Journal of Applied Science and Engineering Innovation, Vol.5 No.1, 2018, pp. 9-12
ISSN (Print): 2331-9062
ISSN (Online): 2331-9070

Research on Task Scheduling Strategy Based on Cloud Computing Environment

Jun Nie

Guangdong University of Science &Technology , Software department, Dongguan , Guangdong Province, 523083, China.

Abstract: This paper studies the task scheduling in cloud computing, analyzes the programming model framework of cloud computing, and proposes a hybrid scheduling algorithm based on genetic algorithm and ant colony algorithm. This algorithm makes full use of the rapid random global search ability of genetic algorithm , But also overcome the problem of the initial pheromone lacking in ant colony resulting in slow solution. The simulation results show that this algorithm has good performance and improves the efficiency of task scheduling in cloud computing environment. It is an effective task scheduling algorithm in cloud computing environment.

Keywords cloud computing; task scheduling; genetic algorithm; ant colony algorithm

INTRODUCTION

Cloud computing system for the general public to provide all kinds of computing or storage services, the server is very large, the user is also very large scale, the task needs to be processed and the size of the data is massive. How to allocate and manage the available resources in the cloud computing system reasonably and efficiently, and implement reasonable and efficient execution processing of application task scheduling requests for all kinds of users, so that all kinds of application task requests can obtain a better scheduling result and become cloud A research hot spot in computing technology. In this paper, the efficiency of task scheduling in cloud computing environment, in order to get the total task execution time and the average task execution time are completed within a short time, this paper presents a hybrid genetic algorithm and ant colony algorithm hybrid scheduling algorithm to Maximize the efficiency of the cloud computing environment, and passed the simulation experiments to verify its good performance and improve the efficiency of task scheduling in the cloud computing environment.

CLOUD COMPUTING PROGRAMMING MODEL ANALYSIS

Google, which is mostly used in the current cloud computing environment, proposes the MapReduce distributed computing mode, which automatically divides a large task into multiple subtasks, and realizes the task scheduling and allocation in a large scale computing node by two steps of Map and Reduce. Most of the information technology vendors proposed cloud programming model used in the development are based on the idea of MapReduce programming tools, it is particularly suitable for generating and processing large-scale data sets, the implementation process shown in Figure 1.

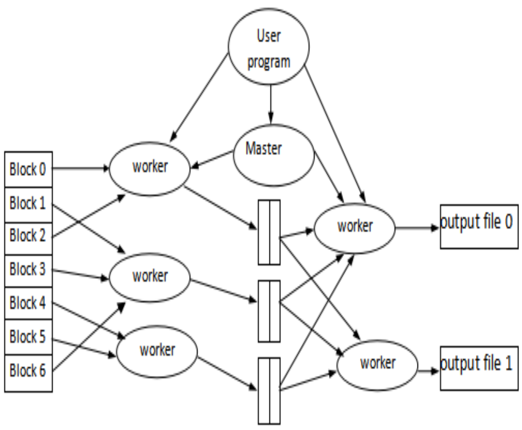

Figure 1 MapReduce implementation process

As can be seen from Figure 1, MapReduce executes seven processes in two phases:

1) Map phase: a large task is first divided into M blocks by the MapReduce library in the user program, each block size is generally 16 ~ 64MB, and then assigned to multiple workers in parallel execution, the output of the middle file.

2)Reduce stage: the Map stage after the results of the summary analysis, the output R files.

Under the MapReduce programming model, scheduling a large number of subtasks is a complex issue. Each task is subdivided into Map phase M blocks and Reduce phase R blocks. M and R should be much larger than the number of worker machines, and each worker performs many different tasks. How to schedule many tasks, it is necessary to make the resources load balancing, but also consider the

Corresponding Author: Jun Nie, Guangdong University of Science &Technology, Dongguang 523083, China.

response time. This is the main issue to be considered in this article. According to the characteristics of scheduling model in cloud computing environment, this paper proposes a new scheduling algorithm suitable for large-scale parallel processing system. This algorithm makes full use of the colony-based and fast searching ability of genetic algorithm. At the same time, by utilizing the advantages of ant colony algorithm such as positive feedback and parallelism, the two algorithms are combined to establish a better allocation scheduling strategy. Through the simulation experiments, it is verified that the task scheduling can reduce the total task completion time, achieve the load balancing and improve the efficiency of task scheduling.

GENETIC ALGORITHM MECHANISM AND MODEL

Genetic Algorithm (GA) was proposed by Holland in 1975 by the theory of biological evolution. Parallelism and global solution space search are the two most prominent features of GA. Aiming at the problem of task scheduling in cloud computing environment, task scheduling with genetic algorithm can get better results.

Chromosome Coding and Decoding

Chromosome coding there are many ways, this article uses resource-task indirect coding. The length of the chromosome is the number of subtasks, and the value of each gene in the chromosome is the resource number allocated to the resource by the subtask corresponding to the position number.

Suppose there are n independent tasks, m available workers, chromosomes of length n + m, the first n integers represent tasks, and n + 1 to n + m represent available resources. The rule is defined as: the nearest task in the queue is assigned to the right resource execution on the right.

There are T tasks, W worker resources, and the number of subtasks to which the tth task is divided is: taskNum (i). The total number of subtasks is subTaskNum:

$$\mathrm{subTaskNum} = \sum_{i=1}^{T} \mathrm{taskNum}(t)$$

If there are 3 large tasks (task), 3 worker resources, each task is divided into several subTask: If task1 is divided into t1.1, t1.2, t1.3 three subTask; task2 is divided into t2 .1, t2.2; task3 is divided into t3.1, t3.2, t3.3, t3.3, t3.4, t3.5, a total of 10 subTask, and then these subTask are numbered, we Using a simple kind of coding is: followed by sub-order coding. Ie the number of the j th subTask in the ith task is m:

$$m = \sum_{k=1}^{i=1} \mathrm{taskNum}(k) + j$$

The length of the chromosome is 10, and the value of each gene is 1 to 3. If one chromosome is {2,3,1,3,2,2,1,2,3,2}, then this chromosome represents The first subTask executes on the second worker, the second subTask executes on the third worker, ..., and the 10th subTask executes on the second worker. The subTask distribution on the worker is obtained by decoding the chromosome. Generate multiple sets of subTask sequences with resource numbers. For example, the above chromosome is decoded as W1 = {3,7}, W2 = {1,5,6,8,10}, W3 = {2,4,9}

Through the decoded sequence and the ETC matrix (ETC [I, j] indicates the time required for the ith task to execute on the jth resource), it is possible to calculate the time it takes for each resource to execute all the tasks on that resource, then all The total task time function is:

$$F_1(x) = \max_{w=1}^{\mathrm{worker}} \sum_{i=1}^{n} \mathrm{worker}(w,i)$$

The completion time of task t:

$\mathrm{tasktime}(t) = \max_{i=1}^{\mathrm{tasknum}} \sum_{j=1}^{k} w(j,i)$, where tasknum represents the number of subtasks that task t is divided into. The average time spent on tasks is: : $F_2(x) = \sum_{i=1}^{TASK} \mathrm{tasktime}(t)/TASK$, where TASK represents the number of tasks.

Genetic Operation

1) fitness function

In the genetic algorithm, the fitness function is related to the convergence speed of the algorithm and the advantages and disadvantages of the solution. In the cloud computer task scheduling, it needs to consider not only the response time of each user, but also the comprehensive satisfaction of users. Therefore, Measured by the average task completion time. In this paper, the task of using the average time function as a genetic algorithm fitness function. which is $F_2(x) = \sum_{i=1}^{TASK} \mathrm{tasktime}(t)/TASK$.

2) Individual choice

Select a pair of chromosome substrings for crossover and mutation operation according to the fitness function value and calculate the selection probability of each individual in the population according to the following formula: $P(i) = 1 - \frac{f(i)}{\sum_{j=1}^{SCALE} f(j)}$, where SCALE is the population size.

3) crossover and mutation operation

Crossover is the most important search operator in genetic algorithms. It mimics the process of genetic recombination in nature and inherits the original good genes to the next generation of individuals and generates new individuals that contain more excellent gene structures. Variation can expand new search space, and population diversity can be maintained through mutation when local population convergence. In this algorithm, we use the sequential intersection method proposed by Davis. We first perform the conventional two-point crossover and then modify the tour routes to maintain the original relative access order.

Mutation method using reverse mutation.The crossover probability and mutation probability are constants c1 and c2, respectively, which are generated by random in practice. If r is less than the crossover probability, the crossover operation is performed, otherwise, the same mutation operation. In this way, comparing the application-degree function values of the generated individuals, only the adaptation values have been improved, otherwise the operations are invalid. After recursive iteration many times to generate several sets of optimization solutions for ant colony algorithm to prepare.

GENETIC ALGORITHM AND ANT COLONY ALGORITHM INTEGRATION

Ant colony optimization (ACO) is a population-based heuristic bionic evolution algorithm proposed by the Italian scholar M.Dorigo et al. By simulating the behavior of colony homing in nature. It is also a kind of global Optimization method, especially suitable for parallel computing, but also has a positive feedback mechanism to simulate the load situation. The disadvantage is the lack of early pheromone, resulting in the accumulation of pheromones occupy a longer period of time initial search.

Genetic Algorithm and Ant Colony Algorithm Integration

Genetic algorithm is inefficient in the latter part of the solution, which will easily lead to a large number of redundant iterations. Therefore, the connection between genetic algorithm and ant colony algorithm is a difficult problem. If the iteration number of genetic algorithm is set to a fixed value, it will be too early or too late Effect of the algorithm. This article uses a dynamic fusion strategy to ensure that the genetic algorithm and ant colony algorithm in the best timing fusion. Strategy is as follows:

1) Set the minimum number of genetic iterations Gmin and the maximum number of genetic iterations Gmax.

2) Statistics the evolution rate of the progeny population in the genetic iteration process and set the minimum evolution rate of the progeny population Gmin-impro-radio。

3) Within a set number of iterations, if continuous Gdie (, the evolution rate of the progeny population is less than Gmin-impro-radio, indicating that genetic algorithm optimization speed is low, so the genetic algorithm can be terminated, enter the ant colony algorithm.

Ant Colony Algorithm Pheromone Settings

This paper draws on the pheromone setting in MMAS (Max-Min Ant System) algorithm proposed by Belgian scholar Thomas Stutzle. The initial value of pheromone is set to be a pheromone constant given according to the scale of the problem solving and the pheromone value converted from the result of genetic algorithm.

1) Genetic algorithm and pheromone conversion

In this paper, we choose the best 15% of the population whose fitness is the best when the genetic algorithm is terminated as the set of genetic solutions, it is $S^{gene}_{15\%better}$. starting with 0, and adding a constant k to each solution. The pheromone update model is $\tau_j^{new} = \rho.\, \tau_j^{old} + \Delta\tau_j$, where (pheromone volatility coefficient).

2) Resource Selection Probability

At time t, the probability that a task is assigned to each resource can be expressed as:

$$P_j^k(t) = \begin{cases} \frac{[\tau_j(t)]^{\sigma}[\varphi]^{\beta}}{\Sigma_N[\tau_N(t)]^{\alpha}}[\varphi]^{\beta}, & j, N \in \text{available resource} \\ 0, \text{other} & \end{cases}$$

, Where is the visibility of resources, is the pheromone weight, is the visibility of resources visibility.

ALGORITHM SIMULATION RESULTS

In this paper, Gridsim network simulation tools to verify the advantages and disadvantages of the GA-ACO and GA, the initial parameters set in Table 1.

Table 1 Gridsim network simulation tools to verify the advantages and disadvantages of the GA-ACO and GA

Algorithm	Project	Value	Algorithm	Project	Value
GA	Population size	100	GA-ACO	Ant colony size	100
	Crossing rate	0.65		σ	3
	Variation rate	0.07		β	5
				ρ	0.2
				φ	0.7

The initial value of each path pheromone in genetic ant colony algorithm is set to be 70. The pheromone value of genetic algorithm to solve the conversion is calculated by adding 3 to the path, Gmin is 40, Gmax is 200. Figure 2 is a continuous graph of the simulation results of the two algorithms.

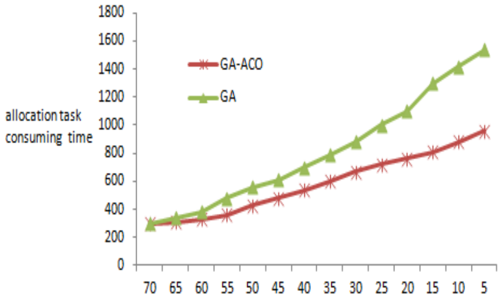

Figure 2 simulation results of the two algorithms

As can be seen from Figure 2, GA-ACO is more efficient than GA (Genetic Algorithm), especially when there are more nodes and less effective nodes, so this algorithm is more effective in cloud computing environment An efficient scheduling algorithm.

CONCLUSION

In this paper, by analyzing the programming model in cloud computing, a task scheduling algorithm based on genetic algorithm and ant colony algorithm is proposed. This algorithm not only sets an important criterion for task completion, but also takes the resource load balance as reference, The effective task height under cloud computing environment has obvious advantages over traditional genetic algorithms.

REFERENCES

Abdulrahman A A,Muhammad I S,Saleh B,et al.A distributed access control architectrure for cloud computing[J].IEEE software,2012,29(2):36-44.

Chen Danwei,Shao Ju,et al.MAH-ABE based prvacy access control in cloud computing[J].Acta electronica Sinica,2014,42(4):821-827.

Feng Chao-sheng,Qin Zhi-guang.Key Techniques of Access Control for Cloud Computing[J]. Acta Electronica Sinica,02(43):312-319(2015).

Ren K,Wang C,Wang Q.Security challenges for the public cloud[J].IEEE Internet comput.

Takabi H,Joshi J B D,Ahn G.Security and privacy chal-lenges in cloud computing environments[J].Security &Privacy,IEEE,2010,8(6):24-31.

Wang Xiao-wei,Zhao Yi-ming.A Task-role-based Access Control Model for Cloud Computing[J].Computer Engineering,38(24):9-13(2012).

Wang Yu-Ding,Yang jia-Hai.Survey on Access Control Technologies for Cloud Computing[J].Journal of Software,26(5):1129-1150(2015)ing,2012,16(1):69-73.

Xiong Jinbo,Yao Zhiqiang. A Composite Document Model and Its Access Control Scheme in Cloud Computing[J]. Journal of Xi'an Jiaotong University, 48(2): 25－31(2014)

Yu Nenghai,Hao Zhuo.,et al.Review of cloud computing security[J].Acta Electronica sinica,2012,41(2):371-381.