Implementation of a Cloud Computer-based Automatic Classification and Processing System for Large Data

Andrew Kim\textsuperscript{1,}\textsuperscript{*}

\textsuperscript{1}The Browning School, New York, New York

**Abstract:** Data technology is an important technology applied in the process of social development, which plays a very important role in data and information operation, storage and analysis. It is directly related to the accuracy of data operation and promotes the efficiency of the operation of a data platform to some extent. This paper analyzes cloud computing technology and large data processing, and puts forward a new automatic data classification and processing system, which can ensure the operation of the system more efficiently.

**Keywords:** cloud computing; Large data; Automatic classification system

1. Introduction

In the background of the development of information technology and network technology, the application of it has strengthened the research and development on data technology to a certain extent. The processing of data is the core of current information technologies and intelligent technology applications. Therefore, on the basis of this, cloud computing technology has been developed, which has promoted the improvement of data computation and processing efficiency to some extent. In the specific application of cloud computing technology, the design of large data classification system can be improved and the operation quality of the system is higher.

2. Large-scale data-processing technology

Large data analysis is an important work in the process of data processing, which plays a very important role in processing data. Under the background of current technology development, the classification, storage, calculation and analysis of data are the core of the data processing, which is directly related to the application of datum, and has great significance for the rational use of digital technology. At present, in the process of large data processing, including the following aspects, to complete the application of data analysis. First, the visual analysis of Analytic Visualizations data reveals the structure of data processing according to the data needs of the requester through processing technology, which enables the user to understand the core applications for processing data. Secondly, the application of Data Mining Algorithms, in the actual large data processing, data mining algorithm can accurately carry out the value mining of data in datasets. The third is predictive analytics, which is the ability to actually define data definitions and parts of performance during data mining to make data analysis more efficient.

Fourth, as a semantic engine, semantic engines are said to be able to perform real analysis of unstructured data engines, thus ensuring more efficient data extraction. Fifth, data quality and data management. In the process of large data processing, the requirement data in the dataset are well defined according to the requirements of data users. Sixth, data storage. In the process of large data processing, data storage is also a very important content. The application of the technology to the storage of data is to protect the data reasonably and make it convenient to use in the future.

In the process of the application of advanced technology, it is very important to apply the technology of large data processing, especially in the design of computational and statistical software, which is directly related to the efficiency of technology application.

3. Brief description of cloud computing technology

From the definition point of view, cloud computing is a large data processing technology. In the beginning of its research, it is both a data computation method and a distributed computing method. In the real process of large data computation, using cloud computing technology, it can decompose big data or large program and feed back the structure of computation to users, which can improve the precision of data operation and analysis. With cloud computing technology, the accuracy of data processing can be improved, and in the actual data computation process,
application to the cloud computer technology can complete
the data operation of more than 10000 units in a few seconds.
In the development of cloud computing technology, distributed computing, utility computing and load balancing, parallel computation, network storage, thermal redundancy and virtualization have been formed, which are very important for data computation and data analysis processing.

4. Large data processing technologies in the context of cloud computing applications

The cloud computing technology is the development product of the large data processing technology, but the data technology will still play an important role in the future production, and it will also influence the result of data treatment and classification. This paper mainly studies the application of cloud technology to the automatic classification of big data, so as to improve the efficiency of automatic data classification, the following is a detailed analysis of this data system.

4.1. Functional analysis of system requirements

In the application of cloud computing technology in large data automatic classification system, we should first complete the functional requirements analysis of the system. The following is a detailed analysis of the functional requirements of large data classification systems for cloud computing.

4.1.1. Data acquisition function analysis

Data acquisition is one of the basic functions in the cloud computing data classification system. In the process of practical application, data capture is the initial module of data analysis. Data processing can be guaranteed to acquire data source only by analyzing data first. In the design of this system, the related personnel of MapReduce program design the data acquisition function, and promote the realization of data retrieval function.

4.1.2. Functional analysis of data processing classification

After the data is obtained, the system will complete the analysis and classification of data and improve the accuracy of the classification. At present, text classification system design, system processing module design and text evaluation system are mainly used in the design of system data classification, so as to ensure that the system can reasonably complete the data analysis and classification function.

4.1.3. Functional analysis of data processing applications

In the process of system construction, the related personnel of mapreduce program also completed the requirement analysis of the data processing application function, and in the actual data analysis application process, completed data application functional module design, promoted the application of cloud computing data classification system.

Through the requirements of data acquisition, data processing and data application, this paper designs and implements the system application of cloud computing technology.

4.2. Major structural design of the system

In the process of designing the automatic classification and processing system for large data, three parts are completed, namely, architectural design, hardware structure design and software design.

4.2.1. Framework design

In the application of cloud computing large data automatic classification and processing system, the design of the system’s organizational structure is also a key part of system design. In the actual system architecture design, we first complete the design of the institutional form. In designing the system, the hybrid mode is selected to design. The three-layer C / S mode and B / S. mode are designed in the practical design process, and the structural form is designed with the mixed mode to realize the database access. These include Web servers and server data access. In addition, in the mixed-mode structure design, the Web server can also provide the functions of database access, data computation and data application, thus realizing the optimization of the system’s main structure.

4.2.2. Structural Design of System Hardware

In the process of the application of large data analysis and processing system, it mainly includes the data acquisition systems, data processors, automatic data storage modules and hardware interface design (figure 1). During the actual design process, the hardware system design is completed, and the efficiency of system operation is improved to a great extent, so as to ensure the reasonable realization of design function. In addition, in the hardware design of this system, we choose to use SOA to construct the model and test the components of the system.

Firstly, the data acquisition device is the core device of the hardware system, which mainly completes the functions of data collection and data transmission. In the hardware design of the data acquisition device, the Ti network control chip and the SCM are chosen as the core components. Its working power supply choice is designed for 5V working supply, and the single chip microcomputer working system provides a separate 3V supply. In addition, in the design of data acquisition device, cloud computing interface is selected as the working interface of the system to ensure the efficiency of system data collection. In the application of cloud computing interface, data transmission can be realized and A / D data signal transformation completed on top of the MCU, which realizes function.

Secondly, the data processor is the core hardware of the whole system, which also has an important influence on the system’s operation. In the design of this system, the IXP2400 processor is selected, and the implementation of the engine
data signal processing and data processing with the device of 1xP24400 can complete the processing of core data.

Thirdly, the application of automatic data storage module design is very important in large data classification system. In the actual design process, the storage module design uses the C8051F series MCU and AT45DB80 hardware, using two systems combined design, thus completes the data storage function well.

4.2.3. System Software Design

The system software design process is also an important part of the data classification and processing system operation, which has a very important influence on the system work quality. In the process of designing the software, it mainly includes the display end design, the service end and the workflow design.

Firstly, the design of netizen network terminal is selected in the system software display end design, so as to ensure more efficient operation of system and to improve the performance of the display terminal to the maximum extent.

Secondly, in the system design process, select SOA server for the service end design. Through SOA client design, the large number of server households can be connected to the cloud interface, and then the concrete application of cloud computing technology in systems design can complete, thus promoting the efficiency of system operation.[4]

Finally, the system software process design includes the steps such as Heritrix data acquisition, data analysis, feature extraction and data classification processing, which realizes the large data classifying processing under the cloud computing technology..

4.3. Optimal analysis of ROS communications

In the design of this experiment, we choose a large data automatic classification and processing system which has been designed as a whole to carry out the test and test, and apply the system designed in this paper to the construction of traffic operation cloud platform in S province. In the actual system design and application process, the system is connected from PaaS layer and interconnected with the traffic GPS system, and system deployment is done well. In the actual system application process, mainly for S province channel logistics data automatic classification analysis, including traffic network, ports, river flow and other aspects of data classification processing, the following figure 3 is a specific statistical map of the system. From top to bottom for traffic network, ports, river flow data statistics and classification of the situation of S Province, traffic logistics can be scheduled and fixed-point statistics, to promote the accuracy of its data classification. With the application of the system, the efficiency of data classification processing on the S transport logistics cloud platform improved by more than 32%.

4. Summary

This paper expounds the application of cloud computing technology in the automatic classification and processing system of large data from the aspects of system architecture design, function designing, software design and hardware design. I hope it will be helpful to the application of data processing system.

References

[1] Pivovarenko Y. (2020) Negative Electrization of the Sargasso Sea as the Cause of Its Anomaly. American Journal of Electromagnetics and Applications, 8 (2), 33-39.

[2] Qian Zhaolou. The Implementation Path of Large Data Automatic Classification System Based on Cloud Computing. Science and Technology Innovation, 2019 (13): 67-68.

[3] Priory of Sion. Research on Design of Automatic Classification Processing System for Large Data in Cloud Computing Environment[J]. Electronic Design Engineering, 2019, 27 (23): 167-170.

[4] Feng Xinyang, Shen Jianjing. Large Data Clustering Algorithm Based on Yarn Cloud Computing Platform and NMF[J]. Information Network Security, 2018 (8): 43-49.

[5] Kim-woo. Design of Large Data Storage Model Based on
HDOOP Cloud Computing Platform[J]. Computer fan, 2018 (016): 215-215

[6] Deepening the Core Architecture of Linux, by Guo Xuxue, People' s Postal and Telecommunications Publishing House, June 2010

[7] Visual SLAM 14 Talk: From Theory to Practice", Gao Xiang Electronic Industry Publishing House March 2, 2018

[8] https://robots.ros.org/

[9] Heise, L., Greene, M., Opper, N., Stavropoulou, M., & Equality, N.A. (2019). Gender inequality and restrictive gender norms: framing the challenges to health. The Lancet, 393, 2440-2454.

[10] Andy FZ, Paymon R, Kevin CC. (2018). Advances in Proximal Interphalangeal Joint Arthroplasty: Biomechanics and Biomaterials. Hand Clinics. 34 (2): 185-194.