

Scheduling Methods for Resource Management in the Cloud computing environment

Chunxia Zhao and Fan Liao

Henan University of TCM, Zhengzhou, China

Corresponding E-mail:springshiner@163.com

Keywords: Cloud computing;Scheduling problem;dynamic

Abstract. The core problem of cloud computing is to deal with the dynamic scheduling problem of physical and virtual resource, which can be regarded as a new problem in cloud computing, in order to achieve energy saving and emission reduction with high performance and low investment as well as some other targets. In this paper, it takes the overview of the cloud computing technology as the cutting point, with the help of the interpretation of computing resource scheduling under the environment of cloud, discussing the implementation of cloud computing task scheduling with optimization method of genetic algorithm.

Introduction

With the development of cloud computing in various areas, there are a variety of problems occurred such as resource management, resource scheduling and security mechanism and so on, which are reflected day by day, while the strategy of scheduling virtual resource has become the core issue of cloud computing research. The essence of scheduling resource is to find out the appropriate resources and make reasonable distribution for a large number of requests, so as to meet the request of task, optimize time and improve the utilization rate of resources as far as possible.

Overview of Cloud Computing Technology

Cloud computing is a new business model and service model, its core concept is to be no longer dependent on the local computer to do the calculations, but it can be operated by the third party to have calculations and store resources. It can distribute the computing tasks into the different data centers that are made up of a large number of physical servers and virtual servers, so as to make all kinds of application obtain computing power, storage space and information services according to the needs.

Cloud computing can integrate various large server cluster into a huge pool of resources to form as cloud, which can provide services through the Internet and make the underlying infrastructure abstract so that researchers can focus on their own areas, who does not need to care about other less important links. The service model of cloud computing can be shown in Fig. 1, including Software as a Service, Platform as a Service and Infrastructure as a Service.

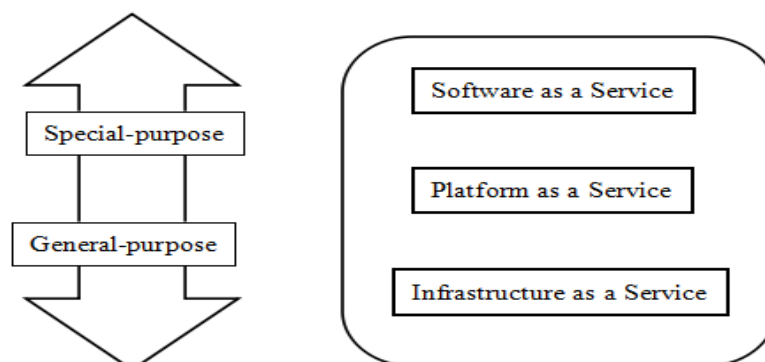


Fig. 1 Model of Cloud Computing Service

Resource Scheduling in Cloud Computing Environment

Resource scheduling refers to the process that applied the available resource to the requested task with resource assignment, so as to ensure the efficiency of finishing task and keep the load of resource balanced with the premise of completing tasks. Cloud computing is evolved on the base of grid computing technology, its resource scheduling method can be referred to grid computing in some extent, but the different scenarios make them have significant difference between the two scheduling methods: grid computing is mainly used in the field of scientific research, which put attention to the efficiency, optimizing the user's requests for the scheduling objective, while cloud computing has to face a large number of users' needs and different applications, and different application requests have different demands on quality of service, therefore, the required resources are also different. In addition, the types of application to cloud computing platform resources are widely distributed with many kinds of factors, which have increased the difficulty to manage resource.

During the process of resource scheduling, as for a large number of resource requests, firstly, it should detect the performance of virtual machine resources, computing the occupancy rate of current resources of the virtual resources as well as the allocation standard deviation function of the resource, real-timely understanding the load of virtual machine, trying to find out more suitable virtual resources to perform tasks, so as to reduce the load of resources and make the correspond request for resource.

Resource scheduling algorithm is the core to realize the optimization of scheduling resource, which is the key of all kinds of resources and service systems that can provide users with the appropriate resources.

Cloud Computing Task Scheduling of Optimal Genetic Algorithm

Design of Optimization Model

In this paper, by using the characteristics of Map/Reduce, it can set up resource scheduling optimization model under the cloud environment. Map/Reduce is by using the distributed parallel algorithm, to deal with a programming model for large data sets in clusters, it is also a kind of common resource scheduling model, which is widely used, Map/Reduce mainly consists of a Map process and a Reduce process, the Map process can divide tasks into smaller sub-tasks and assign them to the working nodes, after the implementation of task, the working nodes can return the results to the main nodes, when the main nodes receive the inputting tasks. While Reduce process can have the reduction of integration in some way and then output the final result, when the master nodes receive the execution result of the working nodes. Combined with the characteristics of Map/Reduce model, this paper takes the optimal time span and balance load as the goal to establish the optimal mapping relationship between tasks and virtual machine with cloud computing resource scheduling model, which can be shown in Fig. 2.

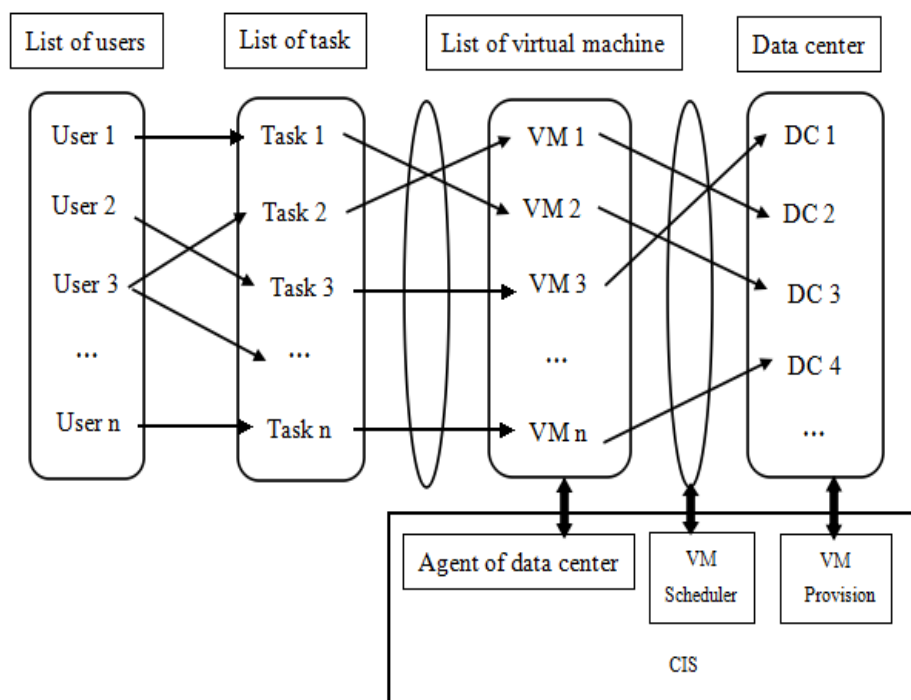


Fig. 2 Resource Scheduling Model of Cloud Computing

The core idea of the optimization model is: when the user submits tasks, these tasks can be decomposed into a number of sets including a plurality of <key, value> according to different user's ID, this process is the mapping phase of the Map/Reduce model; these elements in the collection can be performed in parallel, and the value of key is associated with the user's ID, so as to make response to the users who completes submitting the task scheduling, then through the resource scheduling strategy, with the optimized targets to assign these sub -tasks to the resource nodes of virtual machine. This process is completed by data center server agent; finally after operating the completion of the results, it can make response to users, which is the simplification stage of Map/Reduce model.

The Processing of Index

This paper is mainly based on the fitness function of genetic algorithm to design scheduling algorithm, while scheduling algorithm in cloud environment is close related to the time span of virtual resources nodes and the balance of load, therefore, it needs to have the corresponding treatment over these two indicators, namely, time span and balance of the load balance.

Imaging, there are m sub-task requests on n resource data centers, among them, the number of task that is assigned to each resource is j , the time span that this data center can complete the task can use *time cost* to represent, thus, the formula of calculating *time cost* is as follows:

The three sub- indicators for the balance of load, namely: CPU, the unit of memory and network bandwidth is not the same, so these parameters should be normalized to one unity, so as to eliminate the differences between each sub- index in dimension.

Conclusion

As new type of distributed computing, cloud computing can integrate and manage various types of distributed network resources, however, during the process of unified management, the huge number, strong distribution, plus the dynamic change of resource can have many problems. Therefore, choosing the resource scheduling strategy to improve the organization and scheduling can have a significant meaning not only in resource sharing and utilization efficiency, but also in saving energy and reducing operating costs. Therefore, the resource scheduling becomes one of the key technology of cloud computing.

ACKNOWLEDGMENT

This research is supported by the doctoral research fund of Henan University of TCM (NO.BSJJ2010-13) and the basic and frontier project of Science and Technology Department of Henan Province (NO.142300410428).

Reference

- [1]Jun Ye. 2011. Cosine similarity measures for intuitionistic fuzzy sets and their applications. *Mathematical and Computer Modelling*. vol.53, pp91-97.
- [2]J. Tan, P. Dube, X. Meng, and L. Zhang. 2011. Exploiting Resource Usage Patterns for Better Utilization Prediction.In *Proc. of the 31st International Conference on Distributed Computing Systems Workshops (ICDCSW'11)*. pp14-19.
- [3]C. Mark, D. Niyato, & T. Chen-Khong. 2011. Evolutionary Optimal Virtual Machine Placement and Demand Forecaster for Cloud Computing. *Proceedings of the 2011 IEEE International Conference on Advanced Information Networking and Applications*. pp348-355.
- [4]Bezdek J C, Pal S K. 1992. *Fuzzy models for pattern recognition*. New York: IEEE press. pp27.
- [5] Tseng, F. M., & Tzeng, G.H.. 2002. A fuzzy seasonal ARIMA model for forecasting. *Fuzzy Sets and Systems*. vol.126, pp367-376.