Cloud Resource Adaptive Scheduling Framework and Optimization Strategy Based on Swarm Intelligence

HW Zhao¹, S Zhang¹, Y ruan¹ and XH Jing¹
¹ Shenyang University, Information College, Shenyang, China

Abstract. Resource scheduling framework is very important for the overall cloud service efficiency and service quality, which is the core module of cloud computing. This paper proposes an adaptive cloud computing resource scheduling framework based on swarm intelligence. The proposed adaptive framework includes resource deployment module, scheduling module, recommendation module, optimization module, and monitoring module. The simulation results show that the proposed framework is more efficient than the traditional framework in terms of task execution time, system load balancing and resource service quality, and can improve the service quality of cloud applications while improving resource utilization.

1. Introduction
Cloud computing is a computing model that provides dynamic and virtual resources by service, including cloud services of various functions. The Scheduling framework of cloud computing mainly includes hardware resource layer, virtual resource layer, cloud system management layer and cloud application layer. The hardware resource layer includes large-scale blades, cabinets and tower servers, centralized and distributed storage, and network infrastructure such as routers and switches. The virtual resource layer mainly integrates various resources by IAAS layer to form a virtual resource pool with super capabilities [1][2]. The cloud system management is mainly responsible for security management, task management, resource management, user management, etc. It is the most important level of cloud computing framework. The cloud system management manages the resource usage of the entire virtual resource layer and the service status of the cloud application layer. The cloud application layer is responsible for integrating virtual resources with different functions for different categories of users, such as integrating application resources and network resources into WEB-based access methods. Cloud resources also provide access to the login interface to ensure that authorized users can use cloud resources [3][4]. Therefore, the process of providing the service by the cloud computing service is that the user submits the task to the cloud system through the cloud application layer interface (SAAS/PAAS). After the task is submitted to the cloud system, the cloud application layer is responsible for submitting the task to the cloud system management layer. After the cloud system management layer is processed by the verification process, the resource scheduling component is responsible for allocating resources of the resource pool layer to the task, and finally returns the processing result to the user.

Based on the above research on cloud computing framework and service functions, it can be seen that resource scheduling management is very important for the overall cloud service efficiency and service quality. It is the core module of cloud computing and the key work of this paper. In order to implement the load balancing and improve resource utilization of Cloud Computing system, how to schedule resource becomes a central mechanism in Cloud Computing system.
2. Related works
In recent years, with the rapid development of Internet technology, there has emerged various of network-based applications, such as E-mail and microblog, which brings a lot of convenience for people. At the same time that hundreds of millions of users browse information via the Internet, a huge amount of data have been generated continuously, such as electronic trading records, user access logs, etc. Besides, many large enterprises and organizations also generate a mass of data, including the stock information from stock exchanges and the oceanic data from monitoring stations, etc. In the face of such large data sets, how to efficiently process large-scale data sets and dig out valuable information becomes a point of issue concerned by lots of IT enterprises and scholars.

At present, the cloud virtual resource scheduling problem is a major challenge after cloud security issues. Although the work related to the cloud virtual resource scheduling problem has yielded rich research results, these results are often concentrated on an optimization goal, such as application performance priority, energy consumption priority or cost priority[5].

Most of the cloud computing platforms adopt a preset fixed number of physical devices and pre-configure the size of the virtual resources for the application[6][7]. The overall framework is not suitable for the current flexible and on-demand resource characteristics of cloud computing platforms, and the diversity of cloud services and the complexity of cloud computing environments. Requirements, cloud user requirements, and cloud resource prices all have certain dynamics[8]. As one of the core problems of cloud computing, the efficiency of scheduling algorithm has a direct impact on the operation capacity of the system. Swarm intelligence algorithm, with good coordination and overall stability, is one kind of swarm intelligence algorithms which imitates swarm intelligence in the process of evolution swarm. These algorithms have characteristics of simple structure and strong searching ability, which can improve the accuracy and efficiency of cloud-computing resource scheduling algorithms. This paper studies on how to use biological group intelligent algorithm of cloud-computing resource scheduling problem. This dissertation proposes three improved artificial intelligence algorithm and an adaptive framework for cloud computing resource scheduling model and implements the application of cloud computing framework[9][10].

3. Adaptive optimization of dynamic scheduling framework (DFAOC)

3.1. Basic functions of DFAOC adaptive dynamic scheduling framework
(1) Cloud computing platform portal
Cloud computing platform portal: mainly through WEB mode, responsible for receiving various tasks submitted by users. The platform portal includes various application services such as IAAS , SAAS[11], and PAAS[12].

(2) Resource deployment module
Resource deployment module: The resource deployment module includes application resource deployment, virtual machine resource deployment, and corresponding deployment strategy. The deployment of the application resource refers to the virtual resource that meets the deployment requirements according to the deployment policy, and the application is deployed on a specific one or several virtual resources. The virtual resource deployment refers to the use of the physical resource. In the case, deploy the requested virtual resources to one or several physical resources according to the deployment strategy. After receiving the deployment task, the deployment policy first determines the deployment type, and then binds the application service and the virtual resource according to the corresponding policy, and binds the virtual resource and the physical resource. The application resource deployment and virtual machine resource deployment can be deployed at the same time or at different times. The deployment strategy of this module mainly uses the physical resource scheduling model and the swarm intelligence algorithm to achieve optimal allocation of resource scheduling and improve the efficiency of using various cloud resources. In addition, the resource deployment module
automatically optimizes deployment according to the information that is suitable for optimizing the module, and provides scheduling support for various resources for the scheduling module.

(3) Resource scheduling module
The resource scheduling module needs to ensure the maximization of resource utilization while ensuring the service quality of the cloud computing user. The module is mainly based on the request of the user or the application, and is based on the resource analysis/prediction module, the resource deployment module, and the resource monitoring module. Information, complete the scheduling of cloud resources (virtual scheduling and physical scheduling). This paper mainly uses the virtual resource scheduling model and the swarm intelligence algorithm proposed in the third and fourth chapters.

(4) Adaptive optimization strategy module
The adaptive optimization module includes two main modules: the tuning model and the optimization strategy. The main idea is that, in a time period, the tuning model module obtains the current virtual resource and physical resource load information according to the resource monitoring module, and passes the resource prediction module. Obtain the resource demand information in the next cycle, and perform adaptive comprehensive analysis in combination with various tuning models such as load and energy consumption to determine whether the expected use and demand of various resources exceed the set dynamic threshold. If the expected exceeds the standard, At the same time of initiating the resource warning, the optimization strategy module is used to perform adaptive optimization adjustment such as over-targeting, resource expansion and addition, light load merging, and virtual machine dynamic migration.

(5) Resource analysis and prediction module
The resource analysis and prediction module mainly analyzes and predicts the usage of various resources based on real-time operation optimization and service history information of various resources. The module predicts the use of the PSO-RBF algorithm to predict resource requirements. In addition, the module classifies cloud resources according to resource usage and user dynamic requirements, and sets different indicator weights according to different usage rates of CPU, memory, hard disk, and network.

(6) Resource Monitoring Module
The resource monitoring module mainly performs resource monitoring according to user conditions, resource usage, optimization, and cloud service quality information, including whether the user's QOS and SLA are satisfied, whether the SLO of the application is satisfied, and the virtual threshold is monitored according to the formula. Information such as overload, low load, and failure of resource and physical resource nodes provides support for predictive modules, adaptive optimization modules, and deployment modules.

3.2. Optimization of cloud resource scheduling under DFAOC adaptive process
Based on the DFAOC framework, cloud computing physical resource scheduling optimization problem, each particle represents a physical resource allocation scheme. Currently, typical swarm intelligence algorithms are used to solve continuous optimization problems. Due to the discrete nature of cloud resource scheduling problems, individuals need to be specially coded.

1) Obtain the optimal parameters of the RBF neural network through PSO, and obtain the PSO-RBF optimized neural network;
2) Collect training data and conduct learning training through PSO-RBF neural network;
3) Input the measured data, predict the application resource demand through the PSO-RBF neural network;
4) Initialize the virtual resource pool of the cloud computing system resource according to the prediction result, and dividing the resource pool into the $m$ resource subgroup;
5) Use a uniform method to generate an initial particle swarm, and set the initial position and initial velocity;
6) Calculate the fitness value of each particle, and select the optimal particle as the optimal solution of the individual optimal solution, the sub-population optimal solution and the neighbor group;

7) Updating the velocity and position of the particle swarm, and comparing with the individual optimal solution and the optimal group solution, and updating the optimal solution of the individual, the optimal solution of the sub-population and the optimal solution of the neighbor group according to the result;

8) Select the variable particles generated according to formula to generate the mutated particles, and flying backwards;

9) Judge the termination condition. If it is satisfied, the optimal particle position of the group is calculated, and the optimal scheduling scheme of the cloud computing resource is obtained. Otherwise, the iteration is continued.

4. DFAOC framework verification

This experiment verifies that the DFAOC framework is based on the traditional cloud computing resource scheduling model and the cloud resource adaptive model defined in this paper, and uses the adaptive management process of the DFAOC framework as a method. In the DFAOC framework verification process, the adaptive management process and the physical resource scheduling module are mainly used for verification. Other modules adopt the default configuration to test the scheduling efficiency of cloud resources under the DFAOC framework. In the DFAOC framework, the virtual machine initialization resource deployment module, the adaptive optimization module optimization model, and the analysis recommendation module can be dynamically set and changed, with greater flexibility and scalability, and only the DFAOC framework prototype system can be fully implemented. The overall accurate feasibility verification is carried out, but this is a complicated and heavy work. This part of the work will be reserved for the next step of study. Therefore, this paper uses the simulation key module verification method.

This simulation experiment uses the simulation platform of CloudSim, because the CloudSim can simulate the scheduling strategy in the cloud computing environment. This experiment extends the CloudSim platform according to the DFAOC framework, rewrites a series of classes, and defines Cloud resource basic description model, resource load measurement model, dynamic threshold model, multi-attribute weight model, build the program structure of resource deployment, adaptive optimization, analysis and prediction, and simply set the default processing jump. The programming tool is Myeclipse 9.0, verifying the validity of the DFAOC framework.

5. Conclusion

This paper analyzes the specific scheduling framework and scheduling algorithm involved in the cloud computing resource scheduling system. Firstly, the adaptive scheduling framework DFAOC for adaptive optimization of cloud computing resources is designed, and the specific structure and form of the framework system are given. Finally, based on DFAOC cloud resource scheduling framework, PSO optimized RBF neural network algorithm is simulated on CloudSim platform. The simulation results show that the proposed framework can effectively improve the service quality and resource utilization of applications in cloud computing system.

Acknowledgments

The authors are supported financially by International cooperation project (Project No. S2012ZR0191) and the Natural Science Foundation of Liaoning Province (Project No. 20170540646).

References

[1] Coello C., Carlos A., Lechuga M. S. (2002). MOPSO: a proposal for multiple objective particle swarm optimization, Proceedings of IEEE Congress on Evolutionary Computation. USA, Honolulu: IEEE, 1051-1056.
[2] Cao H., Liu Y. (2010) CBFO: The cooperative optimization of Bacterial Foraging, in Computer Application and System Modeling (ICCASM), 2010 International Conference on, 2, 106-109.

[3] Devi G., Raju G., Sridevi P. (2015). Sridevi Application of Genetic Algorithm for Reduction of Sidelobes from Thinned Arrays, AMSE JOURNALS, 58(1) 32-52.

[4] Doaa M., Faten H., Ninet M. (2012). Ahmed A New Control and Design of PEM Fuel Cell System Powered Diffused Air Aeration System. TELKOMNIKA Indonesian Journal of Electrical Engineering, 10(4): 291-302.

[5] Guo M.M., Yang S.L., Yan H., Kan L.F., Yang B. (2014). MOBILE VIDEO ALARM SYSTEM BASED ON CLOUD COMPUTING, REVIEW OF COMPUTER ENGINEERING STUDIES, 1(2), 5-6.

[6] Hossain S. Infrastructure as a service [J]. Cloud Computing Service & Deployment Models Layers & Management, 2013: 22(7), 26-49.

[7] Tsafrir D, Schuster A, Benyehuda M, et al. Deconstructing Amazon EC2 Spot Instance Pricing. Third International Conference on Cloud Computing Technology and Science [C], 2012: 304-311.

[8] Kippenbrock T, Holloway E, Moore D D. Google Docs [J]. Computers Informatics Nursing Cin, 2010: 138-140.72. Karun A K, Chitharanjan K. A review on hadoop — HDFS infrastructure extensions. Information & Communication Technologies [C]. IEEE, 2013: 132-137.

[9] Browne, Clarke, Mishra. Automatic Verification of Sequential Circuits Using Temporal Logic [J]. IEEE Transactions on Computers, 2006, 35(12): 1035-1044.

[10] Mitra S, Datta S, And T P. Introduction to Physical Polymer Science [J]. Macromolecular Chemistry & Physics, 2010, 207(8): 787-787.

[11] Majhi J, Smid M, Smid M. Multi-criteria geometric optimization problems in layered manufacturing. The Fourteenth Symposium on Computational Geometry [C]. ACM, 1998: 19-28.

[12] WANG G H, LI Q H, LIU A F. Multi-objective optimization cloud workflow scheduling evolutionary genetic algorithm [J]. computer sciences, 2018, (5): 31-37.