balance between cost, efficiency and scalability. Since the 1980s, the distributed computer has become the focus

of research, including a variety of systems, such as middleware, SOA, grid computing, web service, Hadoop platform and so on [8]. Before the emergence of CDC, grid computing is the most typical representative of distributed computing. By connecting the hardware, software and information resources scattered all over the Internet into a huge whole, grid computing can enable people to use the geographically dispersed resources, which will complete a variety of large-scale, complex computing and data processing tasks. Grid computing is an Internet level distributed computing method, which mainly uses the distributed computing resources on the Internet. Grid computing is the closest to CDC, which can achieve centralized parallel processing of large computing tasks. However, the development of grid computing technology contributes a lot, which has become the technical basis of CDC development [9]. Distributed computing is one of the most important supporting technologies of CDC. Taking Google CDC as an example, distributed computing cases mainly include distributed data storage system GFS, distributed data management system Big Table, open source Hadoop platform, etc. In the field of PAAS and SAAS of CDC, distributed computing will be an important technology. With the method of distributed computing, we can release the binding relationship between users and large application systems. Overall, distributed computing breeds CDC. In the CDC environment, distributed computing reshapes the application form and service form of CDC, which provides a simple and feasible computing method for big data applications [10].

2.2. Advantages of IPFS

IPFS provides a new distributed Internet infrastructure. On the infrastructure, we can build many different types of applications. Therefore, IPFS is a global, mountable and versioned file system, which has many advantages [11]. First, decentralization is faster. All the data in IPFS are stored on the user's own computer, which is equivalent to distributing the central server of HTTP to each user. If other users want to get the data, they can extract it from the nearest user's computer. Second, reduce the dependence on the backbone. The transmission means of IPFS is obviously different from that of HTTP. HTTP mainly depends on the backbone network [12-14]. IPFS is mainly transmitted through nodes, which can be transmitted from one node to another. Therefore, IPFS can switch another node immediately even if one node fails. Third, permanent data storage. The storage mode of IPFS is very special, which is a fragmented storage mode. IPFS data can be divided into many parts, which leads to people can not get complete data. Therefore, data can be saved safely and permanently [15].

3. CDC processing

### 3.1. K-nearest neighbor method

K-nearest neighbor (KNN) is a typical ranking classification algorithm. After one judgment, the sorting algorithm can output documents belonging to multiple categories. Through KNN, we can calculate the similarity of each text in the training sample set, which can find k most similar training texts. At the same time, we can select a threshold, which can be sorted according to the score. The similarity between k nearest neighbor training samples and the test sample is shown in Formula 1. K neighbors calculate the weight of each class, as shown in formula 2.

$$Sim(d_x, d_j) = \cos(d_x, d_j) = \frac{\sum_{i=1}^{n} w_{xi} w_{jx}}{\sqrt{\sum_{i=1}^{n} (w_{xi})^2 \sum_{i=1}^{n} (w_{jx})^2}} \tag{1}$$

$$p\left(\overline{x}, c_j\right) = \sum_{di \in KNN} Sim\left(\overline{x}, d_i\right) y(\overline{d_i}, c_j) \tag{2}$$

3.2. CDC architecture

CDC can provide elastic resources on demand, which is a collection of services. The architecture of CDC can be divided into three levels: core service, service management and user access interface, as shown in Figure 1.
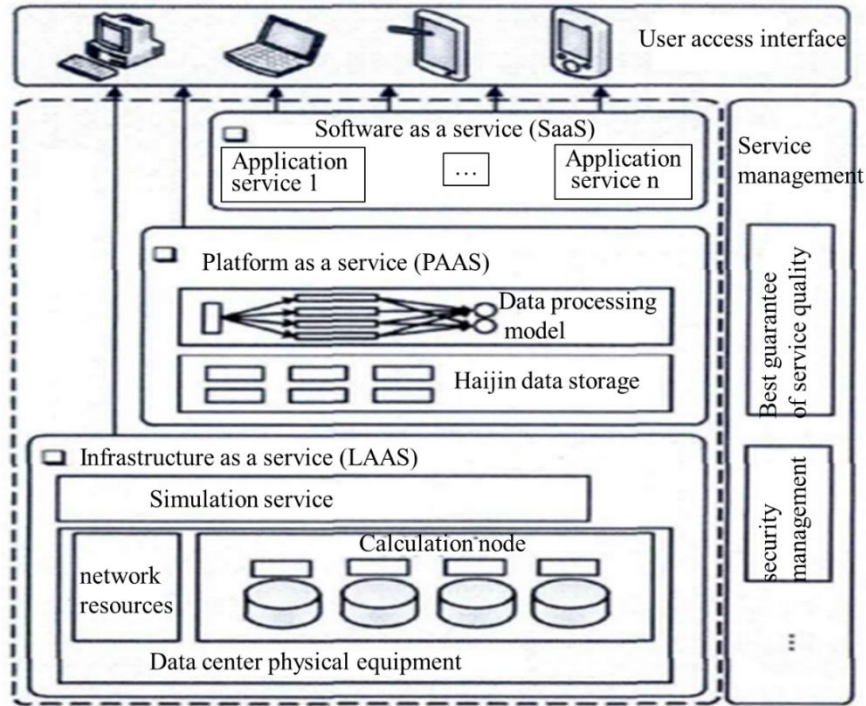
Figure 1. CDC architecture diagram

3.3. File storage verification scheme

File storage verification scheme is the basis for service providers to prove the integrity of their stored data to service consumers. After each service, the service information will be written into the blockchain. Therefore, CDC has become an important computing and storage mode of blockchain. The file storage method is shown in Figure 2.
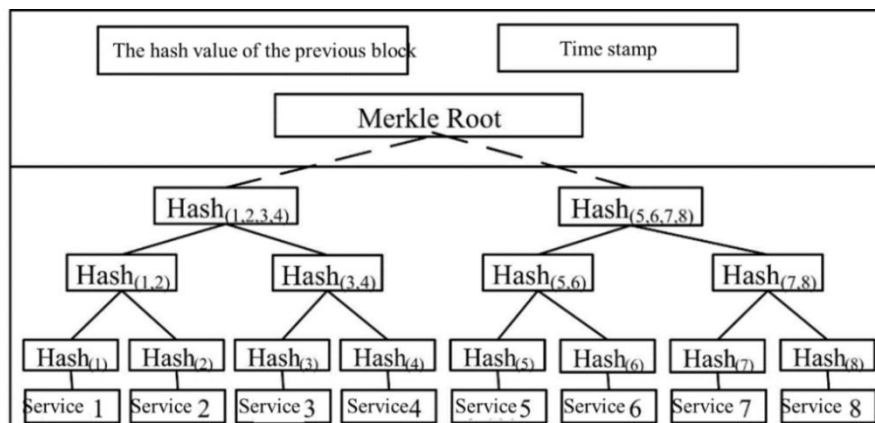
Figure 2. Block structure of this scheme

4. Important technologies of CDC

4.1. Location service based on mobile cloud

As an indispensable supporting technology of mobile CDC, location-based services can provide a variety of location-based services around the architecture of location-based services, such as mainstream location technology, location index, query processing and so on. Location services based on traditional positioning technologies such as GPS cover a wide range, which has been widely used in many fields, such as military, transportation and so on. However, GPS has many problems, such as weak penetration, high positioning energy consumption and so on, which can not fully meet the requirements of new mobile applications such as accurate indoor positioning and user action recognition. Through CDC, we can complete location services of mobile cloud, such as automatic shopping guide service, patient monitoring in smart home, etc. Mobile CDC model has been used to build new location services, which can solve and form an important supporting technology.

4.2. Energy saving technology of mobile terminal
The battery capacity of mobile terminals is growing slowly, and the contradiction between the rapid and rich mobile applications and the limited power of mobile terminals is becoming increasingly prominent. Through CDC, we can achieve energy saving in many aspects, such as data transmission energy saving. The proportion of wireless data transmission energy consumption in the energy consumption of mobile terminals is also increasing. Through cellular network transmission data, we can usually use RRC protocol for the whole process of mobile terminal energy consumption measurement. The results show that there is too much tail energy consumption in the process of data transmission, which reduces the energy utilization of mobile terminals. By changing the time threshold of tail energy consumption, we can reduce the number and time of jumping to the tail energy consumption state. Through transmission scheduling, we can reduce the tail energy consumption. Through the virtual ending mechanism and double queue scheduling algorithm, we can schedule the prefetch data and delay transmission, which can adjust the time threshold.

4.3. Data security and privacy protection
While obtaining rich services of CDC, mobile users will face more security threats such as privacy exposure. This requires us to strengthen the data security and privacy protection of CDC. In mobile CDC environment, users' data and computing tasks migrate through wireless network, which can realize and support online query, multi-user data sharing and so on. In view of the limited computing resources and mobility of mobile terminals, cloud authentication platform can avoid the degradation of service performance caused by multi-user parallel access. Through a series of new cryptographic mechanisms, we can choose between encryption and attribute based encryption. By introducing an access structure to associate ciphertext or user private key with attributes, we can flexibly represent access control policies, which can provide fine-grained access authorization for data.

5. Conclusion
At present, CDC has become an important way of data processing and storage, which can optimize decentralization and other measures. IPFS is a new application based on HTTP, which is an attempt of new technology. Therefore, distributed computing will become the main computing method of computer in the future, which can improve a variety of IT.

References
[1]   Cui Yong, Song Jian, Miao congcongcong, Tang Jun. research progress and trend of mobile CDC [J]. Acta computer Sinica, 2017, 40 (02): 273-295.
[2]   Ding Jian, Wang Huaimin, Shi Peichang, Wu Qingbo, Dai Huadong, Fu Hongyi. Trusted cloud service [J]. Acta computa Sinica, 2015, 38 (01): 133-149.
[3]   Li Jia. Intelligent logistics model reconstruction based on big data CDC [J]. China's circulation economy, 2019, 33 (02): 20-29.
[4]   Lu Xiaobin, Wang Jianya. Analysis of the current situation of CDC Adoption Behavior [J]. Journal of Chinese library, 2015, 41 (01): 92-111.

[5] Lu Xiaobin, Wang Tao. Research on technical improvement and optimization of massive data analysis process by Google's three CDC technologies [J]. Library and information work, 2015, 59 (03): 6-11+102.

[6] Pengxiaosheng, dengdiyuan, chengshijie, wenjinyu, Li Chaohui, Niulin. Key technologies of power big data for smart grid application [J]. Journal of China Electric Engineering, 2015, 35 (03): 503-511.

[7] Qin Rongsheng. Research on the impact of big data and CDC technology on audit [J]. Audit research, 2014 (06): 23-28.

[8] Shi Weisong, zhangxingzhou, wangyifan, zhangqingyang. Edge calculation: present situation and Prospect [J]. Computer research and development, 2019, 56 (01): 69-89.

[9] Shi Weisong. Edge computing: a new computing model in the era of Internet of things [J]. Computer research and development, 2017, 54 (05): 907-924.

[10] Sun Lei, Hu Xuelong, ZhangXiaobin, Li Yun. CDC solutions for biomedical big data processing [J]. Journal of electronic measurement and instruments, 2014, 28 (11): 1190-1197.

[11] Wang Guiling, Han Yanbo, Zhang Zhongmei, Zhu Meiling. Stream data integration and service based on CDC [J]. Acta computa Sinica, 2017, 40 (01): 107-125.

[12] Wang Yu Ding, Yang Jia Hai, Xu Cong, Ling Xiao, Yang Yang. Overview of CDC access control technology [J]. Acta software Sinica, 2015, 26 (05): 1129-1150.

[13] Xu Baomin, Ni Xuguang. Development trend and key technology progress of CDC [J]. Chinese Academy of Sciences, 2015, 30 (02): 170-180.

[14] Yang Qingfeng. Key technology prediction and strategic selection in CDC era [J]. Journal of Chinese Academy of Sciences, 2015, 30 (02): 148-161+169.

[15] Zhou Yuezhi, Zhang Di. Near end CDC: opportunities and challenges in the post CDC era [J]. Acta computa Sinica, 2019, 42 (04): 677-700.