Study on Cloud Computing Resource Scheduling Strategy Based on the Ant Colony Optimization Algorithm

Linan Zhu[1], Qingshui Li[2], and Lingna He[3]

[1] College of Mechanical Engineering, Zhejiang University of technology

[2]College of computer science, Zhejiang University of technology

[3]College of computer science, Zhejiang University of technology

Abstract

In order to replace the traditional Internet software usage patterns and enterprise management mode, this paper proposes a new business calculation mode- cloud computing, resources scheduling strategy is the key technology in cloud computing, Based on the study of cloud computing system structure and the mode of operation, The key research for cloud computing the process of the work scheduling and resource allocation problems based on ant colony algorithm , Detailed analysis and design of the specific implementation for cloud resources scheduling . And in CloudSim simulation environment and simulation experiments, the results show that the algorithm has better scheduling performance and load balance than general algorithm.

Keywords: Cloud computing; Task scheduling; Ant colony algorithm; MapReduce, CloudSim

1. Introduction

Cloud Computing is hotspot for business institutions and research institutions, in last few years. It is mainly about how the computing resources are virtualized, and with scheduler the resources in the logical integration, focus on how to deal with data center resources virtualization, and user submitted to the mission needs and resources to maximum utilization rate for the user to provide services, The study data center of the services they provide types and service mode, How to efficient ly schedule user's tasks, reasonable distribution system resources, to realize the resource load balance is also the key factors of raising the cloud computing platform performance and service quality .

2. Cloud Computing

Cloud computing is in grid computing based on a new calculation model, is the next generation network computing platforms core technologies, It builds virtualization super computer, with on-demand rent way which provides data storage, analysis and scientific computing services through the distributed computing model and the resource pool technology. Cloud computing is also a kind of distributed computing, Through the virtualization technology will be distributed in the network computer resources of idle which combined into one huge resource pool, which is constituted as a super computing capacity of the computer. Calculation node can be put into dynamic system , all kinds of application system according to the need for computing resources, storage space and various software service, fully realize the dynamic autonomy function [1];

In general, cloud computing is a business purpose into forming the field network revolution, it is extension and development of parallel computing, distributed computing, and the grid computing. And increased many new features, the center thought is scattered through the high-speed network interconnection, Through the resource integration of virtual into pool way, to offer users of the WEB way, Users request a calculation, data center according to the task of segmentation and assigned suitable child nodes running, and the results of the calculation results are formed together and then returned to the user.

3. Cloud Computing System Structure

Cloud computing is using parallel computing to solve big problems and the calculated resources can be measured in services to users of the utility computing platform. It employs distributed processing, parallel processing and grid computing and distributed database to improve processing, in the Internet broadband technology and virtualization technology based on the high speed of development was born [2].

Cloud computing platform is a strong network of collaborative work, using virtualization dynamic expansion each node calculation ability, Connected with a lot of computing resources and services operating resources, And their resources through the cloud computing platform combined, constitute a have super computing power and storage capacity computing center, Mainly includes the management system, deploy tools, and resources monitoring module. Cloud computing system structure is as shown in figure 1.

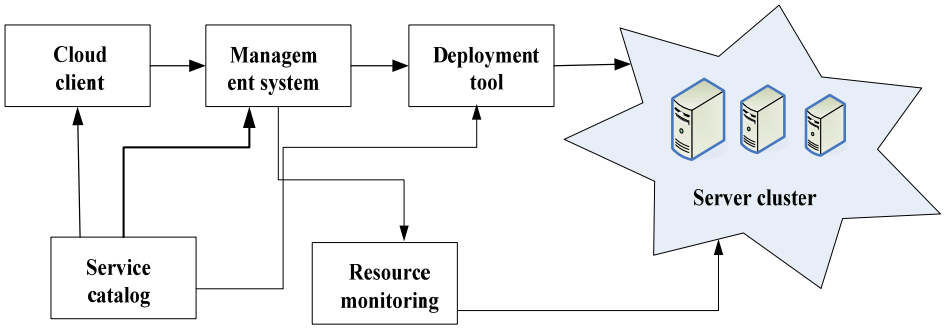

Fig. 1 The Structure of Cloud Computing

4. Task Scheduling Strategy of Cloud Computing

Task scheduling is computer science which has been one of the hottest researches, there are many experts and scholars published papers and journals project to discuss the task scheduling problem. In addition, many emerging disciplines of their research findings are applied to solve the scheduling problem, such as genetic algorithms, neural networks; artificial intelligence and distributed research which regards solving the scheduling problem as its research fields.

Resource scheduling is a crucial question of distribution and in cluster calculation; it gives the user task execution efficiency, the resources of the system numbers and the performance. From scheduling, heuristic scheduling algorithm in Grid Task Scheduling is used in most applications, the most effective, common heuristic scheduling algorithms are: simulated annealing, genetic algorithm, the ant colony algorithm.

4.1 The Task Classification Based On Qos

The Service of Quality (QoS) is internet properties of a security mechanism, in cloud computing environment, QoS is to measure the user's cloud computing application. Service satisfaction with the degree of important factors, cloud computing service function and performance evaluation will no longer be the traditional evaluation standard of service (such as speed, and cost-effective, etc), but with customer satisfaction as the goal, with the service quality to measure. Because the user's diversity, on cloud computing homework scheduling and resource allocation put forward higher request, According to the QoS parameters first task will be required in classification procession, which can be more accurate and timely, the task allocation to the most appropriate resources, generally consider QoS parameters have the following items[4]:

(1) Network bandwidth: when a customer t's communication bandwidth is high, such as multimedia data transmission, it should be priority bandwidth requirements and provide high bandwidth.

(2) Service completion time: for real-time demand higher users, need within the shortest possible time to finish tasks, and respond to user with submitted homework.

(3) the system reliability: to run a number of complex tasks users, need cloud computing center to provide a stable and reliable performance support, such as mass data storage service.

(4) Costs: cloud computing according to the needs to pay , cost is the user' attention focus, for the cheap service to users, cost is a standard.

Therefore, for different user needs, set up different QoS parameters, according to these parameters to measure the user's satisfaction, so as to establish the quantitative evaluation of different standards.

4.2 Mapreduce Level Scheduling

The key technology of Cloud computing MapReduce is step-by-step type processing technology. MapReduce class scheduling is the core of the cloud computing resources scheduling, and is the realization of the logical step calculation realize, all the task scheduling will be realized through this model. In MapReduce programming mode, concurrent processing, fault tolerant processing, load balance problems are abstracted for a function library Through the MapReduce interface, user can put the large-scale computing to be automatic concurrent and distribution implementation. MapReduce programming model calculation of the implementation process can be abstracted as three role [5]: Master, worker and user. Master is a central controller system, responsible for task allocation, load balance, fault tolerant processing, Worker is responsible for receiving task from Master , carries on the data processing and calculation, and responsible for data transmission communication, User is client, input task to realize the Map and Reduce function, control the whole calculation process.

5. Cloud Resources Scheduling Based Ant Colony Optimization

5.1 The Ant Colony Algorithm Principle

Ant colony algorithm is a random search algorithm, in TSP problem study well applied, TSP problem is a given n cities and a salesman starting from a city to visit the city through one and only one last return to the starting point of the shortest path. Introduce the following notation: Given: m is the number of the ants ant colony, $\tau_{ij}(t)$ show that the city i and j the path concentration of pheromone at the time, In the initial moments of each path equal to the amount of information, Set A = C (C is a constant), S express that t time ant k form the position i transferred to j the location of the probability:

$$p_{ij}^{k}(t)=\begin{cases}\dfrac{[\tau_{ij}(t)]^{\alpha}\cdot[\eta_{ij}(t)]^{\beta}}{\sum_{s\in allowed_k}[\tau_{is}(t)]^{\alpha}\cdot[\eta_{is}(t)]^{\beta}} & if\ \ j\in allowed_k \\ 0, & otherwise\end{cases} \quad (1)$$

among them, $allowed_k=\{0,1,.....,n-1\}-tabu_k$ express ant k next step is allowed to choose city, $tabu_k$ for the taboo table, record ant has traversed the city, $\eta_{ij}(t)$ is visibility for the t moment, calculated with heuristic algorithm, General access $h_{ij}=1/d_{ij}$ d_{ij} (i , j=l,2… ,n) which expresses the distance between city i and j, α , β respectively express the information of pheromone and path information transfer probability of impact. Ants in the search process, each move to a city must abide by formula (2) to the local pheromone update.

$$\tau_{ij}(t+1)=(1-\varepsilon)\tau_{ij}(t)+\varepsilon\tau_{o} \quad (2)$$

In formula $\varepsilon\in[0,\ 1]$ is a parameter; τ_0 is a constant.

When all ants complete a traverse, which is according to formula (3) Update the pheromone concentration on the path.

$$\tau_{ij}(t+n)=\rho\tau_{ij}(t)+\Delta\tau_{ij} \quad (3)$$

$$\Delta\tau_{ij}=\sum_{k=1}^{m}\Delta\tau_{ij}^{k} \quad (4)$$

$$\Delta\tau_{ij}^{k}=\begin{cases}\dfrac{Q}{L_k}, & \text{If the } K \text{ of ants in the time between } t \\ & \text{and } t+1 \text{ through } i\ j \\ 0, & \text{otherwise}\end{cases} \quad (5)$$

Where: ρ is residual factor pheromone, the 1- ρ is pheromone evaporation coefficient, To prevent unlimited accumulation of pheromone amount, usually set to 0 $<\rho<1$, $\Delta\tau_{ij}$ is this cycles path(i, j) of the pheromone increment, $\Delta\tau_{ij}^{k}$ express k-ants that in this cycle remain in the path i, j of the amount of information, Where Q is a constant, L_k express the number k of ants in this cycle taking the total length of the path. The shorter the path generated ants in this path, the more on the quantity of information, and we will have more ants' choices of this path [6].

5.2 Algorithm Scheduling Work Process

First of all, to obtain the user's task and according to priority order, classification, Classification that embodies the user's tasks to the requirements of different QoS, and on the basis of QoS classification use ant colony algorithm implementation of resource allocation and scheduling, Once meet the QoS requirements and shortest path , tasks and resources are bind, running tasks. The process is shown in figure 2:

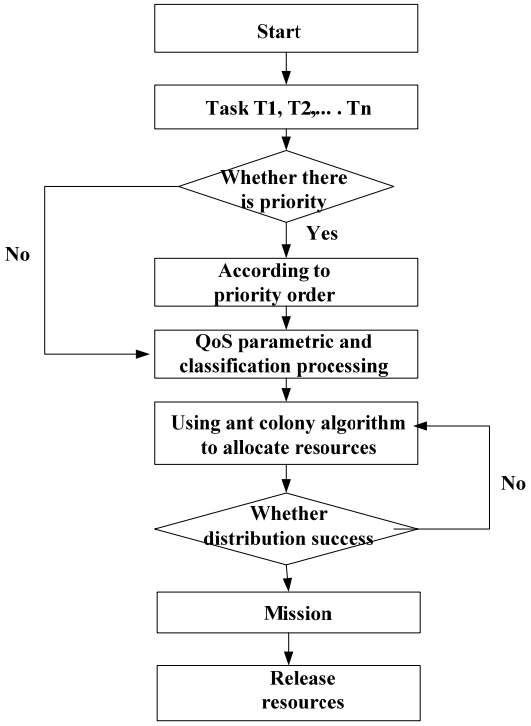

Fig. 2 Algorithm scheduling workflow

6. Experiments Simulation Results and Performance Analysis

This experiment using cloud computing simulation software CloudSim, it is in discrete events-based bag on the SimJava development function library, it inherits the GridSim programming model, in the Windows and Linux system cross-platform run, support computing clouds of research and development, and offers the following new features: (1) support the large cloud computing infrastructure of the modeling and simulation; (2) a self-contained support data center, service agent, scheduling and allocation strategy platform.

Emulator CloudSim simulation process: First establish a data center, then the data center creates a series of resources such as CPU, memory, bandwidth, etc. Then send registered message to CIS to be registered, once registered message can be used by DatacenterBroker management information interaction process, The flow chart is shown in figure 3:

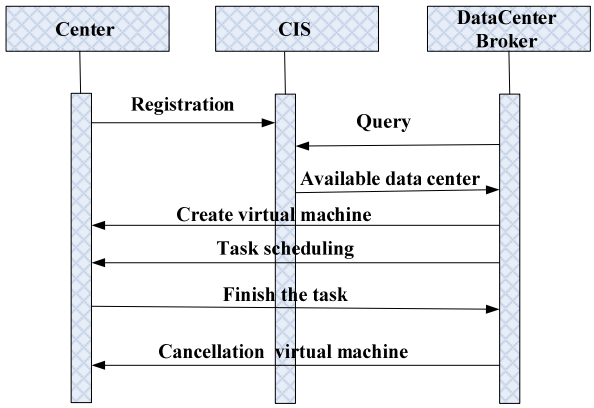

Fig. 3 The workflow of CloudSim emulator

In this experiment we set the number of task from 20 to 100, the number of node calculation of 8, In order to to show distinction, we designed the QoS attribute of node set up large gap, mainly including the CPU, memory and network bandwidth. Application of ant colony optimization (ACO) and random distribution algorithm (RA) respectively carry out 10 times and take an average, task execution time spent as shown in figure 4:

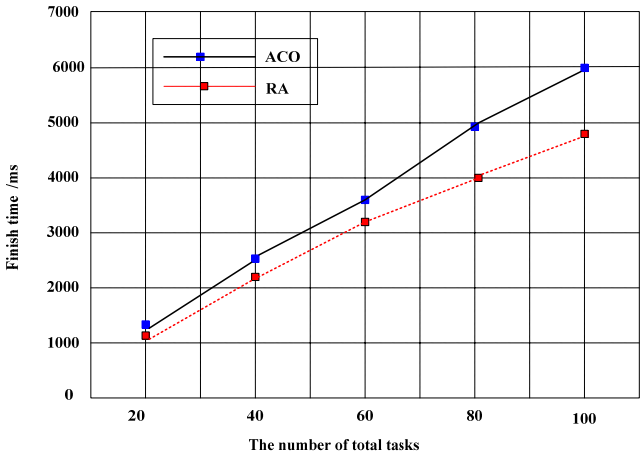

Fig. 4 The makespan of the two scheduling algorithm

It can be seen from the figure, with the increase of the quantity task, through the ant colony optimization

algorithm performs all the tasks, it takes the time less than general algorithm. Due to the ant colony algorithm which chooses target path through the pheromone strength, so when the task amount is less (such as 20), this algorithm implementation effect is not obvious, But when the task quantity achieved 80, two algorithms‘s execution time nearly one seconds. This indicates that within a certain range, with the increase of the number of task, ant colony optimization algorithm is better than general allocation resources and the execution time overhead is longer. This algorithm in cloud computing tasks in the debugging of the application is correct.

Conclusion

This paper makes research and elaboration on the cloud computing technology, and analyzes the cloud computing system structure and the realization of mechanism, resources scheduling strategy is the key technology in cloud computing. Therefore the use of ant colony algorithm for the basic model, detailed analysis and design of the cloud resource scheduling the concrete realization, And in the simulation software CloudSim simulation experiment, from the results we can see that, the algorithm for calculating node distribution and load balancing has good performance.

References

[1] China cloud computing. Liupeng: cloud computing the definition and characteristics [EB/OL]. http://www.chinacloud.cn/ the 2009-2-25
[2] Application Architecture for Cloud Computing. IBM, WTHIE PAPER
[3] Weiss. Computing in the Clouds, networker, Dec, (2007) vol.11:16-25
[4] WangXiaoChuan, JinShiYao, XiaMingBo. (2007)Web cluster of cybernetics QoS quantitative based on distributed control, Journal of software, November, Vol. 18 (11) : 2810-2818
[5] J.Broberg, R.Buyya, and Z.Tari. MetaCDN: (2008)Harnessing’Storage Clouds’for High Performance Content Delivery, Technical Report GRIDS-TR-2008-11, Grid Computing and Distributed Systems Laboratory, The University of Melbourne, Australia,
[6] Yang Jingyu,Gao shang. (2006) Swarm intelligence algorithms and its application, Water conservancy and hydropower press.

Linan Zhu received the degree in Computer Science & Technology from Zhejiang University of Technology, in 2004. He is a research student of College of Mechanical Engineering, Zhejiang University of Technology. Currently, he is a Lecturer at Zhejiang University of Technology. His interests include Cloud Manufacturing Networked Manufacturing Resource scheduling and PSS.

Qingshui li received master degree from Zhejiang University in 2002.He is a associate professor at Zhejiang University of Technology. His interests include HCI personalized recommendation data minning and Extenics.