Analysis of Large Data Mining Platform Based on Cloud Computing

Fei Qi[1,] *, Furong Yang[2]

[1] School of Information Engineering, Longdong University, Qingyang 745000, Gansu, China

[2] College of Foreign Languages, Longdong University, Qingyang 745000, Gansu, China

Keywords: cloud computing; large data; platform; research and analysis

Abstract: With the development of the times and the progress of society, the era of big data has arrived, and has gradually become an important manifestation of the height of a country's strategic development. Big data is the basic condition of finance and innovation in all walks of life. The emergence of big data can be applied in data analysis in a timely manner, which is important for China to transform from knowledge economy era to intelligent economy era. By analyzing the data, cloud computing is relatively small, a lot of researches are about research process of big data and manufacturing, while ignoring the application of computing platform for big data and cloud, eventually leading to the researches on big data and the cloud is still at a basic level, so it is urgent to conduct research and analysis of large data platform based on cloud computing.

1. Introduction

At the present stage of cloud computing technology has the characteristics of dynamic resource allocation and scheduling of certain, which can overcome the slow speed, low efficiency and high energy consumption of traditional data transmission in a certain extent, can effectively reduce the cost of data, to further improve the efficiency of data mining. This paper will analyze the large data mining system of cloud computing, and study the existing problems of large data based on cloud computing at present and the corresponding countermeasures.

2. The development of the cloud

In the era of rapid development of the network, the Internet is seen as a blue sky, then spread on the network all kinds of countless resources is cloud, big data in the Internet is the "cloud", "cloud" big data some calculation is called cloud computing. That is to say, for cloud computing, we should say that cloud computing is the way and method of computing, and big data is the raw material of computing.

With the development of society and technology, many important technological revolutions appear gradually. There are virtual technology, NoSQL database technology and hadop in cloud computing. These three technologies have promoted the revolution and development of big data era to a certain extent. Among them, NoSQL database technology is mainly used to store data well, and its storage capacity and storage capacity are greatly improved compared with the past. Many units and the company on the Internet to build their own "cloud", the data and information stored in the cloud, ensure that relevant personnel can according to their own needs, using different computer equipment whenever and wherever possible, to obtain information on the data, a good backup, free to worry about data loss.

3. The status of big data and the importance of cloud computing in large data mining

First of all, in the face of the huge data storage case, big data not only has the characteristics of large capacity, it also has certain real-time, highly developed in the present stage of the network today, automatic sensing technology is constantly to provide us with more analysis of real-time data, and these data is long-term growth. In addition, the number of large data in large data is unstructured, and there is no connection between them. Secondly, because the data itself does not

produce certain value, it needs to be decomposed and analyzed to a certain extent, so that it can help units and companies effectively transform the analysis results into productive forces, so that we can enhance the core competitiveness of enterprises.

With the development of the global economy, the amount of data is increasing gradually, and the low value density characteristics of a lot of data are developing gradually. In the era of the big data age, the importance of its data is of great significance. The value of big data is not simply displayed by itself, but by data mining, data can be obtained from low value density data and get its potential value. When data are mined, we often need data acquisition and statistics for many times, and solve and optimize them very well, and we need a lot of time to access large scale data and consume time. For a long time and complex data calculation to a certain extent is low efficiency, energy consumption is slow and slow, but the calculation of dynamic resource allocation and scheduling as well as the application of the platform in the cloud, which can meet the performance requirements of mining calculation. That is to say, the current large data mining platform in the process of establishment and development can not be separated from the help of cloud computing technology.

Cloud computing refers to the distribution of complex computing tasks was constructed by computer in the "cloud" large, computing and storage ability and application ability of users are able to demand in accordance with the distribution of good, in order to improve the efficiency of obtaining better data. But the general data mining is to process a large number of incomplete, noisy and fuzzy data, and extract more valuable information hidden in the data by optimizing screening. Because of the complexity of many data, we need some storage space to calculate. It can effectively control data storage cost and operation control, so as to improve data mining efficiency and better transform traditional data mining ability.

4. The principle of cloud computing

In cloud computing, many computers and servers are not arranged at random, but need to be distributed in a distributed manner. When cloud computing is carried out, computer resources can be integrated to a certain extent, and the corresponding computing is decomposed into distributed computers. So that we can be able to implement a certain calculation. In cloud computing, resolution is also a very important content. Because the content of cloud computing is very much, the speed is slow in the process of running, and it takes longer time. The use of network can split large data stored in cloud computing to form many small programs, and then divide small programs into every computer to process data, so as to effectively reduce data acquisition time. The cloud computing technology effectively combines network computing, distributed computing and conformal computing, which is more effectively applied to various fields.

5. Application Research of large data platform based on Cloud Computing

At the present stage, the cost of cloud computing processing mode adopted by various industry fields in China is relatively high, and it is not able to expand the requirements according to users' usage. With the development of society, the amount of application data of many users is increasing constantly, and the degree of data processing is also improving. This will lead to this kind of big data processing mode to fail to meet the requirements of customers to a certain extent. After years of research, based on cloud computing technology and big data platform can be very good to solve the above problems, according to the application of data service user characteristics of different levels to provide safe, reliable and efficient, in order to better the user data were used to improve the quality of data. Big data after cloud computing has good practical function in the actual application process. It has virtualization and transparency for resources to better meet customer requirements. The new computing processing mode represented by cloud computing is more capable of processing, effectively improving storage space, storage reliability and storage security, so as to better meet the needs of customers.

The analysis and research of large data platform through the use of cloud computing, capable of processing capacity of the platform and storage capacity are improved, and can use the flexible use

of a variety of data processing mode, enable it to face more data format and data operating system in the practical application, provide more management means and mechanism. To support large data can quickly become functional target data, and play an important role in the future system environment and application configuration.

6. Parallel data mining platform based on Cloud Computing

According to the research and Analysis on the present stage of complicated data, data mining and data processing need to run in high speed and high efficiency condition, so the need for a large-scale computing, cloud computing platform to achieve this goal, through multiple virtual machines will be in accordance with the resources need to be allocated to the user, to provide the resources for the future the rate of. On the platform of cloud based data mining architecture is mainly to use the concept of partition in the database, the data sheet after more conducive to storage in each point, with a central hope to summarize and maintain the information of each node, and between the boundary of each algorithm is not fixed, for but by the application of different algorithms, can be used in various distributed environment, so as to ensure more flexible and efficient use of these traditional data mining platform.

Parallel data mining platform based on cloud computing can effectively help cloud computing to store large amounts of data for better calculation, so as to facilitate the solution of massive data and high efficiency requirements of large data. At the present stage, the research method of the parallel data mining platform for cloud computing has made some achievements. At present China has a number of researchers using a large data mining platform based on Cloud Calculation of means, realizes the preprocessing and classification of relevant data and various types of data mining algorithm, mining platform performance to a certain extent, the distributed data were accelerated, which some enterprises and units can be calculated by cloud operation. For example, China's Mobile Corporation has developed a parallel data mining tool based on cloud computing, data development and operation processing is easy for programmers to a certain extent, the greater the further realization of the storage and analysis of massive data, to the data processing and mining, to ensure the data storage company and improve. Its core part needs to establish a large cluster data file, and it is a simplified distributed programming model, which can be used to process and generate large data sets efficiently.

7. Status and Countermeasures of data mining platform based on Cloud Computing

Based on data mining platform based on cloud computing, cloud computing and data mining are perfectly integrated, which effectively integrate dynamic resource allocation efficiency and powerful data mining of cloud computing. The structure of data mining platform for cloud computing is complex and rich in content. Cloud computing has used a lot of algorithms and algorithms to analyze and study all kinds of data, and to a certain extent, it can achieve better data mining results. At the present stage, cloud computing provides data management subsystem and account management subsystem for data mining platform, which can be responsible for more different tasks and ensure close combination between them.

Cloud mining platform also includes a server and database based on group, these groups are able to use the platform owned by local or remote server to the service for more time, effective utilization and allocation of data resources, and the database of information platform and its data storage. Data mining platform based on cloud data has certain advantages, but because of the relatively low technology of cloud computing, it is not mature enough. Based on cloud computing technology, we can build data mining platform. There are still many problems to solve in this process. In the process of cloud computing, many data mining algorithms still face great challenges in parallelism. They need to overcome some uncertain factors such as data mining methods and results to a certain extent. Because data mining is built on a certain basis of cloud computing, and cloud computing itself is a relatively unstable software. In the face of cloud computing, we need to improve the correctness, security and reliability.

In order to solve the above problems, we must pay more attention to the professionalization and personalization of the data mining cloud computing platform. According to the problems that each company and unit have, study and analyze the problem and find the most suitable answer. The corresponding service platform is effectively built according to the different service objects. When the large data mining platform for cloud computing is established, the corresponding service platform must be built according to the different service objects. And in data mining algorithm selection, we should study its versatile and variability as much as possible, to a certain extent, be able to carry out the inspection and visualization, and protect some privacy data.

At the present stage, the research and development of the data mining platform system based on cloud computing and its application have achieved good results to a certain extent. This type of system has the characteristics of very high efficiency and big data, but because this kind of cloud computing technology is still at the primary level of development, facing many difficulties, we need to continue our research and analysis. In the process of development, the security of cloud service software must be overcome, and the uncertainty of data mining algorithms and results must be overcome. Therefore, in the process of building data mining platform based on cloud computing, we must pay attention to the combination of actual effect and pay attention to the personalized and universal nature of design, so as to increase the privacy protection of privacy data.

Conclusion: In conclusion, with the development of social science and technology, the era of big data has come quietly, and to a certain extent, it can decide the height of a country's strategic development. For the units and enterprises themselves, data is the basic condition of their internal finance and innovation. The emergence of big data can be applied in data analysis of units in a timely manner, to a certain extent, promoting the transformation of China's intelligent economic era. During the research and analysis of big data and cloud computing in China, many data research is too focused on the research and manufacturing process of big data, but neglects the research on the application of big data and cloud computing platform. This paper analyzes the large data mining system of cloud computing, and studies the existing problems of large data based on cloud computing at present and the corresponding countermeasures. At present, cloud computing technology has some characteristics of dynamic resource allocation and scheduling, which can overcome the slow transmission speed, low efficiency and high energy consumption in traditional data, effectively reduce the cost of data, and further improve data mining efficiency.

References

[1] Xiao-Hui L I. Analysis on the Application of Data Mining Based on Cloud Computing [J]. Journal of Changchun University, 2012.

[2] Yu L, Zheng J, Shen W C, et al. BC-PDM: data mining, social network analysis and text mining system based on cloud computing[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012:1496-1499.

[3] Wang Z Q, Li H L. Research of massive Web log data mining based on cloud computing[C]// Fifth International Conference on Computational and Information Sciences. IEEE, 2013:591-594.

[4] Qing H E, Zhuang F Z, Li Z, et al. PDMiner: a cloud computing based parallel and distributed data mining toolkit platform[J]. Scientia Sinica, 2014, 44(7):871.

[5] Wang Y, Zhao Y W. Transplantation of Data Mining Algorithms to Cloud Computing Platform when Dealing Big Data [J]. 2017.

[6] Wang Y, Wang L, Liu H, et al. Large-Scale Clinical Data Management and Analysis System Based on Cloud Computing[M]// Frontier and Future Development of Information Technology in Medicine and Education. Springer Netherlands, 2014:1575-1583.