Research on data distribution and load balancing scheduling framework in cloud environment based on fuzzy logic concept

Dongmei Li
(School of Information Engineering and Artificial Intelligence, Henan Open University, Zhengzhou 450046, China)

Abstract: Framework based on fuzzy logic concept is proposed to solve the problem of data distribution and load balancing scheduling in cloud environment. The framework comprehensively considers the system load, data access mode and node resources, aiming at the balanced distribution of data and the reasonable scheduling of load. By fuzzy processing the resource status information in the cloud environment, the advantage of fuzzy logic can reflect the actual situation more accurately. At the same time, considering the size, type and access frequency of data and other factors, the data is distributed to different server nodes to achieve the balance of data distribution. Based on the load balancing algorithm, the task allocation strategy is dynamically adjusted according to the resource status information and task load of each server node to improve the performance and reliability of the system. The effectiveness and superiority of the framework in the cloud environment are proved by the experimental evaluation. This research has practical significance and application value for data management and load scheduling in cloud environment.

Key words: cloud computing; Data distribution; Load balancing

I. Introduction

With the worldwide popularity of the Internet and the explosive growth of cloud computing services, more and more applications are being deployed on cloud computing platforms, resulting in a sharp increase in the amount of data on cloud computing platforms. This has made the cloud computing platform enter a new stage of accelerated development, and the platform's data storage, data transmission and data processing capabilities are facing great challenges. At home and abroad, there have been many researches on load balancing scheduling and data distribution in cloud environment.

Domestic researches mainly focus on the improvement and optimization of load balancing algorithms, including load sensing based load balancing algorithm, dynamic weight based load balancing algorithm and predictive model based load balancing algorithm. These algorithms try to dynamically adjust the resource allocation according to the current load situation of the system to achieve load balancing. The performance and resource utilization of the system can be improved by dynamic adjustment according to the real-time load situation. However, the algorithm design is complex and requires high real-time performance, and may be limited by the accuracy of load perception and prediction.

The foreign research mainly focuses on the improvement and optimization of load balancing algorithm. For example, load balancing algorithm based on traffic prediction, load balancing algorithm based on adaptive scheduling and load balancing algorithm based on multi-objective optimization and so on. These algorithms try to achieve load balancing through prediction and optimization techniques to improve the performance and resource utilization of the system. Although the system can schedule resources according to the prediction and optimization results, it has certain intelligence and flexibility. However, the algorithm complexity is high, and the implementation and management cost may be high.

This paper presents a scheduling framework for data distribution and load balancing in cloud environment based on fuzzy logic concept. The fuzzy logic theory is applied to the fuzzy processing of the resource state information in the cloud environment, and the data distribution is considered, and the load balancing algorithm is combined to improve the performance and reliability of the system.

II. Design and implementation of data distribution and load balancing scheduling framework in cloud environment

This paper puts forward a cloud environment of data distribution and load balance scheduling framework mainly divided into data analysis and prediction, fuzzy logic modeling, adaptive conditions, scheduling execution, monitoring and feedback these five steps.

1. Data analysis and prediction

Analyze and forecast the data in the cloud environment to determine the data distribution and load, mainly using machine learning and data mining technology for data analysis and prediction. The service layer, attribute layer and indicator layer of cloud computing provide rich tools and resources to support the realization of data analysis and prediction. At these levels, machine learning and data mining technologies can be used for data analysis and prediction, specifically as follows:

(1) Service layer: Cloud computing service layer provides various types of computing resources and storage resources, each user can use these resources to train machine learning models and data mining. For example, users can deploy machine learning frameworks and algorithms on virtual servers, use cloud storage services to store and manage data sets, and use automated SaaS tools for data processing and visualization.

(2) Property layer: The cloud computing property layer provides a variety of performance metrics that can be used to analyze and optimize the performance of machine learning and data mining. For example, users can use monitoring indicators (rental cost, available

resources, quality of service) to monitor the resource usage and health status of the system in real time, log indicators to track the events and errors of the system, performance indicators to optimize the response time and throughput of the system, and cost indicators to control the use cost of cloud computing resources.

(3) Indicator layer: The cloud computing indicator layer provides physical resources and system indicators, which can be used for data analysis and prediction of machine learning and data mining. For example, users can use physical resource indicators to determine the computational resource requirements for the training and inference of machine learning models, use system indicators to analyze the characteristics and relationships of data sets, and use machine learning algorithms and data mining techniques to perform data analysis and prediction.

2. Fuzzy logic modeling

Based on the results of data analysis and prediction, the data distribution and load situation are modeled as a set of fuzzy logic rules, and these rules are stored in the fuzzy control rule base. The fuzzy model mainly uses particle swarm optimization algorithm to find the optimal value of its solution space, and then obtains the optimal identification effect of the model. Particle swarm optimization algorithm is a parallel optimization algorithm based on swarm intelligence. In the solution space, each particle learns from its own optimal value and the group optimal value at the same time, carries out dynamic update and group information sharing, and finds the optimal solution. For a dimensional solution space, there are two particles, and each particle iterates according to equation (1) :

$V_i^t = \omega V_i + c_1 r_1 (P_{ibest} - X_i) + c_2 r_2 (P_{gbest} - X_i)$

$X_i^t = X_i + V_i^t$

Where: i=1,2,...N; X_i, X_i^t is the position vector before and after the update; V_i, V_i^t is the velocity vector of the particle before and after updating;

ω is the inertia weight; r_1 and r_2 are interval [0,1] random numbers, c_1 and c_2 are cognitive and social factors; After several iterations, P_{ibest} is the individual optimal value found by the i particle in the solution space; P_{gbest} is the optimal value found by the group.

(1) Adaptive adjustment

According to the current load and data characteristics, adaptively adjust the data distribution and load balancing scheduling strategy, mainly using fuzzy control theory and adaptive control method to achieve. The following steps are usually required to realize the adaptive adjustment of data distribution and load balancing scheduling policy:

① Adjust the model indicators: for the established fuzzy model, the adjustment model includes the status and performance indicators of various resources. Use fuzzy control theory to model the state and performance indicators of each resource in the system into fuzzy variables.

② Fuzzy processing: the status and performance indicators of each resource in the system are fuzzy processing, and they are transformed into fuzzy variables. This can better deal with the dynamic changes and uncertainties of the system.

③ Optimize fuzzy controller: Optimize fuzzy controller, fuzzy variables as input, output control strategy. Fuzzy controller usually includes fuzzy rule base, fuzzy inference machine and fuzzy output module.

④ Parameter learning: Using adaptive control method for parameter learning, according to the current load and data characteristics, adjust the parameters of the fuzzy controller, in order to achieve the optimal scheduling strategy.

⑤ Scheduling strategy optimization: According to the output of the fuzzy controller, the data distribution and load balancing scheduling strategy optimization. For example, according to the current load situation, dynamic allocation of computing resources, adjust the data transmission policy, etc., to achieve the optimal load balancing and data distribution.

Through these steps, the data distribution and load balancing scheduling policies can be adjusted adaptively to achieve the optimal performance of the system.

3. Schedule execution

Execute the data distribution and load balancing scheduling policies after adaptive adjustment, and apply the data distribution and load balancing scheduling to each node in the cloud environment. The scheduling execution usually includes three levels: user, broker and infrastructure:

(1) User level: The user level refers to the end users of the system, who need to submit tasks and obtain the results of task execution. In scheduling execution, users need to submit tasks to the system, and specify the task priority, resource requirements and execution period and other information. The user also needs to monitor the execution of the task, obtain the execution result, and carry out follow-up processing. In the process of scheduling execution, users can interact with each other through the system interface or API interface, and modify the related parameters of the task as required.

(2) Broker level: The broker level refers to the scheduler in the scheduling system, which is responsible for scheduling and assigning tasks submitted by users. In the scheduling execution, the broker needs to formulate the optimal scheduling strategy according to the priority of the task, resource demand, execution deadline and current system status and other information, and assign the task to the appropriate computing resources. Brokers also need to monitor the execution of tasks and make dynamic adjustments based on real-time data and feedback information to ensure the stability and performance optimization of the system. In the scheduling execution process, the broker can use various algorithms and technologies, such as load balancing, task queuing, predictive control and other methods, to achieve the optimization of task scheduling.

(3) Infrastructure level: Infrastructure level refers to the computing resources and network environment in the scheduling system,

including servers, storage devices, network equipment, etc. In the scheduling execution, the infrastructure needs to call the corresponding computing resources and storage devices according to the tasks assigned by the broker, execute the tasks and return the results. The infrastructure also needs to monitor the execution of the tasks and make dynamic adjustments based on real-time data and feedback information to ensure the stability of the system and performance optimization. In the scheduling and execution process, the infrastructure can use various technologies and tools, such as containerization, virtualization, load balancing and other methods, to achieve the optimal utilization of computing resources.

Coordination and cooperation among these three levels to achieve the optimization of task scheduling is the core content of scheduling execution. At the same time, the scheduling execution needs to consider the stability of the system, performance optimization and user needs and other factors to ensure the effectiveness and reliability of the system.

4 Monitor feedback

Monitor and feed back the results of the scheduling execution so that adjustments and optimizations can be made to the system. Use real-time monitoring and feedback mechanism to achieve. Monitor the implementation and effect of the load balancing scheduling policy, including the distribution of requests and the load balancing of nodes. This can be done by monitoring the working status and log information of the load balancer or scheduler. Monitoring feedback can include visual charts and statistics of the load balancer scheduling, such as the distribution of requests, node load balancing metrics, etc., fed back to the adaptive adjustment model so that the effectiveness of the load balancer scheduling policy can be evaluated and adjusted.

III. Conclusion

The feature and innovation of the framework proposed in this paper is that the fuzzy logic theory is used to fuzzy the resource state information in the cloud environment, so that the system can reflect the actual situation more accurately, so as to achieve more precise load balancing. In addition, the data distribution is also considered. According to the size, type and access frequency of the data, the data is distributed to different server nodes to achieve the balance of data distribution. Finally, the framework combines the load balancing algorithm to dynamically adjust the task allocation strategy, so that the task can be better scheduled and executed in the cloud environment, and improve the performance and reliability of the system.

References:

[1] Rui Liu,Yongan Zhang,Jianfeng Wang. Review of Load Balancing Algorithms in Cloud Computing Environment [J]. Computer Engineering and Design,2019,40(6):1402-1408.

[2] Zhigang Li,Yan Wang,Xiaolei Xie. A Cloud Computing load Balancing Algorithm based on Fuzzy Logic [J]. Computer Science,2017,44(12):238-242.

[3] Na Liu. Overview of Data distribution Strategy in Cloud Computing environment [J]. Software Guide,2019,18(11):54-58.

[4] Li Ma, Yanan Zhong,Zhijun Hu. Research on Data Distribution Algorithm based on Spark [J]. Exploration of Computer Science and Technolo gy,2018,12(7):1061-1070.

[5]Santosh Kummar Maurya; Suraj Malik; Neeraj Kumar.Virtual machine tree task scheduling for load balancing in cloud computing[J].Indonesian Journal of Electrical Engineering and Computer Science,2023,Vol.30(1):388-393.

About the author: Li Dongmei, female, born in Jinzhong, Shanxi Province, Master candidate, Associate Professor, Henan Open University. Her research interests include cloud computing, machine learning, software technology, etc.

Fund Project: Youth Backbone Teacher Training Program of Higher Vocational Schools in Henan Province in 2020. 2023 School-level education and teaching reform research and practice project of Henan Open University, “Research on the Construction of teaching System of” Coordinated Development and Multiple Interaction “between Higher Vocational Education and higher education Continuing Education”, project number: 2023JGXMZ004.