Journal of Theoretical and Applied Information Technology
20th May 2013. Vol. 51 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A METHOD OF SYNTHETIC SCHEDULING BASING ON THE CLOUD COMPUTING

ZHAO HONGWEI
Asstt Prof., School of Information Engineering, ShengYang University, ShenYang ,Liaoning,China
E-mail: zhwzys@163.com

ABSTRACT

The traditional Scheduling method in the Cloud Computing system can only provide load distribution for resource Scheduling. In order to improve the utilization ratio of the resource as well as handling up rate of the Cloud Computing system, a method of Synthetic Scheduling basing on the Cloud Computing has been designed and implemented after the study on the Cloud Computing system. It can provide a better method to solve the problem of scheduling in the Cloud Computing system .Firstly, a layered loading balancing scheduling mode has been proposed. Secondly, a comprehensive resource distribution algorithm has been designed and implemented in consideration of respective local resource counts, each join points' performance and current load distribution. Finally, the result of the experiment indicates that the scheduling system can improve the efficiency of dispatching resource and the utilization ratio of distributed Cloud Computing resource.

Keywords: *Distribution; Cloud Computing; Load; Load balance*

1. INTRODUCTION

Cloud Computing is a type of integrated resource and resource environment,which includes various related resource and resource integrated with computing power, data information and knowledge, software and people, aiming at organizing the geographically spreading computers into a "virtual supercomputer" with the use of internet. Cloud Computing is a special type of distribute Computing, which improves the application of the Web resource technology, provides a range of resource sharing to the extreme for enterprises and is on the way to become the most effective way to increase the overall level and capacity of enterprise.

At present, there exists a great deal of research on Cloud Computing in all aspects, ie., relatively well-known systems, such as google, Hadoop, EC2, and so on, with the developing aim to take the effective use of the geographical distributed resource, in which effective scheduling strategy seems very critical for optimizing resource utilization rate.

Similarly, in the process of resource schedule of Cloud Computing, we also need to coordinate the use of the distributed resource to carry out transparent and automatic resource adjustment by a number of local agents. In order to implement the balanced distribution [1] of the Cloud Computing system and to improve the utilization ratio of the resource as well as handling up rate of the system, how to realize the resource distribution has become the core of the mechanism of Cloud Computing system, The distribution of resource depends on the access to the load information of calculation nodes and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

processing technology. Therefore, how to access the load information of nodes, how to measure and evaluate the load condition of local agents with the use of above mentioned information for resource distribution have become the important load-balancing system research.

2. RELATED WORK

In Cloud Computing [2],[3] system, resource is dynamically generated with the size of the load of each local agent changing dynamically, thus only dynamic load balancing scheduling strategy is usually put into consideration instead of the static load-balancing scheduling method.. In general, dynamic load balancing scheduling can be divided into two broad categories, centralized scheduling and distributed scheduling [4], All of the resource of the first one is submitted to the global agent, which will be in charge of collecting load information of the relative local agent to determine the load-balancing scheduling program. In this mode, local agent does not carry out scheduling but to distribute the resource assigned by the global agent and submit the resource back to the global agent when it ends, the main advantage of which lies in its relatively simple realization and the disadvantage of which lies the high costs of scheduling for the global agent would become the bottleneck of the system in the environment of large–scale Cloud Computing with a great number of nodes. Each of the local agent of the second one can receive resource and carry out scheduling, realizing load balancing operation according to some of the load information in the local scope, by which each computer would broadcast its load information to others on a regular basis to update those of the local maintenance Load vectors, the biggest advantage of which is good scalability and the drawback of which is the long time to wait for resource as a result of the large amount of communication among nodes.

Aiming at the characteristics of the change of Cloud Computing load [5], this paper first proposes the layered load balancing scheduling model on the basis of the analysis of the load-balancing scheduling model [6],[7], then brings about the structure of this system, and at last designs and achieves a type of resource distribution algorithm which comprehensively taking the number and the performance of relative local agent resource as well as the current load situation into account.

3. SCHEDULING ARCHITECTURE

3.1 Layered scheduling architecture

According to the pros and cons of the above mentioned two kinds of scheduling modes, we have brought out a layered scheduling model,in which the global agent is in charge of collecting load information of the relative local agent and all of the resource is submitted to the global agent, but different from centralized scheduling, not all of these tasks are saved in the global agent resource submitted queue waiting for scheduling, but are directly assigned to local agents by global agent in accordance with load balancing and scheduled by local agents. Thus the global agent will not interfere with the resource and its load reduce , which avoids becoming the bottleneck in the system with its less resource waiting time, in order to achieve a simpler realization than distributed scheduling. From the view of the whole Cloud Computing system, taking centralized scheduling in local parts and the distributed scheduling in global ones would not only maintain the advantages of centralized scheduling, but also make up for the deficiencies of it in the use of distributed scheduling on the overall situation layered

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

scheduling system structure is composed of the following parts:

1) resource process Module, receiving resource requests and achieving a reasonable dynamic resource composition according to the status of each node portfolio of resources.

2) Scheduling Module, a receiving module to a distributor, in charge of dynamic collection of load information on various nodes, setting up the distribution levels of nodes and transmit the information to resource distribution module on a regular basis by the analysis on the performance of node, node CPU utilization, memory usage and I / O usage, and so on. Monitoring module, monitoring whether the local agent overload or delay too long to start re-scheduling strategy.

4) Transmission module, transmitting information of each node and integrated performance level of resource by the use of HTTP transmission based on Cloud Computing technology, since the majority of Internet firewall and proxy will not undermine the HTTP transmission.

3.2 Layered Scheduling Algorithm

Layered scheduling algorithm(LSA) which is based on Cloud Computing can solve a mass of resource's finding problem. LSA can be shown as followed:

$$f^*(n) = g^*(n) + h^*(n) \quad (1)$$

f*(n) denotes the cost of best resourcewhich crosses from start to the goal cross node n; g*(n) denotes the cost of best resourcewhich crosses from start to the node n.; h*(n) is the cost function of the best resourcewhich crosses from node n to the goal. h*(n) is not a fixed value but a circle function. If the original value can't find the goal, then h*(n) will search again based on new circle condition until meets the circle condition. Output the first resource which is the best for single node, then record all resource that be found in circle searching. When the map's environment changes, if the first resource can't lead to the goal, it needn't compute new resource at once, because it can choose the other paths, which are in the record to detect whether it can get to the goal. If all of the paths can't get to the goal because of the changed environment, then call LSA again to search the goal in the new environment. So LSA can reduce CPU computing resource. And LSA is right for colony node, because it supply multi-resources in order to avoid the jam during searching.

3.3 Realization of synthetic scheduling

In Cloud Computing environment, there are a great many Cloud Computing nodes and resource and a much complicated matching relationship between resource application and resource. For example, a user may contain a resource request and multiple ones as well; some individual resource may be composed of multiple sub-resources which having some kind of dependent relationship among them or independent of each other. Diversified resource requests submitted by different users may be implemented in the same Cloud Computing resource, forming the relationship of competition; also, a resource processing may need to simultaneously or successively use multiple Cloud Computing resource. The relationships among resources and resources, resource and Cloud Computing resource nodes, resource nodes and nodes will eventually affect the synthesis of Cloud Computing.

When users submit resource requests to Cloud Computing system, Cloud Computing resource location and a copy of positioning will first check appropriately to determine what kind of sub-resources existing in Cloud Computing system and what to be re-developed if there is no

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

direct resource resource available to Cloud Computing system while the requested resource associated by a number of related resources, invoking directly the resources existing in Cloud Computing system and redeveloping those non-existent ones following the requirements of Cloud Computing system(The specific process of developing Cloud Computings is omitted here) Secondly, the user's request is to be described as a process composed of multiple sub-resources with the use of pre-defined correlation structure, a sequential structure forming when a resource needs the implementation result of a related resource as its prerequisite; a cluster structure constituting if a resource requires implementation results of more than one resource as its prerequisite, and a branch structure forming if the implementation result of a resource works as prerequisite of other two or more than two resources.(all sub-resources being available in this article)

Through the identification of associated structures among all sub-resources, match scheduling can be performed according to the resource to be implemented. In Cloud Computing environment, because of the large-scale characteristic of resource and the copy management of resource as well, there may be more than one requested resource resource to meet some requirement, which may obtain different abilities and pay different costs during the implementation in the Cloud Computing resource at the same time, ie., there exists difference in the quality of resource offered by the same qualified resource resource, which is significant sometimes. Therefore, the optimal match scheduling of resource should be taken into account for Cloud Computing synthesis resource scheduling. After match scheduling done, resource can be performed, whose implementation process being under the control of the local resource management system. Following the implementation process, the occupied resource should be returned to internet resource management part, while the Cloud Computing management module deliver performance results and related information to the submitter or directly to the next resource node to be implemented the, with the use of RSL language to describe the resource information and related parameters to transmit information.

In summary, this paper presents the strategy of Cloud Computing synthetic resource scheduling as the following:

To determine the sub-resources included in the resource submitted by users;

To determine what kinds of sub-resources already existing, what to redevelop according to Cloud Computing demand and how to develop.

To determine the associated relationship among the various sub-resources, with scheduling process including:

a. In accordance with the implementation of the different resources from the execution resources to be implemented, adjusting the order of resource scheduling to make different resources matching carried out at the same time, thereby reducing the processing time of Cloud Computing's request on users, known as order schedule.

b. finding the appropriate resource resource for the resources to be performed (ie, select the copy resource), known as the match schedule;

c. implementing the matched resources, followed by transmitting scheduling request and data information to the resource to be scheduled.

4. EXPERIMENT AND THE ANALYSIS OF RESULTS

In this paper, the project kit, Cloud Computing Simtoolkit4.0 [8], has been used in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the simulation experiment, mainly because Cloud Computing Sim[8] acting on the simulation test focusing on the scheduling strategy in the Cloud Computing environment, providing the various basic function components of Cloud Computing and simulating the various basic actions of the function components, which making the developers achieve scheduling simulation with ease by this simulation tool. And related Cloud Computing association has been simulated. In the simulation experiment, response time has been compared between running and turning means of the resource distribution algorithm and that mentioned in this paper considering load balancing, the two curves are as follows, (shown as figure 1)

a: the theoretical value

b: means of resource synthetic scheduling presented in this paper

The Experiment shows that the synthetic scheduling method proposed in this paper, ie., resources being dynamically allocated, can enable all local agents load balanced, accelerate the speed of resource scheduling shorten efficiently the completion time of resource, and reduce the impact of resource delay on the improving the parallel efficiency of overall Cloud Computing system

Figure 1: The comparison chart of theory and practical

5. CONCLUSION

In this paper, a method of Synthetic Scheduling supporting Cloud Computing is proposed in this paper, followed by the analysis of the specific model and technology related to Cloud Computing, together with a layered scheduling model and the structure of load-balancing system, and then a method of Synthetic Scheduling comprehensively considering the task number and current load performance of various local agents has been designed and realized. The main purpose of this method is to improve the efficiency of the implementation of Cloud Computing system, by which to solve the problems of Cloud Computing system such as scheduling delays and so on. Finally, the superiority of the method proposed in this paper to other methods is verified by the experiment.

REFRENCES:

[1] J.Andrew, T.J.Naughton: Dynamic task scheduling using genetic algorithms for heterogeneous distributed computing. In Proceedings of the 9th International Workshop on Nature Inspired Distributed Computing, IPDPS, April, (2005):p.189-197.

[2] Brian Hayes: Cloud Computing Communications of the ACM. Vol. 51, July (2009), p.9 — 11

[3] C Dary,l. Plummer, W.David: Cloud Computing Confusion Leads to Opportunity. Gartner Research. Jun. (2008)

[4] M.Maheswaran, H.J.Siegel: A Dynamic Matching and Scheduling Algorithm for Heterogeneous Computing Systems. In Seventh Heterogeneous Computing Workshop,IEEE Computer Society Press, (1998): p.57-69.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[5] Y.X.He, Z.Y.Liu and A.L.Deng: Load distribution and balancing strategy of distributed system based on workstations. Computer Engineering, (1999), 25(11):p.11-13.

[6] D.M.Li, H.H.Shi and L.Q. Gu: Layered load balancing scheduling model based on Rules .Computer Science, (2003),30(10):p.16-20

[7] L.M.Gu, Research on Cluster Server load balancing technology [J]. Micro Computer Information, 2007, 23(12) vol: 20-23.

[8] Buyya, R., Manzur, M. Cloud ComputingSim: A Toolkit for the Modeling and Simulation of Distributed Resource Management and Scheduling for Cloud Computing Computing[J]. The Journal of Concurrency and Computation: Practice and Experience, Volume 14, Issue 13-15, Wiley Press, 2002, 14(13-15):1175-1220