

Research on the Cloud Computing Load Balance Degree of Priority Scheduling Algorithm based on Convex Optimization Theory

Rong Wang

Hubei University of Science and Technology,
Xianning, Hubei Province, 437100 China

Bin Rao *

Hubei University of Science and Technology,
Xianning, Hubei Province, 437100 China

*Corresponding Author: Bin Rao

Abstract. In this research paper, we research on the cloud computing load balance degree of priority scheduling algorithm based on convex optimization theory. Clustering is an important problem is the load balancing in the system. Load balancing cluster system goal is according to the performance of the processor to allocate commensurate with the task. Cloud services in one of the core technology of virtualization technology can make a variety of computing resources dynamically organize, computer hardware architecture and software system will separate. The theoretical and numerical analysis show the effectiveness and feasibility of our propose method. The detailed discussion will be conducted in the following subsections with examples and mathematical inductions.

Keywords: Cloud Computing; Load Balance; Convex Optimization; Priority Scheduling.

Introduction

In recent years about load balancing algorithms emerge in endlessly, and more and more toward intelligent direction. Researchers more load information known and unknown forecasting information as the judgment standard of load balancing. Yet the former bring extra resources enormous overhead and hysteresis, often do more harm than good and the latter also because too much subjective information into which reduces the algorithm accuracy. Because of the lack of theoretical analysis of load balancing nature, limit load balance technology further development. The nature of the load balancing problem can be understood as system's internal

energy equilibrium problems. Any system has its physical meaning, the nature of cloud computing cluster system. Clustering is an important problem is the load balancing in the system. Load balancing cluster system goal is according to the performance of the processor to allocate commensurate with the task, to minimize the execution time of the application. Load balancing strategy can be divided into static load balancing and dynamic load balance. Static load balancing algorithm is based on the system of prior knowledge to make a decision, run-time load redistribution [1-2].

Information collected at this stage after the node selection algorithms are all based on single node load information for sorting or classified to choose the best nodes, without considering cluster overall or regional load conditions. This was suitable in the traditional cluster, but for the partition management area of swarm is not appropriate. An area as a whole, for example, has been in a saturated state, but probably already release the spare resources, when using the traditional algorithm for task assignment, are likely to choose the node, thus causing partial load overload affecting the overall performance of the system [3]. Cloud services in one of the core technology of virtualization technology can make a variety of computing resources dynamically organize, computer hardware architecture and software system will separate, make retractable transparent computing system structure. Virtualization technology can be widely used in modern computers, mainly has the following two advantages: on the one hand, can aggregate a variety of computing resources,

improve the utilization of computing resources, and can according to user's personalized needs to provide the use of computing resources environment, thus the flexibility to build to meet various application requirements of computing environment [4].

Monitor node status, the reasonable allocation of resources, load balancing mechanism is becoming the key factor of swarm group of reliable operation. In a cloud environment, the traditional load balancing algorithm will face three problems: the node load condition after the definition, load information acquisition, information of the node selection algorithm. Gathered group is composed of numerous physical server and tenant's virtual machine instance distribution in different physical servers, because the tenant demand is different which cause its purchase of the virtual machine configuration and performance is different, make part of the performance of the physical server has not been fully utilized. Cloud data center is a large-scale data operation and mass storage function of cluster resources. Traditional data center frame through switches connected in a single cluster of communication, access layer switches have two layers, the upper switch with load balancing device.

In this paper, we conduct research on the cloud computing load balance degree of priority scheduling algorithm based on convex optimization theory. Single point in the centralized monitoring architecture, the monitoring server node is mainly from monitored data access to load on server node information. System administrators with the method of batch deployment will be deployed in the cluster agent process server nodes to load data broadcast to the monitoring server. Monitoring server deployed on a single node, collect the monitored node in the cluster agent send the load data, through the preprocessor analysis of these data, then through drawing tool show in the form of Web operations staff or put first to cluster corresponding database, the data show by querying the database operation.

Centralized single point monitoring architecture is mainly the monitoring server directly to monitor the monitored object. Monitoring server to be monitored cluster monitoring, monitoring of physical nodes to run on collecting load daemon, its load data obtained by periodically run a Linux command, finally sent to the monitoring server. In the following sections, we will discuss the issues in detail.

The Proposed Algorithm

The Convex Optimization Theory. Perturbation theory is an important branch of modern applied mathematics and it is to solve the nonlinear differential equation with variable coefficients, the approximate analytical solution of a kind of effective method. Its central idea is: depending on the speed change between state variables, and to classify them according to a certain time scale. Forced singular perturbation approach is used to artificially changing fast state variables on the first derivative of adding small parameters, with small parameter tends to zero value, you can put the order reduction of higher order differential equations into some low order differential equation, then state variables series form of generative containing perturbation parameter, the solution for asymptotic approximation with perturbation parameter.

By the general uncertainty relation in the process of information transmission, set up network is proposed to fit, said learning ability training relative error of the training sample set and test sample set test generalization ability, relative error and the number of hidden nodes of network structure and characterization of the complexity of sample meet the general uncertainty relation between multiple correlation coefficient and based on this relation, the export of a given sample set, under the request of meet the approximation error, follow the principle of network structure design of the Jane's coordinated network generalization ability and learning ability and the formula of optimal number of hidden nodes and finally pointed out

that in order to make the training of the network has good generalization ability best stop method which is shown in the formula one.

$$\Delta p \cdot \Delta x \geq (Sh) / (2 \log_2 (1 + M / N))$$

(1)

From the information theory point of view, the network learning and testing process can be regarded as the transmission process of information. Some of the structure of the network is an information system. The so-called network learning and testing process is according to different network structure, adjust the network weights of threshold parameters, such as to approximate simulation function, the approximation error small as possible which is shown as follows.

$$R = \sqrt{1 - R^* / R_{yy}^*}$$

(2)

But the truth is: no matter what kind of network structure and how to adjust the design parameters, when learning to a certain stage, if learning training sample set of average relative error is reduced, and the increasing average testing relative error instead of testing sample set, which appeared over fitting phenomenon. That is network is proposed to fit and its learning ability and generalization ability is the change between the reverse, only in a fitting beginning point, is the best match between them. As the training sample set is always limited, therefore, the network of training set samples of the generalization error is inevitable.

The Balance Degree of Priority Scheduling.

In recent years, with mass data processing and the parallel development of distributed computing, virtual machine cluster technology is also in slowly rises. As a new topic in the field of high performance computing and information services, virtual machine technology application from the theory of cluster parallel computing, distributed computing, grid computing, can meet the demand of service integration and dynamic from multiple aspects. From huge amounts of data processing to provide computing resources

service, to save costs, etc., the virtual machine cluster technology has made the prominent contribution. In addition, as a leading cloud computing technology, the virtual machine left cloud cluster technology with extensible, such as green energy saving and dynamic deployment features. Cluster and virtual machine technology is one of the main research content of load balancing, especially the load balancing in cloud computing environment to start new changes have taken place, in the form of load balancing based on software becomes more and more important. In the figure one, we illustrate the general pattern.

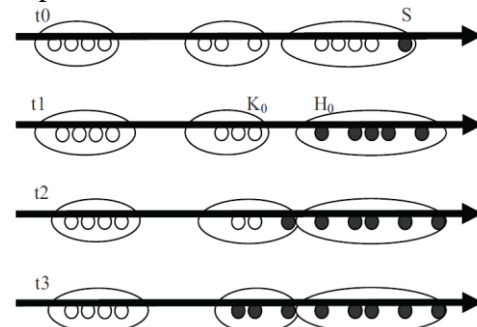


Fig. 1 The General Pattern of Priority Scheduling

Weights is embody processor performance of abstract, tradition that weights are set by the system administrator by experiences an integer value, weight big server can handle more web request. If the system administrator set of weights is unreasonable, can cause within each cluster server load distribution is not balanced. In view of this, the domestic many institutions have done related research of dynamic load balancing, and its dynamic characteristic is through constant feedback and correction server to achieve the static weight, and to form the value of dynamic change and changes with the server performance, effectively balanced server cluster. Due to the weights of combination with data and significant testing time is long, give a relatively optimal weight calculation method. Through the definition of load is given and the design and implementation based on dynamic feedback load balancing algorithm, maximization and the rational use of existing resources, give full play

to the virtual machine cluster strength, not a moment exists because of the heavy load of smooth implementation of the virtual machine delayed tasks, or because of some machines too idle to resource didn't get to good use. In the following formula, we show the entropy value calculation.

$$Q = \Delta S = S_1 - S_0 = k(LNn_1) \quad (3)$$

Network address translation way is through an external IP address mapping for multiple internal IP addresses, dynamic using one of the internal address every time, achieve the goal of load balance. The implementation way by many hardware vendors are integrated into the switch, the servers are strategy is based on the number of connections or response time to select the actual server and forward the request. The complexity of the generalized set describes general set inside the severity of the general difference of each flag values. When within a collection of individuals certain properties have definite value in determining the moment, and individuals with different characteristic values known, from the total complexity minus the flag values in the same part of the complex degree of the complexity of the set can be generalized internal state:

$$C = N \ln N - M \ln N \quad (4)$$

The LVS load balance the advantage of the underlying implementation forward requests the efficiency is very high, is widely applied to various types of hardware products. Defect is mainly dependent on the static algorithm which fails to reflect the status of the current server load. Computer cluster is usually better than a single computer has faster computing speed and reliability, and on the cost performance is much better than the single server such as workstation or the super computer. Cluster technology using cheap server as a node, at very low cost to realize high speed calculation, mass data processing, meet the demand of the explosive growth of the current information processing. In terms of the

high availability cluster system with the character with not interrupt service is widely used, the heartbeat detection technology is presented for high availability.

The Cloud Computing Load Balance. High performance computing cluster, mainly focused on the cluster computing capacity, parallel algorithm for solving the problem of complex scientific computing research and development related software. High performance computing cluster of many nodes via Ethernet connection research high-performance computing the two main problems: large-scale scientific computing and storage and processing mass data problems. Cloud computing according to whether the public service can be divided into public cloud, private and hybrid cloud. Public and private cloud technically there is no essential difference, and only operate and use object is different, the former refers to the enterprise to use other units operating cloud services.

While the latter is the enterprise own operating and using cloud services platform. The general dynamic feedback algorithm is a certain amount of time between all the server nodes on the monitoring results are sent to the load balancer, such as literature. But doing so makes all nodes at the same time access to the load balancer, largely consumed the resources of the load balancer. And there are a lot of server load in this interval weight change is not big, even no change. Unable to control request task specific time of arrival and quantity, the fixed forwarding mode itself may cause request load distribution is extremely uneven. Because of the weight of the load is through a certain algorithm, rather than the server the real situation, so the value of the system are estimated and the real value will more or less there is a gap, so we choose the lightest server also may not be the true sense the lightest server. Therefore, the proposed methodology is urgent and necessary.

Conclusion

In this research paper, we research on the cloud computing load balance degree of priority scheduling algorithm based on convex optimization theory. Cloud storage to network in many different types of storage devices by applying the software integration, and work together, common external providing services such as data storage and access model. Compared with cloud computing systems, cloud storage is mainly focused on by allocation strategy to achieve the improvement of the performance of the storage and access. Due to difference users interested in its focus on the part of the overall file, only in this case, to the overall backup file may not only waste a lot of storage resources. Our proposed method solves the problem well. In the future, we will research more on the optimization part to modify the objective functions used in this paper to obtain better result.

References

- [1] Mao Y, Chen X, Li X. Max–Min Task Scheduling Algorithm for Load Balance in Cloud Computing[J]. *Advances in Intelligent Systems & Computing*, 2014.
- [2] Zheng H, Guo P, Li Q, et al. On a Load Balancing Algorithm Based on Bandwidth Constraints in Cloud Computing[J]. *Journal of Southwest China Normal University*, 2014.
- [3] Jianfang C, Junjie C, Qingshan Z. An optimized scheduling algorithm on a cloud workflow using a discrete particle swarm[J]. *Cybernetics & Information Technologies*, 2014, 14.
- [4] Chiu C F, Hsu S J, Jan S R, et al. Task Scheduling Based on Load Approximation in Cloud Computing Environment[J]. *Lecture Notes in Electrical Engineering*, 2014:803-808.