

Research on a Resource Scheduling Scheme of Cloud Compute Source Load Balancing Based on Genetic Algorithm

Yuan BaoLi

ShiJiaZhuang Vocational Technology Institute, ShiJiaZhuang, HeBei province , 050081, China

stevenybl@qq.com

Keywords: Genetic Algorithms, Cloud Computing, Resource Scheduling, Load Balancing

Abstract. The schedule of resources and tasks in cloud computing have an important impact on the overall performance and operating of cloud computing. To solve the prone uneven load node when the cloud performs massively parallel tasks, which led to the entire cloud computing system performance degradation and inefficient problem, this paper research on how genetic algorithm achieve load balancing in cloud computing on the basis of resource scheduling and genetic algorithm analysis cloud.

Introduction

Cloud computing is a new business model, the calculated mass distribution of tasks in the resource pool consisting of computer hardware, user-demand customized way to enjoy the services provided by cloud computing, such as computing power, storage space and information services. Cloud of huge commercial potential, is the biggest reason it is highly touted by the industry. However, due to the number of computers cloud computing platform is very large, complex composition, configuration differences greater resources to achieve such a large scale, heterogeneous, non-centralized resource scheduling, it is difficult to achieve load balancing of resources. The likely cause uneven load efficiency and lower system throughput, as well as the daily operation of system instability, seriously affecting the quality of service of cloud computing.

Cloud computing is customer-oriented, very frequent scheduling of resources, resource join or leave at any time, thus scheduling policy should not only consider the first dispatch, but also consider the re-deployment of resources after the failure, which requires scheduling algorithm has a strong flexible line and adaptability while being able to quickly respond to user requests. But big data load balancing scheduling is NP-complete problem [1], the general load balancing scheduling algorithms can not be solved, heuristics and often only for a particular problem can be satisfied with the solution, once the external conditions change, the solution the quality will rapidly decline, the performance is very unstable. How to achieve load balancing cloud computing resources scheduling, it is being more and more attention, and made a lot of new methods and ideas, but so far there is no good scheduling algorithms to meet cloud computing for load balancing and time the complexity of the requirements.

Cloud Computing

Cloud computing as distributed computing in one of its most basic concepts to calculate various tasks discretely distributed in a large number of computer resource pool configuration, make all kinds of applications can get the computing power according to their actual needs, as well as storage space different software services. Cloud computing is virtualization, utility computing, IAAS (infrastructure as a service), PAAS (platform as a service), SAAS (software as a service) concepts such as evolution and commercialization achieve mixed results [8]. Virtualization is the key technology to achieve cloud computing services, any service on the cloud, virtualization technology through virtual implementation. IAAS, PAAS and SAAS model is the mainstream of the current cloud services, they focus on different applications, but they need support for virtualization technology and cloud resources to solve the problem of load balancing scheduling. And finally through SLA (Service Level Agreement) to agreed customer service has promise and QoS [2].

The Overview of Load Balancing

Load balancing in many professional fields have applications, load balancing computer is accompanied by the emergence of the people's demand for computing power and clustering technology generated, are cluster system performance, and with the development of technology and the development of the cluster [9]. Load balancing algorithm scheduling mechanism, it is possible to solve the load imbalance cluster environment issues, and promote the rational allocation of resources and improve service quality.

The Genetic Algorithm (GA)

Genetic algorithms are inspired by Darwin's theory of natural selection and the natural process of biological evolution and the analog design of a computational model. It does this by using 8 simple coding techniques to represent a variety of complex structure, and a set of encoding represents a simple natural selection and survival of the fittest genetic manipulation to guide the study and determine the direction of the search. Typically the algorithm will be represented as a population group of binary string, wherein each of the binary string of an individual. The individual is a solution corresponding to the problem, but the population is a solution set of problems. Starting from an initial population of genetic algorithm, using the fitness selection strategy to select individuals in the population-based, choose the operation, then use the crossover and mutation to produce the next generation population. Thus mimic the evolution of life, the evolution of the population down from generation to generation until the termination condition to meet the expectations of the algorithm [3].

Selection, crossover and mutation is a genetic algorithm basic operations. Crossover is also known as recombinant means combining parent's information to generate new individuals, usually by binary-coded crossover implementation. Mutation is a small probability event, through the implementation of certain rules for selecting individual variation operation, the offspring of the rich diversity and prevent local optimum algorithm in order to obtain higher quality individuals. Crossover and mutation operation generates new individuals will form a new population.

There are two ways to use the genetic resources to implement load balancing algorithm to dispatch. One is based on the system because the task entry and exit frequently, recent generations of solution is often the assumption that the main areas in which the system load, genetic algorithm solution recently selected by generations of selection operation, the use of crossover and mutation operation generates a new difference of a larger solution. The new solution because of the recent no task transferred, and thus the load will be relatively low, so that the load is moved from the main load area of the system to other areas of the system load is low. Another solution is based on the utility, a good solution can be well-balanced system load, the utility of the solution will be higher, which is operating a selected probability of selection is also higher. And when a solution has been selected because often fail when it since the utility will reduce the probability of being selected as the utility Solutions decreases.

The Resource Scheduling Scheme Design of Cloud Computing Resource Load Balancing Based on Genetic Algorithm

Cloud computing environment task scheduling situation can be described as follows: The total number of resources is P, corresponding to the set $R = \{r_1, r_2, \dots, r_p\}$; total number of jobs submitted by the user is M, corresponding to the collection: $HJ_i, h > \dots, JM\}$; Let M jobs in total are divided into N tasks, corresponding to the set $T = \{T_1, T_2, \dots, T_n\}$, the first job is divided $J_m T_{Num}(J_m)$ task, then the M corresponding to the total number of job tasks:

$$N = T_{total}(M) = \sum_{m=1}^M T_{Num}(J_m)$$

Cloud computing task scheduling environment is to solve the efficient execution of N interdependent parallel tasks on a limited P resource issues while to fully meet the user in task completion time, bandwidth, reliability, and cost, etc. expectations [4].

Coding. For task scheduling problem, many existing encoding, commonly used can be roughly divided into two categories: direct and indirect coding. In contrast, direct coding is simple, but carry less amount of information on the indirect encoding requires certain information to be encoded conversion, can carry more information, more targeted. In order to improve the algorithm search capabilities, taking into account the division of the job under a cloud environment, and constraints related to each other sequence, the paper preclude the use of scheduling mix coding method. That chromosome is divided into two parts, the left part of the substring represents resources assignments, called distribution substring; the right part of the substring represented scheduling order, called the scheduling sub-string. The length of each chromosome set, namely 2 times the total number of tasks for the 2L.

Generate initial population. This article preclude the generation of the initial population using a random manner. In the process of generating the initial population, it needs to detect program constraints and exclude invalid program. Check whether the program effective method: Let the set Sf as completed tasks set, followed by checking on each resource is currently ready to perform the task Ti, Ti does not exist if the task dependencies or all dependent tasks are present in the set Sf in, Description Task Ti can be implemented, the implementation of Ti and put it into the set Sf. Repeat the above process, if the end all tasks to be executed, the program is an effective program, otherwise it is invalid program.

Fitness function. Fitness function used to determine the direction of the evolution of the genetic algorithm, used to assess the extent of the merits of each individual group. Select the fitness function is particularly important for the efficiency of genetic algorithms. In the cloud computing environment, task scheduling need to meet the diverse needs of users on the quality of service, while more traditional genetic algorithms to a single goal as the fitness function, does not apply to the task scheduler cloud computing environments. In the cloud computing environment, user satisfaction with the service quality of service QoS can be used to measure this standard. In this paper, the characteristics of cloud computing business model, select the job completion time, bandwidth, cost and reliability of these four objectives to quantify the degree of satisfaction of different users.

Choose. Select the operation is intended to promote good individual chromosomes with a higher probability of genetic to the next generation. This article preclude the use of roulette wheel selection method, set the size of the population is S, for a fitness for the fi individual chromosomes, the choice of probability Qi

$$Q_i = f_i / \sum_{j=1}^S f_j$$

Cross. Crossover operator of genetic algorithm is an important operator can search for new solutions in a very wide space. For the assignment and scheduling substring article preclude the use of two different cross-ways: allocation and scheduling substring sub-string for distribution, we use single-point cross-cross pattern. For scheduling sub-string, the paper and preclude the use of a similar cross-section matching PMX (Partial Mapped Crossover) operator.

Variation. Mutation in a certain extent, overcome the premature convergence of genetic algorithms can guarantee the diversity of the population of individuals. In this paper, allocation and scheduling substring sub-string with a different variation of ways: allocation and scheduling substring sub-string for distribution, the paper way preclude the use of random operations and that is randomly selected locus mutation, randomly selects it is like allelic replacement. For scheduling sub-string, the paper preclude the variation method for setting an integer value d as a gene mutation digit randomly select a variation point, to point to the variation as a starting point and then d consecutive gene loci value way completely permutations using produce a new individual, for

breach of task variation generated between mutual restraint individual relationship, where it directly discarded.

Algorithm steps. According to the basic process of genetic algorithm, can be obtained based on genetic algorithm QoS constraints, follow these steps [5]:

Step 1: initial population. Generating random manner initial population $P(0)$, calculate the fitness of each individual of the initial population, setting evolutionary iteration counter $k = 0$, $P(k)$ represents the k -generation population, the $P(0)$ as the parent populations.

Step 2: Termination condition. When the evolutionary reaches a predetermined maximum evolution generation or discover individual state remains in a stable state algorithm terminates, the output search results; otherwise continue.

Step 3: Select options. Roulette method using cross-operating on the population $P(k)$ the selection operation and it has strong adaptability to elect individuals to form a new group, the next step.

Step 4: Cross operation. Preclude the use of crossover from the group after the selection operation generated randomly selected two individuals crossover operation, generating two new individuals.

Steps 5: mutation. Preclude the use of mutation method selected individuals mutation operation, will produce a new individual into compliance with the requirements of the population $P(k + 1)$.

Step 6; update evolutionary iteration counter $k = k + 1$, go to Step 2.

Conclusions

Along with the progress of development of the Internet and technology, cloud computing has shifted from initial concept slowly approached our lives, have an impact on the way we live and work. But cloud computing does not look so nice, before it achieve full commercialization needs to solve many problems. Load balancing is a problem anycluster system will encounter, and cloud computing is no exception. And because of the special nature of cloud computing, it puts forward higher requirements for load balancing. From the characteristics of cloud computing, combining the characteristics and operation mechanism of genetic algorithm analysis, we found that the pheromone cloud computing model can play a great role in promoting genetic resource scheduling process fast balancing scheduling for cloud computing resources. Thus proposed load balancing slave group scheduling model based on genetic algorithm of cloud computing and various details of the algorithm are improved.

References

- [1] Cao Zhibo: Summary of Research, Vol.11 (2012) No29, p.4024-4025.
- [2] Li Jianfeng: Computer Applications, Vol. 4 (2010) No 13, p.34-36
- [3] Larry Rudolph: Lecture Notes in Computer Science, Vol. 3 (2013) No 22, p.93-96
- [4] BUYYAAR: Future Generation Computer Systems, Vol. 10 (2013). No 10, p.12-14.
- [5] ALI S, SIEGEL H J: Journal of Science and Engineering, Vol. 6 (2013) No 13, p.45-47