Application of Big Data Technology in Computer Network Information Management

Aliya Awakeli*, Yikelamu Awakeli

*575263189@qq.com

Xinjiang Vocational & Technical College of Communications, Xinjiang, China

Abstract. With the wide application of computer network technology, it provides many conveniences for modern people's life and work. People can reasonably use computer-related information and data technology to develop the value of various information resources according to their actual needs. Moreover, the application of big data technology can complete the processing and analysis of information resources with the help of computer equipment. Based on this, this paper introduces the concept of big data technology and its main application fields in network computer network information management, and introduces the application of big data technology in computer network information management from the aspects of cloud computing and data backup, so as to provide corresponding reference for computer network information management.

Keywords: big data technology; computer; network information management

1. Introduction

Nowadays, the application of big data technology in computers is becoming more and more important, and it has developed into the mainstream. Computer information system has provided convenience for people in both life and work. Big data technology can summarize network information data, discuss the long-term slow development process of things, and provide people with more targeted services. However, when people use the convenience provided by big data technology and the rapid development of network technology, how to deal with and develop massive information data has become the focus of many people's research. Therefore, how to apply big data technology to computer network information management has also attracted the attention and attention of many scholars.

2 The meaning of big data and its main application fields

2.1 The meaning of big data

Big data technology is a modern technology corresponding to the formation of massive information data in the information age. When it is difficult for people to summarize and analyze information data in a short time, it can efficiently save manpower, material resources and time for data analysis and strengthen people's ability to control information. In recent years, big data technology has developed rapidly and has been widely used in various industries, especially in computer information systems. In addition, big data technology

contains large-scale information data, which also provides efficient thinking for people to study the development trend of things, but during the practical application, there may be problems of data loss or inaccurate analysis direction. [2]

2.2 Main application fields of big data

First, in the information system that people mainly use at present, the biggest content of cost input is the transmission of information data. With the rapid development of the new era, the traditional information data transmission technology can no longer meet the actual needs of users, which makes a significant difference between the current information storage system and the large-scale growth of information data, and the concept of big data also comes into being. The emergence and wide application of big data technology has pushed the transmission and management of computer equipment information systems into a new stage, and big data has put forward a new concept. [2] And because of the reasonable calculation method of big data and the integrity of analysis information, the information data transmission capacity and storage capacity of the computer system are greatly enhanced, thus enhancing the stability, security and compatibility of computer network information management and control.

2.3 Impact of big data technology on data analysis

As far as data is concerned, storing the knowledge that can be transmitted is a very basic problem, and the more difficult problem is the analysis and application of data. For the database composed of rich types and huge amounts of data, during the actual use, users only need to extract and analyze their valuable data in order to obtain very ideal results. In the era of big data, the amount of information has greatly increased, which in turn poses a certain degree of challenge to information refining. For this reason, more and more users put forward new data analysis methods. By giving priority to data screening, processing and research, users' valuable news can be extracted as soon as possible. Therefore, the analysis speed of big data technology hormone user information data has effectively improved the efficiency of information data application and ensured the quality of data analysis to the greatest extent. [3] Therefore, the data put forward more novel requirements and arithmetic methods in time, which led to technical reforms in computer network information transmission and analysis, which had an enlightening effect on the future development of information technology and accelerated the development and innovation of computer information technology management.

2.4 Security of big data

With the rapid establishment of big data platform, it provides convenience for people to obtain the necessary data, and then achieve personal goals through data research or results. However, because big data itself has high application value, through the study of data, we can infer the user groups, business activities and consumption patterns related to industry secrets. In particular, massive information and data usually involve personal privacy. Whether from a commercial point of view or a social point of view, big data research may have a certain degree of impact on social stability. Therefore, in the era of big data, the rapid development of computer security technology can be promoted, and the state can predict some social trends by studying some specific social information data. The computer network information

management system also needs to improve the protection awareness of users' information data through big data technology to promote the improvement of information security system. It is precisely because big data has extremely high commercial value and social value that it is necessary to strengthen the working system for data protection, so as to prevent big data from being used by illegal elements and financial threats to social stability.

3 Application of Big Data Technology in Computer Network Information Management

3.1 Website log mining algorithm based on cloud computing

For large-scale portals, the daily visits may not be less than 100 million person-times, and the total amount of website log files may reach tens or even hundreds of GB. Therefore, it is necessary to adopt a website log mining algorithm suitable for cloud computing to provide help for computer network information management. In this regard, we can adopt a user browsing preference path mining algorithm based on useful preference. With the expansion of the Internet and the increasing number of websites, PBPU algorithm needs to invest a lot of time and cost to process computer network information data on a serial basis. If it is adapted to be parallel, the time efficiency will be significantly improved by realizing parallel processing through multiple processing devices. Therefore, this paper uses Hadoop, an open source cloud computing infrastructure, to achieve the fusion of PBPU algorithm and MapReduce, and subdivides the large-scale tasks that run independently in a centralized environment into multiple subtasks that can be processed in parallel, and then hand them over to each processing device in a group for execution at the same time, so that the calculation and storage consumption can be shared equally. The calculation flow of PBPU algorithm based on MaprReduce can be subdivided into two steps. One is to preprocess the web log file with MapReduce programming structure, and the other is to complete the parallel execution of PBPU algorithm through MapReduce programming structure. All processes are implemented by MapReduce process, and all tasks are further subdivided into several subtasks, and then distributed to different service nodes and executed. During the actual implementation of the parallel PBPU algorithm, the web log files stored in the distributed file system are preprocessed, and the irrelevant contents are eliminated. [4] The frequency statistics of one of the access paths are made through the MapReduce programming structure, and then the access spelling is corrected through the MapReduce programming structure, and the corrected sub-paths that meet the user-defined useful scale preference are merged, and then output. The specific process is as follows:

First, data preprocessing based on MapReduce. There are generally two forms of network logs, namely CLF and ECLM, among which the latter is frequently used. Time/Date represents the time requested by the user behind the network server; The URL address of the page requested by the user whose URL meets the standard; Refeer represents the URL of the referenced page. If the user directly enters the URL to visit the page or visits it before operation, the content of this column is empty; Agent browses all operating systems on behalf of users. Not all the raw data in the network log are necessary for analysis. Therefore, the data preprocessing process for Web logs before submitting mining algorithms includes two processes: data cleaning and frequency scanning. In the process of analyzing the user's browsing preference path, computer

network information data management only pays attention to the access interface and reference interface, so it is necessary to preprocess the original data and eliminate the information data that has nothing to do with the conversation. [5] The daily visits of computer network information management system can reach hundreds of millions of people, forming a huge number of files. We can try to use MapReduce model to achieve high parallelism, build a data analysis pipeline, and realize parallel preprocessing of massive log files in order to obtain more ideal results. Within the framework of MapReduce, the map mode is responsible for refining all access records Referer and URL, and transmitting them as output. The input and output modes in the Map stage are shown in Table 1.

Table 1 Input and output of the Map arrangement link

Map stage Input	Map stage Output
Record the offset in the file, record 1	(Referer1, URL1)
Record the offset in the file, record 2	(Referer2, URL2)
……	
Record the offset in the file, record n	(Referern, URLn)

The output of the Map function is analyzed by MapReduce, and then transmitted to the reduce function. During the period, the key/value is arranged and grouped, that is, the outputs with the same (referrer, URL) are merged, where m represents the number of different (referrer, URL) and n1, n2 ... nm represents the corresponding (referrer, URL) access frequency. Table 2 shows the input and output of Reduce link. [6]

Table 2 Input and output of the Reduce arrangement link

Reduce stage Input	Reduce stage Output
$(Referer_1,\ URL_1)$	$\square Referer_1,\ (URL_1,\ n_1))$
$(Referer_2,\ URL_2)$	$\square Referer_2,\ (URL_2,\ n_2))$
……	
$(Referer_m,\ URL_m)$	$\square Referer_m,\ (URL_m,\ n_m))$

Reduce combines all the values contained in the key, and then sends them to the output file by means of (Referer, (URL, n)), and finally saves the input file in HDFS, as shown in Table 2.

After data preprocessing, the management system only extracts two valuable fields, Referer and URL, and excludes other information unrelated to mining, and at the same time merges the same record content, which greatly reduces the total amount of log files. The data preprocessing link code is as follows:

Input: web log file f

Output: (Referer,list(URL, n))//n represents the access frequency of the path Referer URL.

Traverse all in the f file

//Define Map function

//inputkey: The total offset of this record in all files (in bytes)   Value: the actual content

of all records.

```
        //Output key:Referer value:list(URL)
    FOR EACH Record IN f
                    //Record belongs to one of the access records in f
    Key=Referer
    Value=URL
    Emit(Key,    Value)
    END FOR
        //Define Reduce function
      //input key:Referer value:list(URL)
    //Outputkey:Referer
    List[i]={1}//The array List[] records the access frequency of all paths < Referer,URL >
    FOR EACH atom IN list(URL)//Atom is a < Refereri,URL > record.
            IF URLi uniform with the URLk(i≠k)
                List[i]++
    key=Refereri
            values=(URLi,list[i])
            Emit(key,values)
        END IF
    END FOR
```

Secondly, PBPU algorithm based on MapReduce. MapReduce parallel computing process belongs to the master node in the cluster, which subdivides the files that are processed by the processing pipeline and stored in HDFS to record the relationship between page connections into different file blocks, and then transmits them to the nodes running Map tasks in the cluster. After this node obtains the scene module for storing the connection relationship file, it converts all the records into a certain format of < key,value > and transmits it to the master node. The node running the Reduce task then reads the intermediate results output by the Map task, summarizes the intermediate results by combining the key values entered, and finally stores the < key, value > pairs in a certain format in HDFS. [7]

3.2 Practical application of data backup technology

In the new era, the business environment is becoming more and more complex, so enterprises need to strengthen the protection and backup of information data and enrich the fault-tolerant space of the system. A sound disaster recovery system can screen data, protect and save them by classification, and minimize the probability of risks.

At present, computer network information management systems tend to ignore data backup. According to relevant research and investigation, less than 20% of enterprises use computer network information management systems and data backup systems, and less than 5% use disaster recovery systems of data backup systems. It can be seen that enterprises lack understanding of the application of data backup system and disaster recovery system. Therefore, it is necessary to actively popularize the value of data backup technology nationwide and strengthen enterprise information protection. [8] At the same time, to ensure the security of enterprise information data, it is also necessary to know the comprehensive details of backup while backing up information data regularly. In terms of data backup, the most frequently used media are magnetic tape, mobile hard disk and USB flash drive. Enterprises need to select the corresponding data storage media according to the actual situation. At the same time, it should be stored in different places as far as possible, and sound computer equipment and remote communication equipment should be arranged in the storage location, so as to quickly start the data backup system after an emergency and avoid data damage. In addition, it is necessary to build an efficient connection between the data backup system and the enterprise management system to improve the recovery speed of damaged data and ensure the smooth development of relevant work in enterprises. At the same time, it is necessary to transmit data information to the inside and outside of the mobile hard disk, compress the information data by remote technology, make it into a CD-ROM for storage, and manage it separately to strengthen the security of enterprise information.

4. Conclusion

The application of big data technology in computer network information management can greatly reduce the cost of management, improve the overall quality and efficiency of management, and ensure the security of information data. Therefore, enterprises need to make clear the meaning and application fields of big data technology, and apply it reasonably in network information management to effectively improve the quality of network management and give full play to the value of information data.

References

[1]Yu Qing. Application Analysis of Big Data Technology in Computer Network Information Management[J]. Shihezi Science and Technology, 2022(01):23-25.

[2]Cheng Yaohua. Application of Big Data Technology in Computer Network Information Management — Comment on "Project Management Division Course of Information System"[J]. China Oils and Fats, 2022,47(1):166.

[3]Miao Jingfeng, Li Qiang. Research on the Application of Big Data Technology in Computer Network Information Security[J]. China New Telecommunications, 2021,23(12):77-78.

[4]Li Xiaokang. Application of Big Data Technology in Computer Network Information Security[J]. Wireless Internet Technology, 2021,18(7):86-87.

[5]Liu Shiping, Li Shan, Yu Yang. Application of Big Data Technology in Computer Network Information Management —Comment on "Big Data Management: Technology, Method and Practice of Data Integration"[J]. Modern Radar, 2021,43(1):92.

[6]Song Shilei. Application of Big Data Technology in Computer Network Information

Management[J]. Information Recording Materials, 2020,21(12):172-173..

[7]He Binying. Application of Big Data Technology in Computer Network Information Security —
Comment on "Computer Network Information Security"[J]. Plating & Finishing, 2020,42(3):47.

[8]Cheng Lingyun. Research on the Application of Big Data Technology in Computer Network
Information Management[J]. Satellite TV & IP Multimedia, 2020(5):63-64.