Research on resource allocation strategy of data center based on cloud computing

Shanshan Ji*
School of Tianjin University of Technology, Tianjin, China

*Corresponding author e-mail: jishanshan@tjut.edu.cn

Abstract. With the development of Internet and information technology, cloud computing has attracted extensive attention from industry and academia. The large scale of resources, concurrent execution of multiple tasks and dynamic changes of application resource requests make the resource allocation of data center face severe challenges. To solve the problem of low balance of traditional resource allocation, this paper focuses on the resource allocation optimization of data center, and proposes the resource allocation strategy of data center based on cloud computing, so as to complete the effective resource allocation and assignment. This paper also verifies the designed resource allocation method through example research. The research shows that the distribution balance degree of resource allocation strategy based on cloud computing is significantly higher than the control group, which proves that the designed resource allocation strategy can solve the problem of low balance of traditional resource allocation.

1. Introduction
With the increasing popularity of Internet technology, Internet of Things technology, 5G communication technology and intelligent terminal equipment, diversified data shows geometric growth, and the resource allocation of data center has been paid attention to and concerned in many aspects. As a central organization for data statistics, analysis and processing, the data center realizes the organic integration of a mass of data resources through a special network virtual dedicated line. The data center is the basis of cloud computing services. People's requirements for data center are increasing, and the users of data center are also gradually increasing. Due to the different load of data center resources in different periods of time, the uneven resource allocation of data center often occurs. Facing many users, the data center has too much load, large data resource base and complex types, which brings new problems and challenges to the resource allocation of data center [1], and the resource allocation strategy is an important means to realize the balanced resource allocation.

Since the concept of cloud computing was put forward in 2006, it has led the third technological revolution of the Internet. Cloud computing has become the focus of the transformation and strategic layout of major information technology enterprises around the world. Compared with the traditional resource allocation, the resource allocation strategy based on cloud computing needs to consider the problem of balance. Because the traditional resource allocation method of data center has the phenomenon of low balance, they can't realize balanced resource allocation, resulting in low resource allocation efficiency. Therefore, it is necessary to optimize the resource allocation strategy to meet the
2021 2nd International Conference on Applied Physics and Computing (ICAPC 2021)  
Journal of Physics: Conference Series 2083 (2021) 042085  
doi:10.1088/1742-6596/2083/4/042085

personalized needs of users in many ways. In the past, data center resources focused on the input quantity of users, and there were still some deficiencies in the balanced allocation of resource channels for clients. The resource allocation strategy of data center based on cloud computing can realize load balancing through distributed computing, reasonably allocate resources, and then improve the balance degree. To solve the problem of resource allocation as soon as possible and make the resource allocation of data center more balanced and reasonable, this paper designs the resource allocation strategy of data center based on cloud computing, which is committed to greatly improving the resource allocation balance of data center through cloud computing.

2. Cloud Computing
Cloud computing not only represents the innovation and optimization of science and technology, but also represents a new concept and service method. In the cloud environment, cloud users don’t need to be bound and restricted by time and place. They can use terminal devices anytime and anywhere to directly obtain the information and resources they want to query from cloud services through the cloud network [2]. Through cloud computing, data processing can be completed in only a few seconds. The application of cloud computing in the data center can effectively improve the computing and analysis ability of data center, which is of great significance to the resource allocation of data center.

Cloud computing technology is a data technology means widely used in various fields. At the same time, it is also a computing platform environment characterized by distributed structure developed in recent years. In cloud computing technology, data center is not only the fundamental support and core, but also the main location for centralized data storage and management. Cloud computing is characterized by low cost and high efficiency. Generally, cloud computing platform includes four basic hierarchies: upper computer display layer, middle layer, analysis and processing layer and server terminal layer. Through the practical application of cloud computing, it can effectively improve the interconnection between various data, information and other resources and host computer, and realize the efficient utilization of various virtual resources [3]. How to effectively allocate resources has become an important problem to be solved in cloud computing, and it is also a hot topic in cloud computing research.

3. Resource allocation problem
As the data center is becoming large-scale and globalized, it is difficult to allocate and manage data center resources. On the one hand, the resource consumption of data center increases sharply, and the resource utilization rate is generally low, resulting in high operating costs of data center [4]; On the other hand, the resource allocation of data center is often unbalanced. The unbalanced load leads to excessive virtual machine migration, which consumes the network resources, and the unbalanced use of resources in the server leads to the waste of resources [5].

4. Resource allocation model construction based on cloud computing

4.1. Resource allocation objective function
In this paper, the first step of constructing the resource allocation strategy of data center is to propose the objective function of resource allocation. Considering that in the data center, each node performs different resource allocation tasks and cannot achieve balanced allocation, and some network nodes are allocated too many resource tasks, which will lead to excessive node energy consumption and failure of allocation tasks or node functions [6], in order to effectively prevent this phenomenon, a distribution of node delay minimum objective function is:

$$Z = \sum_{i=1}^{H} B_i + \sum TS_{ij}$$

Where \( i \) is the number of resource allocation; \( H \) is the resource set of data center; \( T \) is resource allocation time of data center; \( S \) is the signal-to-noise ratio of data center resource receiver.
4.2. Resource allocation parameters

According to the objective function of resource allocation in the data center, since cloud computing requires few estimated parameters, this paper makes a maximum posteriori hypothesis based on the load balancing assumption of cloud computing to make the resource attributes independent of each other. Set the maximum posteriori hypothesis and objective function of resource allocation attributes as follows:

\[
M = \frac{p(c_i)}{p(x)} \prod_{j=1}^{n} p(x_j|c_i) Z
\]

\[
N = \eta \ln B = \sum_{\nu} \{\eta \log p(v|n(v))\}
\]

Where \( p \) is the number of servers in the data center; \( c_i \) is the maximum resource capacity of servers under the resource allocation times of data center; \( x \) is the class with the highest posterior probability of resource allocation; \( n \) is the amount of resources wasted during resource allocation; \( j \) is the number of virtual machines; \( Z \) is the minimum objective function; \( \eta \) is a parameter of resource allocation structure characteristics; \( B \) is the intermediate vector; \( v \) is the auxiliary vector. The resource allocation parameters are calibrated based on the representation vector with high discrimination, and the resource allocation balance degree is improved to a certain extent by integrating the allocation balance factors into the objective function of the representation vector [7].

4.3. Resource allocation fitness ranking

After the calibration of resource allocation parameters, the fitness of the dynamic resource allocation is calculated, and the dynamic allocation of data resources with different priorities is realized by ordering the fitness of dynamic resource allocation in the data center. Cloud computing is used to divide the fitness level of dynamic resource allocation, so as to avoid conflicts and low priority queue hunger in the process of dynamic resource allocation. Suppose that calculation expression of the fitness of dynamic allocation of data center resources is:

\[
W = \frac{i}{\sum_{i=1}^{n} i}
\]

According to the above formula, the fitness of dynamic resource allocation of data center can be obtained. On this basis, the data center resource dynamic allocation fitness ranking is carried out based on the cloud computing allocation fitness ranking mechanism. Combined with the sorting mechanism, the fitness of dynamic resource allocation of data center is used as the selection operator in this allocation. While preserving the diversity of dynamic allocation of data center resources by means of mutation evolution, the similarity between dynamic allocation of resources is distinguished by sorting the fitness of dynamic allocation of data center resources.

4.4. Resource allocation implementation

Build a complete resource allocation method of data center based on cloud computing through the fitness ranking of dynamic allocation of data center resources, find the optimal solution based on cloud computing, and realize the balanced resource allocation of data center [8], as shown in Figure 1.
The objective function $\omega = (P_{min}, A, O, Z, l, f, D, P_{max})$ represents the resource allocation of data center, where $O$ is the resource allocation path; $l$ is the average connectivity of resource allocation network; $f$ is the communication resource quantity on the resource allocation paths; $D$ is the congestion degree of data center nodes. Based on the objective function of resource allocation based on cloud computing, through the proportional fairness algorithm to ensure the balance of resource allocation, sets the expression for the boundary function $\partial$, and obtain the complete resource allocation expression based on cloud computing:

$$\partial = e_1 p_1(t) + e_2 p_2(t) + \cdots + e_k p_k(t)$$

$$\rho_i = \begin{cases} 
\text{m. t max}(\omega^*) \\
\text{s. t max}E(B) \\
\text{s. t 2 max}E(\gamma) \\
\text{s. t 3 P}(\omega) > \{(1 - u_1)u_2 - s\}/ \partial 
\end{cases}$$

Where $e_1$ is resource allocation transmission power; $p_1$ is the average speed of each data center resource combination; $t$ is the maximum cellular signal-to-noise ratio; $k$ is the number of resource nodes; $E$ is the characteristic of resource allocation; $\gamma$ is the analytical parameter of resource allocation behavior; $u_1$ is the probability of high quality resource allocation; $u_2$ is the probability of low quality resource allocation. The optimal solution with the maximum value of $\rho_i$ is to perform resource allocation.

5. Experimental study

5.1. Experimental design

In order to verify the rationality and practicability of the designed resource allocation strategy, through the method of experimental study, this paper takes a data center as the experimental object, we randomly selects two real data sets (Tyhusge data set and Cfisnja data set) as the data center resources, selects a circular area with a side length of 10 km, and sets 200 communication nodes. The distribution balance degree of resource allocation by cloud computing method and traditional method is tested by Kerterley software. The former is recorded as the experimental group and the latter as the control group. Set the
allocation times to 10 times, and take every 2 times as a recording node to record the experimental results. The higher the degree of distribution balance is, the higher the rationality of the distribution strategy is.

5.2. Experimental results
According to Table 1, in the 10th experiment, the highest value of distribution balance degree in the experimental group is 1.57, in the 4th experiment, the highest value of distribution balance degree in the control group is 0.78, and the experimental group is 0.79 higher than the control group; In the 2th experiment, the lowest value of distribution balance degree in the experimental group is 1.31. In the 8th experiment, the lowest value of distribution balance degree in the control group is 0.55, and the experimental group is 0.76 higher than the control group. Therefore, the distribution balance degree of the resource allocation strategy designed in this paper is higher than the control group, which can realize the balanced distribution of data center resources and has practical application value.

| frequency | experimental group | control group |
|-----------|--------------------|---------------|
| 2         | 1.31               | 0.63          |
| 4         | 1.40               | 0.78          |
| 6         | 1.36               | 0.60          |
| 8         | 1.35               | 0.55          |
| 10        | 1.57               | 0.65          |

6. Conclusion
This research makes cloud computing from theory to practice. Based on cloud computing, load balancing and resource allocation can be solved. It is no longer a simple concept at first. Various technologies of cloud computing have made breakthrough progress. This paper designs the resource allocation strategy of data center based on cloud computing, constructs the objective function, verifies the practicability and rationality of the designed resource allocation strategy by means of example analysis, and proves that it can solve the defects existing in the traditional resource allocation, prevent the waste of resources caused by unreasonable load allocation, and realize the balanced allocation of data center resources.

Although the experimental analysis is carried out in this paper, there are still some shortcomings. The accuracy of the experimental analysis results is not tested. In the future, we can supplement the research in this regard, deeply study the resource allocation and optimization of data center, and provide constructive suggestions to improve the quality of resource allocation.

References
[1] Luo, P. & X. Wang. Research on the resource allocation algorithm based on forecasting in mobile cloud computing [J]. Journal of Intelligent & Fuzzy Systems, 2018, 35(2): 1315-1324.
[2] Geelan, J. Twenty-one experts define cloud computing [J]. Cloud Computing Journal, 2009(4): 1-5.
[3] Mala, K. & S. Sarbjeet. A review of metaheuristic scheduling techniques in cloud computing[J]. Egyptian Informatics Journal, 2015, 16(3): 275-295.
[4] Al-Dulaimy, A., W. Itani, Zekri A, et al. Power management in virtualized data centers: state of the art [J]. Journal of Cloud Computing, 2016, 5(1): 6.
[5] Wei, W., Y Liu, & W Yang. A fast approximation algorithm for the general resource placement problem in cloud computing platform [J]. Journal of Computer Reserch and Development, 2017, 9(3): 17-33.
[6] Gawali, M.B. & S.K. Shinde. Task scheduling and resource allocation in cloud computing using a heuristic approach [J]. Journal of Cloud Computing, 2018, 7(1): 1-16.
[7] Vicentini, C., S. Altair, V. Eduardo, et al. SDN-based and multitenant-aware resource
provisioning mechanism for cloud-based big data streaming [J]. Journal of Network and Computer Applications, 2018.

[8] Cheng, M.Y. & Prayogo. Symbiotic Organisms Search: A new metaheuristic optimization algorithm [J]. Pergamon, 2014(139): 98-112.