

Research Progress of Cloud Computing Task Scheduling Technology

Xiuhuan Zang^{1, a)}, Jin Sun^{2, b)}, Jiantao Zhao^{2, c)} and Wenjing Zeng^{1, d)}

¹ *State Grid Economic And Technological Research Institute Co.LTD, Beijing, 102209, China.*

² *North China Electric Power University, Beijing, 102206, China.*

^{a)} zangxiuhuan@chinasperi.sgcc.com.cn

^{b)} sunjin@eplian.com

^{c)} zhaojiantao@eplian.com

^{d)} zengwenjing@chinasperi.sgcc.com.cn

Abstract. Cloud computing is a distributed computing model that enables developers to automatically deploy applications during task allocation and storage allocation. Cloud computing is the sharing of a virtual computer resource pool and the storage, storage and computing power of the device. In cloud computing environment, task scheduling is a key problem, which needs to consider various factors that restrict user tasks, and is responsible for selecting the most appropriate cloud computing resources for user tasks. Task scheduling algorithm is a NP complete problem which plays an important role in cloud computing. This paper studies the application of job scheduling algorithm in cloud computing environment, compares and analyzes various classical algorithms and optimization algorithms, and summarizes the advantages and disadvantages of each algorithm.

Keywords: Cloud computing, Task scheduling, Technology

1. INTRODUCTION

1.1 Research Background of Cloud Computing

With the rapid development of processing and storage technology and the huge success of Internet popularization, computing resources are cheaper, stronger and more common than ever before. This technology trend enables a new computing model called cloud computing to be realized. In this mode, resources are provided as universal tools, and users can rent and release them on the Internet on demand. Cloud computing relies on the use and sharing of virtual and physical pool of resources, not local or personal hardware systems and software systems. The cloud computing task is only a small piece of work and should be executed in a specific time. The cloud task scheduler gets information about the state of available resources from the cloud information service manager, so that it can be assigned to a specific task to complete the task.

1.2 The Basic Problems of Cloud Computing Tasks

Cloud computing provides a flexible execution environment for multi cloud users and provides a variety of granularity services for the specified quality of service level. The quality of service is the collective effort of service performance, which determines the satisfaction of the user to the service. The quality of service is based on qualitative measures such as completion time, waiting time, execution price, packet loss rate, throughput, and reliability.

Task scheduling is an important problem that affects the performance of cloud computing environment. Task scheduling problem is a NP complete problem. The ultimate goal of task scheduling is to maximize the utilization of resources and minimize the waiting time of the task. Dispatchers should schedule tasks to improve the quality of service while maintaining the balance between productivity and fairness. The tremendous growth of virtualization and cloud computing technology reflects the increasing need for virtual machine services.

The key question is how to allocate users' tasks efficiently and reasonably to different resources according to the service quality requirements of cloud computing centers and users. Various scheduling algorithms have been applied to various data workloads and measured with different performance metrics to evaluate performance.

In addition to the task scheduling of traditional systems, task scheduling needs to meet the following new features in cloud computing environment:

(1) Diversity and preference

Different users have different needs for cloud computing resources, and each user often has their own preferences. That is, the task scheduling in the cloud computing environment also needs to meet the diversity and preference. Therefore, cloud computing needs to consider the user service quality requirements more complex, so as to fully meet the needs of different users, to the greatest extent possible to protect the quality of customer service. These problems can be solved by some solution to combinatorial optimization, such as heuristics, workflow-based methods, and so on.

(2) Maximum profitability

The task scheduling strategy in the traditional system generally only considers how to satisfy user service quality requirements to the maximum, but seldom considers how the service provider should obtain the maximum benefit. When scheduling cloud computing tasks, various cloud computing resources need to be effectively and reasonably utilized. We also need to ensure that cloud computing service providers receive the maximum benefit, which is their maximum profitability. Therefore, we can use the queue model approach to solve the task scheduling problem in the cloud computing environment.

2. ANALYSIS OF THE PRESENT SITUATION

The current classical algorithms of cloud computing have been quite a lot, and have been widely applied to the actual construction. However, most classical algorithms are still not perfect, there are various flaws, and there is still room for optimization. Some scholars propose a new algorithm model and optimize the current algorithm according to the current actual demand, and some progress has been made. This article gives an example of a newer algorithm that has its own advantages and is worth learning and researching.

2.1 Classical Algorithms

2.1.1 First Come First Service Method (FCFS)

The simplest way of FCFS is to accumulate tasks and queue them when resources are busy, and then allocate resources to tasks according to their arrival time. The round robin method uses the same FCFS technology to schedule tasks, but allocates resources to each fixed time slot task. Then the task into the waiting state, waiting in line for the next implementation of the opportunity.

2.1.2 Min-Min Method

The Min-Min method picks the smallest task out of all the available tasks and assigns them to the machine for the task's minimum completion time, regardless of the load of the machine on which the task is to be completed.

2.1.3 Max-Min Method

Max-Min method is to choose the longest task, first in the fastest machine arrangement. Smaller tasks need to wait longer ^[1]. It increases the completion time and system throughput, but does not consider load balancing. Improved cost-based algorithm, according to the cost of different resources to arrange tasks, the cost varies with the complexity of the task ^[2].

2.1.4 Simulated Annealing Method

The simulated annealing method is an iterative method that is similar to a genetic algorithm, starting with a single solution (map) chosen from a random distribution. Repeat again to get a better version. After the mutation, the new completion time is analyzed and if it is found to be lower than the previous one, the old working time will be replaced by the new one.

Other classic algorithms include Round Robin Scheduling Algorithm (RR), Resource Scheduling Algorithm (RSA), and the most suitable task scheduling algorithm (MFTF). These classic algorithms have greatly enriched the choice of cloud service providers and further promoted the development of cloud computing.

2.2 Optimization Algorithms

2.2.1 An Improved SJF Scheduling Algorithm in Cloud Computing Environment

Literature^[3] proposed an improved shortest job priority algorithm (MSJF). The algorithm minimizes the completion time of the last task and minimizes the average response time. MSJF has two functions, one is to calculate the average length of the task, and the other is the load balancing between virtual machines. An important advantage of MSJF is sending the longest task to the fastest machine.

2.2.2 Ant Colony Algorithm Based Cloud Computing Multi - objective Optimization Scheduling Algorithm

Literature^[4] proposed a multi-objective optimization scheduling method based on resource cost model (PBACO). The method takes the maximum production time and the user budget cost as the constraints of the optimization problem, and realizes the multi-objective optimization of performance and cost. An improved ant colony algorithm is proposed to solve this problem. Two constraint functions are used to evaluate and provide feedback on performance and budget costs. The two constraint functions enable the algorithm to adjust the quality of the solution on the basis of feedback in time to achieve the optimal solution.

2.2.3 An Improved Artificial Bee Colony Algorithm

Literature^[5] proposed an improved artificial bee colony algorithm. The algorithm first proposed a quality of service evaluation model (QoE) based on the Fuzzy Analytic Hierarchy Process (FAHP) for service composition (SC). Then, an improved artificial bee colony optimization algorithm (IABC) based on QoE SC problem is proposed. By introducing the current global optimal solution to improve the reconnaissance bee update mechanism, to speed up the convergence rate, and ultimately improve the quality of solution. Compared with the original ABC, PSO and DE, IABC has better performance on QoE-based SC problems.

2.2.4 Workflow Application Scheduling Algorithm Based on Key Path

Literature^[6] proposed a workflow-based application scheduling algorithm based on critical path. The algorithm is a workflow scheduling algorithm based on the critical path in cloud computing environment. Its main purpose is to shorten the duration of the workflow and reduce the overall execution costs by finding critical path tasks. Scheduling workflow application tasks on resources to reduce the total execution cost of workflow applications.

2.2.5 An Effective Algorithm of Workflow Scheduling in Cloud Computing Environment

Literature^[7] proposed an efficient workflow scheduling algorithm (EWSA). Is to maximize the use of resources, while meeting the application deadline, you can handle a large number of applications. The goal of this algorithm is to dynamically estimate the execution time of all tasks. The algorithm also creates a suitable virtual machine with minimal resources so that the entire application can execute within the deadline.

2.2.6 Multi-Agent Genetic Algorithm

Literature^[8] proposed a multi-agent genetic algorithm (MAGA). The algorithm is a hybrid algorithm of genetic algorithm, its performance is far superior to the traditional genetic algorithm. The algorithm solves the problem of cloud computing load balancing by designing the load balancing model based on virtualized resource management and comparing with various genetic algorithms.

3. PROBLEMS IN CLOUD COMPUTING TASK SCHEDULING TECHNOLOGY

3.1 Classic Algorithm Analysis

The proposed classic algorithms are earlier, are relatively authoritative algorithm, has its own advantages. However, when applied to the cloud environment, these algorithms will have their own shortcomings. The advantages and disadvantages of the classic algorithm are shown in Table 1 below:

TABLE 1. The advantages and disadvantages of the classic algorithm..

Algorithm name	advantages	disadvantages
First come first service algorithm	Easy to implement	There are no other scheduling criteria
Min-Min method	Better completion time than other algorithms	The load balance is poor and the QoS factor is not considered
Max- Min method		
Simulated annealing algorithm	1. Always trying to find a better solution 2. Larger solution space and better completion time	QoS factors and heterogeneous environments need to be considered

3.2 Advantages of Optimization Algorithm

It is these shortcomings of the classical algorithm, to some extent, promoted the birth of the optimization algorithm. Based on the classical algorithm, the optimization algorithm inherits the advantages and makes up for the shortcomings according to the requirements of the actual cloud environment. The optimization algorithm has its advantages over the classical algorithm. The following will introduce the above optimization algorithm where the specific advantages.

3.2.1 An Improved SJF Scheduling Algorithm in Cloud Computing Environment

- 1) It not only focuses on the completion time of the task, but also considers the completion time of all the tasks^[9].
- 2) Send the longest task to the fastest machine.
- 3) Compared with SJF and FCFS, MSJF more effectively improve the response time and completion time.

3.2.2 Ant Colony Algorithm Based Cloud Computing Multi - objective Optimization Scheduling Algorithm

- 1) The PBACO resource cost model can be used to define the task requirements of a resource in detail. This model reflects the relationship between resource cost and task cost^[10].
- 2) Based on this model, a multi-objective optimization model is proposed. Its main goal is to optimize the scheduling of performance and user costs.
- 3) An improved ant colony algorithm is proposed to solve and optimize the problem^[11]. To prevent the ant colony algorithm from falling into the local optimal solution, the method uses performance and budget constraint functions to evaluate costs and provide feedback on the quality of the solution. Then adjust the quality of the solution based on the results of the evaluation and feedback.

3.2.3 An Improved Artificial Bee Colony Algorithm

By introducing the current global optimal solution, the algorithm improves the updating mechanism of reconnaissance bees, speeds up the convergence speed, and finally improves the quality of the solution^[12]. Compared with the original ABC, PSO, and DE, IABC has better performance on the QoE based SC problem.

3.2.4 Workflow Application Scheduling Algorithm Based on Key Path

Compared with Min-Min and Max-Min scheduling algorithms, this algorithm minimizes the execution cost of workflow applications and fulfills all the tasks of workflow applications within the deadline. Compared with other existing heuristic scheduling algorithms, it can reduce costs and adapt to workflow applications of different sizes and types.

3.2.5 An Effective Algorithm of Workflow Scheduling in Cloud Computing Environment

- 1) A clever strategy to estimate the execution time of a task.
- 2) The algorithm dynamically creates and schedules effective mechanisms for virtual machines.
- 3) The algorithm minimizes the number of virtual machines required to execute a workflow.

3.2.6 Multi-Agent Genetic Algorithm

The algorithm can achieve better load balancing performance, greatly improve the convergence time and optimize the results, especially in dealing with large-scale, high-dimensional, complex and dynamic optimization problems, than the traditional genetic algorithm algorithm has obvious advantages. Moreover, it is more suitable than the general genetic algorithm to deal with the problem of high-dimensional function optimization^[13].

With the continuous development of cloud computing, the demand for cloud resource scheduling is also getting higher and higher. Task scheduling is an important indicator of cloud computing performance. Some traditional scheduling algorithms are far from satisfying the pursuit of cloud computing performance. Therefore, some improved scheduling algorithms are introduced in this paper in order to satisfy the people's pursuit. However, these algorithms still have some shortcomings and hope to be improved in the future research.

4. CONCLUSION

This article summarizes a number of related papers, providing the current progress in cloud computing task scheduling and resource allocation techniques. Most of the authors in the literature under study focus on improving and calculating the quality of service. All of these algorithms are based on the previous algorithms and are obviously better than the original algorithm in terms of performance, efficiency and service quality. Improve and further meet the needs of enterprises and users.

REFERENCES

1. Huang, Y., Huang, P.C.: Task Mapping and Routing Scheme Based On Minimization of Completion Time in Cloud Services. Dongguan University of technology, 2016. 121-126.
2. Neeta, P., Deepak, A.: A Review - different Scheduling Algorithms in Cloud Computing Environment. Intelligent Systems and Control (ISCO), 2017 11th International Conference on: IEEE, 2017.
3. Mokhtar, A.A., Atyaf, D., Asma, A., Al-Hashmi, A., Basit, D.Suresha.: An Improved SJF Scheduling Algorithm in Cloud Computing Environment. 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT): IEEE, 2016.
4. Zuo, L.Y., Shu, L., Dong, S.B., Zhu, C.S., Takahiro H.: A Multi-Objective Optimization Scheduling Method Based on the Ant Colony Algorithm in Cloud Computing. IEEE Journals & Magazines, 2015.
5. Xu, B., Qi, J., Wang, K., Wang, Y., Hu, X.X., Sun, Y.F.: An Improved Artificial Bee Colony Algorithm for Cloud Computing Service Composition. 2015 11th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QSHINE), 2015.
6. Jailalita., Sarbjeet, S., Maitreyee, D.: Critical Path Based Scheduling Algorithm for Workflow Applications in Cloud Computing. 2016 International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Spring), 2016.
7. Mainak, A., Tarachand, A.: Efficient Algorithm for Workflow Scheduling in Cloud Computing Environment. 2016 Ninth International Conference on Contemporary Computing (IC3), 2016.
8. Zhu, K., Song, H.G., Liu, L.J., Gao, J.Z., Cheng, G.J.: Hybrid Genetic Algorithm for Cloud Computing Applications. 2011 IEEE Asia-Pacific Services Computing Conference, 2011, 182 – 187.
9. Singh., Umang, S., Ayushi.: CloudSim Simulator Used for Load balancing in Cloud Computing. International Journal of Emerging Technology and Advanced Engineering, (2016) 246-255.
10. Wan, F.J., Wang, X.P., Qin, S.L.: The Resource Cost Model of Activity-Based Costing: Theory and case study. Journal of hubei university of economics, 2012-7(1).
11. Bai, Y.P., Song, J.: An Improved Ant Colony Algorithm and Its Application in the Shortest Path Problem. North Central University, 2013. 4-7.
12. Wang, B.: Improved Artificial Bee Colony Algorithm Based on Local Optimal Solution. Beijing Institute of Technology, 2013. 1-5.
13. Zheng, Z.N., Wang, R., Zhong, H., Zhang, X.J.: An Approach for Cloud Resource Scheduling based on Parallel Genetic Algorithm. Computer Research and Development (ICCRD), 2011 3rd International Conference on, vol.2, pp.444-447, 11-13 March 2011.