Construction of accounting big data analysis platform driven by cloud computing technology

Xiaoyun Yan
Guangdong University of Science & Technology, Guangdong, 523083,China

ABSTRACT

In the current market competition, if enterprises want to occupy a favorable position, they must use information technology tools, give full play to their own production and management advantages, improve market competitiveness and resist risks. The emergence of cloud computing brings hope to the current financial management information construction model. In this paper, an accounting big data analysis platform driven by cloud computing technology is established, and an accounting big data cloud computing task scheduling algorithm based on improved FA(firefly algorithm) is proposed. Aiming at the problem that traditional FA is prone to premature and fall into local optimum, chaotic disturbance strategy is introduced to disturb the part with low target value, and fireflies flying out of the search space are bounced back according to the rebound theory based on real physics in firefly position update, so as to maintain the optimization ability of the population. The research results show that this algorithm can shorten the total execution time of cloud computing tasks by about 21.06%. It is proved that the improved FA has better optimization ability and real-time performance.

Keywords: Cloud computing; Accounting; Big data

1. INTRODUCTION

Cloud computing is a new data computing method, which mainly connects the clients through the network system, and then integrates the information and data resources in the clients into a huge data resource pool. After that, the resources can be allocated and managed according to the actual needs. On the premise of massive data, cloud computing can analyze data more efficiently and collect, store and process more useful information for users. The storage and application of big data, due to the existence of a large number of semi-structured and unstructured data, are too different from the storage of traditional data, which requires us to analyze it more deeply[1]. Therefore, how to quickly, conveniently and accurately search, mine and acquire the required knowledge from the ocean of big data has become the need of the development of the times.

In the current market competition, if enterprises want to occupy a favorable position, they must use information technology tools, give full play to their own production and management advantages, improve market competitiveness and resist risks. The emergence of cloud computing brings hope to the current financial management information construction model [2-3]. Literature [4] thinks that the advanced technology and systematic work of cloud computing can make cloud computing better serve enterprises, and studies the concept, characteristics, key technologies and deployment modes of cloud computing. Literature [5] analyzes some key factors that enterprise users should consider when renting cloud computing services, and gives the index system and evaluation method to help enterprises decide whether to adopt cloud services based on their own situation. Literature [6] discusses the feasibility of enterprise applying cloud computing to information construction.

*109759252@qq.com

Seventh International Conference on Mechatronics and Intelligent Robotics (ICMIR 2023), edited by Srikanta Patnaik, Tao Shen, Proc. of SPIE Vol. 12779, 127792I · © 2023 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.2689382

The arrival of the era of big data provides rich information resources for the financial analysis of accounting department managers. At present, with the development and maturity of IT technologies such as big data and cloud computing, its application in accounting practice is becoming more and more extensive, and it is urgent to optimize or establish a new accounting platform to meet business needs. In view of this, this paper introduces big data theory into the construction of accounting information system, expands the connotation of accounting data, enriches the theoretical system of accounting informatization to a certain extent, and discusses and designs an accounting big data analysis platform driven by cloud computing technology from the perspective of actual needs of enterprises, providing new ideas for realizing low-cost accounting control of enterprises.

2. RESEARCH METHOD

2.1 Overall framework of accounting big data analysis platform

Cloud computing is a service related to information technology and software, relying on the Internet. Through servers, storage and networks, cloud computing can be used by users according to their needs in the form of services. Network information retrieval involves artificial intelligence, database technology, information science, natural language processing, machine learning and many other knowledge and disciplines. The accounting big data formed by the combination of big data and accounting has played a strong role in promoting the development of accounting. And accounting big data is also generated on the basis of big data and original accounting data. It is not only pure accounting data, but also information and data after processing, so the form of data is not fixed [7-8.]

In the existing traditional accounting system, the accounting is single, that is, only the information that can be measured in money can be processed, but the processing ability of intangible assets and human capital and other non-financial information is very limited. When an enterprise develops to a certain scale, the application of new technology will inevitably produce a huge amount of information, and how to use the accounting system to combine experience with information and refine it through deduction, comparison and induction to realize the transformation of value into knowledge will be a big challenge for enterprises. After building the basic cloud service platform, enterprises only need to pay for the services provided by service providers, and no longer pay for the maintenance of hardware and equipment at all, saving costs again.

The construction of accounting big data analysis platform driven by cloud computing technology must follow the requirements of accounting standards, accounting system and other relevant laws and regulations, and integrate the effective original accounting data and processed accounting data into accounting big data, and collect, store, analyze and apply it to explore its value. Generally speaking, its analysis model includes three layers: the top layer, the middle layer and the data center layer; It is subdivided into five modules, the top layer includes two modules, the workflow and user interface subsystem, and the middle layer includes parallel data and data preprocessing subsystem. In addition, there is a data center, cloud computing data center, which is not included.

Traditional auditing often limits the scope of auditing to the financial information reflected in the electronic account set, and the processed data information is often incomplete, which makes it impossible to fully understand the audited object. The establishment of financial sharing center enables enterprises to see the dynamic situation of capital flow in real time, and the application of cloud computing and other technologies enables enterprises to calculate the stock or demand of funds. The docking of financial cloud platform and online banking enables the group to timely dispatch funds to units in need, and centralize the remaining idle funds for unified management to increase profitability[9]. The construction framework of accounting big data analysis platform driven by cloud computing technology mainly includes data source, data processing, data storage, data warehouse and data visualization, and the standardization and security mechanism of accounting big data run through the platform construction, as shown in Figure 1.

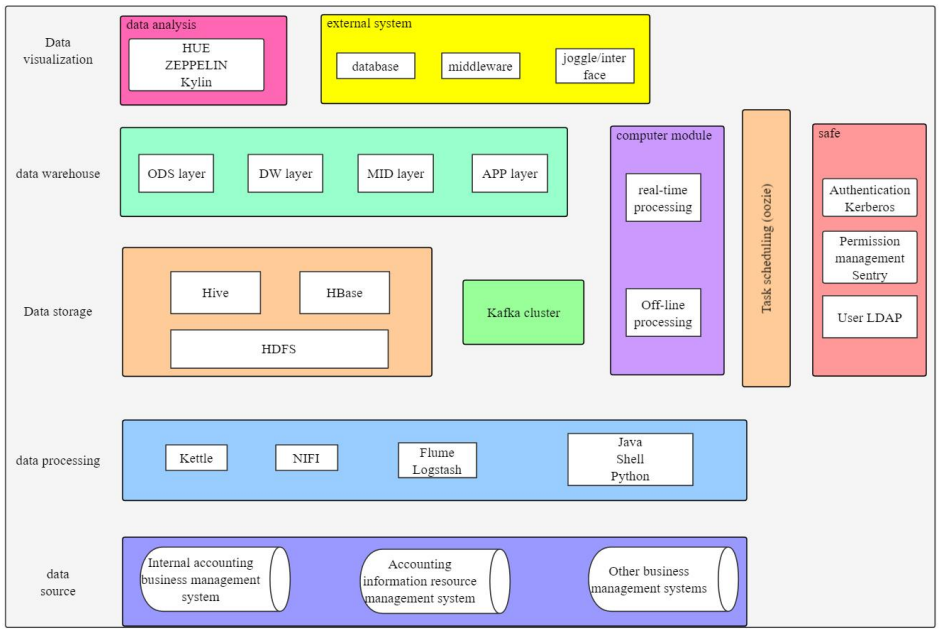

Figure 1. Cloud computing technology-driven accounting big data analysis platform architecture

Among them, the cloud service platform layer: network devices, storage devices, operating systems, etc. built by system management, stores the processed data in different databases, human resource accounting data, product sales accounting data, customer accounting data, etc. are uniformly stored in the basic database, and method base, model base, knowledge base, etc. are uniformly stored in the analysis database, providing convenience for later analysis and use.

In financial financing decision-making, models such as capital cost, financial risk and capital demand prediction are called to analyze the capital structure and use of enterprises at first. Secondly, the weighted average cost of capital is used to analyze the effect of capital use, and then the demand for funds is predicted by the sales percentage. Finally, the financing decision is made according to the knowledge law in the knowledge base. Activity-based costing is usually used to predict operating costs, and cluster analysis is used to test the influence degree of different factors on costs, and then the corresponding model is automatically selected according to the changing trend of historical output and cost data, so as to realize the requirement of forecasting future costs.

Financial forecasting is to predict and manage the financial risk, operating conditions and costs of enterprises by using linear regression method through financial analysis results and accounting information of industries. After analyzing and processing the collected information, we can find the indicators of risk early warning, and then quantify the indicators to form a comprehensive monitoring system of financial risks. On the basis of the monitoring system, a comprehensive early warning and monitoring model can be established in a scientific way, and a comprehensive system for the detection, management and evaluation of enterprise financial risks can be realized [10]. The decision-makers of enterprises only need to input relevant keywords in the search bar by using instant query on the mobile terminal, and the system will automatically present the corresponding accounting information according to the captured keywords, and explain the relevant accounting information by using the existing models in the system, so as to provide intelligent decision support for information users.

2.2 Task scheduling in accounting big data cloud platform

In the big data environment, too extensive data sources and diverse data forms, in the process of data collection, the collection of the same kind of data often requires multiple channels to operate together to obtain. And the data obtained from different channels are often quite different. According to the correlation of many events, data mining technology excavates the set of general rules to explain data, and uses training and self-learning to extract new relationships hidden in data. Based on this service model, enterprises can build an accounting big data analysis platform driven by cloud computing technology, without defining the storage scalability of accounting data, and without managing and controlling the infrastructure and network of the cloud, but they can control the deployment of applications and configure the

corresponding environment to improve the return on investment of enterprises in software. At the same time, it should be noted that building software based on PaaS mode will be a long-term development trend.

PaaS implemented by cloud computing technology provides services such as automatic deployment, running monitoring and flexible scheduling for applications and middleware. With the application of VM virtual resource pool and Open Stack resource pool[11], it realizes multi-tenancy, service isolation, high availability and flexible expansion of public resources. With the continuous growth of business on the cloud and the migration of old systems to the cloud, it is becoming more and more important for resource managers to schedule tasks reasonably and use resources efficiently. In this paper, a task scheduling algorithm for accounting big data cloud computing based on improved FA(firefly algorithm) is proposed. Aiming at the problem that traditional FA is easy to fall into local optimization, chaotic disturbance strategy is introduced to disturb the part with lower target value, and fireflies flying out of the search space are bounced back according to the theory of real physical rebound in firefly location update, so as to maintain the optimization ability of the population.

In FA algorithm, the moving operation caused by mutual attraction between fireflies determines the optimization ability of the algorithm. Brightness and attraction are the two most basic concepts in FA algorithm, which are the basis for solving optimization problems. β represents the attraction between fireflies X_i, X_j, which is a measure of the distance r_{ij} between them, and it changes with r_{ij}. The smaller r_{ij}, the greater the attraction β, whereas the smaller the attraction β, the specific form is shown in (1).

$$I(X_i) \propto -f(X_i) \tag{1}$$

$$\beta = \beta_0 * \exp\left(-\gamma * r_{ij}^2\right) \tag{2}$$

Where γ stands for absorption factor and is a constant.

FA is applied to accounting big data cloud computing task scheduling. In this paper, fireflies are coded by natural numbers, and the dimension of fireflies' position is equal to the number of cloud computing tasks to be scheduled. Assuming that there are m tasks and n assignable resources in the accounting cloud environment, fireflies can be encoded into m-dimensional vectors:

$$X = (x_1, x_2, \cdots, x_m) \tag{3}$$

Where: $1 \le x_i \le n, 1 \le i \le m$, each one-dimensional subscript of firefly represents a task number, and each one-dimensional component x_i represents the resource number assigned to this task by the cloud environment.

In the initial stage of FA iteration, fireflies keep a large step size and a good global optimization ability, so as to avoid the premature maturity of fireflies and falling into local minima. When the neighbor set is empty, fireflies i try to move 20 times by random step size, and choose an optimal position to update the position of fireflies i during these 20 times, which can avoid the situation that a group of fireflies lose their mobility with great probability.

$$s(t) = s_{\max} * e^{-p*q^{1.5}} + s_{\min} \tag{4}$$

$$q = \frac{t}{t_{\max}} \tag{5}$$

$$x_i(t+1) = x_i(t) + rand(1, D) * rand(1) \tag{6}$$

t represents the number of iterations; $t_{\max}$ stands for the maximum number of iterations; $s_{\max}, s_{\min}$ represents the maximum moving step size and the minimum moving step size respectively; p is the step change amplitude adjustment factor; D stands for the dimension of fireflies.

Chaos is a common phenomenon in nature, and most nonlinear systems will have different degrees of chaotic motion [12]. In order to prevent the convergence rate from slowing down due to the lack of population diversity in the late iteration, chaotic sequences are introduced to disturb fireflies and enhance their search ability. In this paper, Tent mapping is selected to generate chaotic sequences, and its formula is:

$$x_{n+1} = \begin{cases} 2x_n, & x \in [0,0.5] \\ 2(1-x_n), & x \in [0.5,1] \end{cases} \tag{7}$$

When the convergence speed of the population begins to slow down, the fireflies are sorted according to the fitness function value, and the 20% fireflies with low fitness value in the current population are disturbed by Tent chaotic map. The basic execution steps based on the improved FA are shown in Figure 2 below.

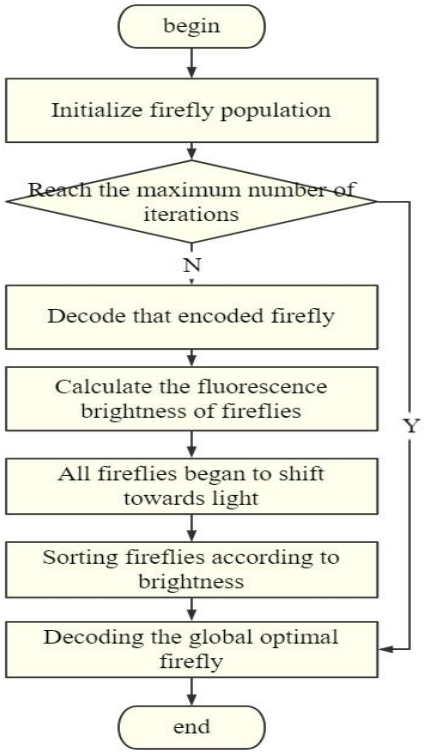

Figure 2. Process based on improved FA

3. EXPERIMENTAL ANALYSIS

The DatacenterBroker class of CloudSim platform provides virtual machine proxy function, and users can extend the function of the proxy class by adding a method that implements a custom scheduling algorithm to it, thus completing the test experiment. The running environment of this experiment is: the CPU of the processor is I 52.8 GHz, the memory is 8 GB, the operating system is Windows 10, and the development environment is Matlab.

The parameters of the algorithm are as follows: the individual population size of fireflies is 100, the number of iterations is 300, the fluorescein volatilization factor is 0.3, the fluorescein updating rate is 0.5, the initial fluorescein quantity of fireflies is 4, the initial step size is 0.1, the maximum moving step size is 1, and the minimum moving step size is 0.01.

In order to illustrate the advantages of the improved algorithm, Figure 3 shows the convergence curve before and after the improvement of the algorithm. It can be clearly concluded that the improved FA adds chaotic local search operation on the basis of the basic FA, and the chaotic local search operator has strong local mining ability, so it can improve the local search operator of the algorithm.

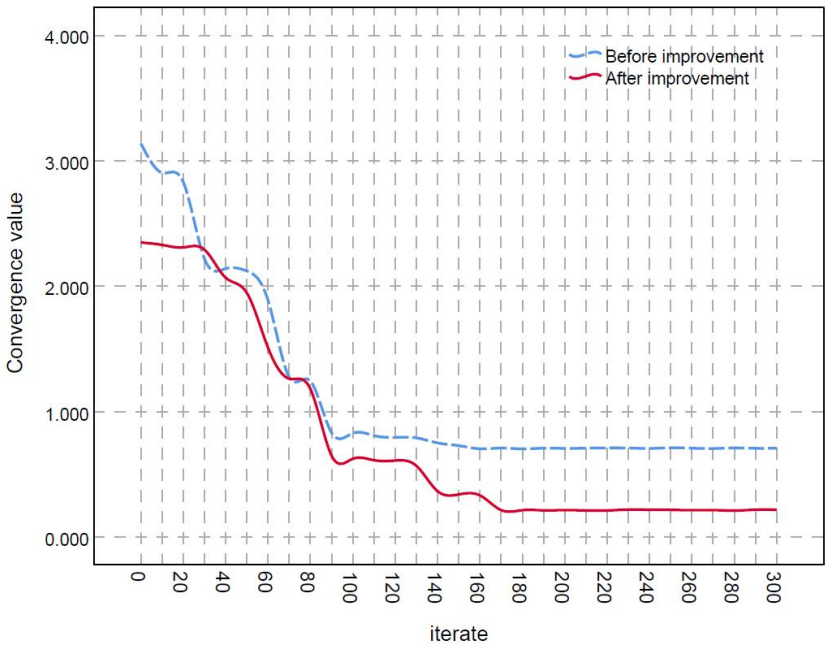

Figure 3. Convergence curve before and after improvement of the algorithm

5,000 cloud computing tasks are distributed to 10 ~ 15 resource nodes, and the experimental results are shown in Figure 4.

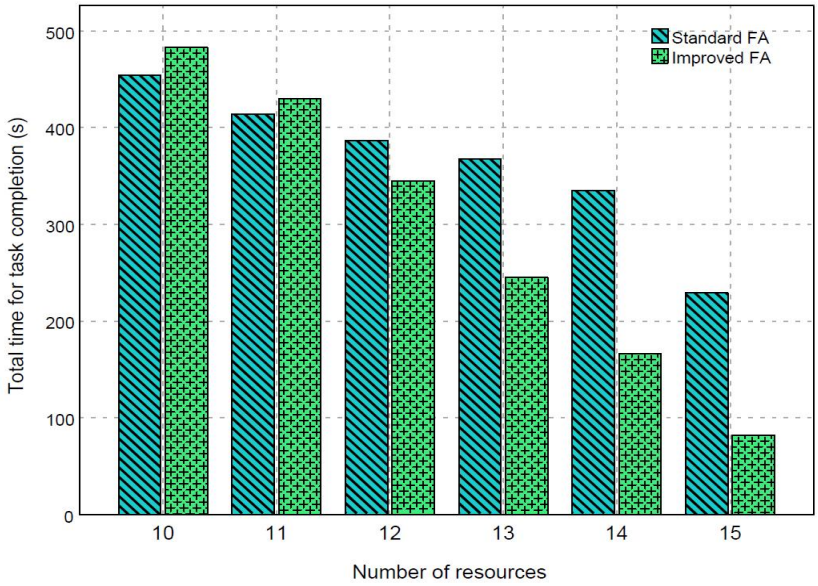

Figure 4. Variation curve of total task completion time with the number of resources

With the increase of computing resources, the total task completion time gradually decreases, which is because the increase of resources leads to the decrease of the intensity of task competition resources. By comparison, it can be seen that this algorithm can shorten the total execution time of cloud computing tasks by about 21.06%. It is proved that the improved FA has better optimization ability and real-time performance.

4. CONCLUSION

On the premise of massive data, cloud computing can analyze data more efficiently and collect, store and process more useful information for users. The storage and application of big data are different from the storage of traditional data because of the existence of a large number of semi-structured and unstructured data, which requires us to analyze it more deeply. The accounting big data formed by the combination of big data and accounting has played a strong role in promoting the development of accounting. In this paper, an accounting big data analysis platform driven by cloud computing technology is established, and a cloud computing task scheduling algorithm for accounting big data based on improved FA is proposed. The research results show that this algorithm can shorten the total execution time of cloud computing tasks by about 21.06%. It is proved that the improved FA has better optimization ability and real-time performance.

REFERENCES

[1] Chen, J. L., Construction of an accounting big data analysis platform based on cloud computing. Accounting Newsletter, (6), 5 (2017).

[2] Wu, D. J., Some thoughts on the construction of big data platform for CPA industry. China Certified Public Accountant, (12), 2 (2018).

[3] Zhu, B., The application of big data accounting analysis platform supported by cloud computing technology in the field of finance —— Comment on "internet plus Accounting Archives Management". Chemical Reagents, 42(2), 1 (2020).

[4] Lai, L. Z., Construction of cloud audit platform system under the background of big data. Finance and Accounting, (3), 3 (2017).

[5] Lai, L. Z., Research on the audit ecosystem framework of accounting firms based on cloud audit platform. China Certified Public Accountant, (12), 4 (2015).

[6] Sun W., Analysis of the challenges and opportunities of big data to financial accounting. Business Research, 000(011), 137-138 (2018).

[7] Xia, S. T., Research on the application of cloud accounting under the background of big data. Economics, 4(4), 64-65 (2021).

[8] Zhang, Y. Q., Research on the operation mode of enterprise financial integration under cloud computing-taking State Grid as an example. Friends of Accounting, (24), 3 (2018).

[9] Wu, K. F., Liu, W. T., Li, Y. H., Su, Y. P., Xiao, Z., Hu, S. L., Power Big Data Analysis Technology and Application Based on Cloud Computing. China Electric Power, (2), 111-116,127 (2015).

[10] Wang, Z. Q., Wu, F. W., Cheng, C., Yang, Y. H., Research on Meteorological Big Data Analysis Technology in Urban Planning. Urban Development Research, 27(11), 6 (2020).

[11] Ammari, A. C., Labidi, W., Mnif, F., Yuan, H., Zhou, M. C., Sarrab, M., Firefly algorithm and learning-based geographical task scheduling for operational cost minimization in distributed green data centers. Neurocomputing, (14), 490 (2022).

[12] Peng, H., Xiao, W., Han, Y., Jiang, A., Xu, Z., Li, M., Multi-strategy firefly algorithm with selective ensemble for complex engineering optimization problems. Applied Soft Computing, (120), 120 (2022).