2013 International Conference on Computational and Information Sciences

Job Scheduling Algorithm In Cloud Environment

Hongbo Yu
School of Computer Engineering
Huaihai Institute of Technology
Lianyungang, China
blackbull99@163.com

Yihua Lan*,Xingang Zhang, Zhidu Liu
School of Computer and Information Technology,
Nanyang Normal University
Nanyang, China
lanhua_2000@sian.com,ZXG550@163.com,
nyliuzd@126.com

Chao Yin, Lindong Li
School of Computer Science and Technology
Huazhong University of Science and Technology
Wuhan, China
yinchao408@gmail.com, gogo1207@qq.com

***Abstract*—This paper researches on the cloud computing technology systematically and comprehensively, and proposes a scheduling algorithm in the cloud environment based on the Berg model, which is a sociological justice distribution theory. This algorithm adapts to the commercialization and virtualization features of cloud computing, differs from traditional job scheduling algorithm's character - focusing on efficiency, and establishes dual fairness constraints under the cloud environment. For tasks with multiple Qos properties, this paper gives multiple generally expect in processing tasks by applying people element analysis theory. Job scheduling simulation program has been achieved on the extend CloudSim platform, which is recompiled and generated. And the algorithm simulation validation and comparative analysis of the experiments showed that the algorithm can effectively perform user tasks meanwhile reflect better fairness.**

Keywords-Cloud Computing; Scheduling Algorithm; fairness constraints; CloudSim

I. INTRODUCTION

Cloud computing [1] is a development of grid computing [2], parallel computing and distributed computing. It is an emerging commercial computing model, and has a good developing prospect. Grid computing, which generally accompanied with large-scale computing needs, is a product of a computing model to integrate the scattered resources and enable resource sharing and collaborative work, its appearance solves a lot of complex issues in many field.

In the end of 2007, cloud computing began to develop. It distributes the calculating tasks in a resource pool of a large number of computers. That gives a way for various application systems to get computing power, storage space and a variety of software services according to the needed. Cloud computing is a business model that packages various resources from data center into an outside services through the Internet.

The purpose of the job scheduling [3] in cloud environments is to achieve optimal scheduling of jobs submitted by users, and to maximize the throughput of the system. The main objectives are the optimal span, quality of service, load balancing and economic principles:

1) Optimal span

Span, which is the most important and common targets, refers to the length of scheduling, which is the running time from the start of the first task to the finish of the last one. Optimal span is the common goal of the user and the grid system.

2) Quality of Service(QoS)

While the grid system provide users the computing and storage services, user's demand for resources reflect from QoS. During task scheduler, the task management and scheduling system must protect the grid application's QoS.

3) Load balancing

Load balancing [4] is a critical issue in the development of parallel and distributed computing applications. Grid task scheduling involves cross-domain and large-scale applications, load balancing direct impact on the utilization of system resources.

4) Economy principles

Resources are widely distributed geographically in a grid environment, and each resource belongs to a different organization, which has its own resource management mechanisms and policies. According to the principles of market economy in real life, the use of different resources does not have the same cost. Resource management and task scheduling driven by market economy must make the consumer both sides (resource users and resource providers) mutual benefit, that's the principle to maintain the stability of the grid system.

By using simulation tools CloudSim and combining the basic mechanism of scheduling algorithm, the paper conducts in-depth discussion and analysis of key technologies in building scheduling mechanism based on cloud platform. The contributions of this paper are described as follows:

* Corresponding author.

978-0-7695-5004-6/13 $26.00 © 2013 IEEE
DOI 10.1109/ICCIS.2013.432

• By introduction of the sociology wealth distribution theory - Berg model of distributive justice, two fairness constraints are appended in the cloud environment : First, the selection process of resources using generally expect constrained. Second, the fairness of the allocation judgment and model self-correcting using allocation results fairness evaluation function. By using those constraints, resource allocation and job scheduling are achieved in cloud computing environment.

•The idea that job scheduling algorithm derived from the fairness perspective of using resources are introduced, which makes up traditional job scheduling algorithm that characteristically focuses on efficiency;

• Analyzing and expanding the simulation platform CloudSim. Comply job scheduling algorithm simulation on an expanded and recompile platform.

II. RELATED WORK

A. Cloud computing

From the end of 2007, cloud computing began to be widespread concern. Google, IBM, Microsoft, Amazon and other IT vendors promote cloud computing technology and the popularity of products with unprecedented speed and scale. Cloud computing is the development of parallel computing, distributed computing and grid computing, is commercial implementations of these computer science concepts. Cloud computing is the hybrid evolution and jumped results of concepts such as virtualization technology, utility computing and so on.

Cloud computing is Internet-based super computing model that centralizes together large amount of information and processor resources on the PC, mobile phones and other devices, and dispatch them working together. It is an emerging method that shares infrastructure. The method can connected huge pool of systems together to provide a variety of IT services. Many factors driving the demand for this type of environment, including connected devices, real-time data streams, SOA adoption, as well as search, open collaboration, social networking and mobile commerce web2.0 applications' dramatic growth. In addition, the improving performance of digital components makes a significant increase in the size of the IT environment, so as to further strengthen the need of unified cloud management.

Cloud computing is dragged by the scale economies, provide for external users on the Internet a group that has features such as abstract virtualization, dynamically scalable, computing resources management capacity, storage capacity, platform and services calculation. It has the following definition: "cloud computing distributes computing tasks in a resource pool constitutes by a large number of computers, various application systems can allocate the calculate power, the storage space and a variety of software services according to the needs." [5].

Cloud computing is still in its infancy, there are numerous types of vendors in the development of different cloud computing services. Currently, there are serval forms of cloud computing applications as follows:

B. Job scheduling

The job scheduling is a dispatch method that assigns suitable task of suitable jobs to the appropriate tasks server. There contains a two-step process. Firstly, we select the job and the task in this job. Same with other task allocation work, job scheduling is also a complex task. Bad task allocation may lead to the increase in network traffic, the overload of some tasks server and lead to an efficiency decline, so on. Moreover, the distribution of tasks or a consistent pattern of different business may require different algorithms in order to meet the demand. Job scheduling becomes independent as a plugin component, allowing the user to provide different implementations according to their needs.

1) FIFO scheduling

In Hadoop [6] job scheduling, JobTracker plays a very critical role. JobTracker is a master control node running the MapReduce application process, its core function is job scheduling, which control the job execution order and the allocation of computing resources. That control is directly related to the Hadoop platform's overall performance and system resource utilization.

2) Fair Scheduler

The fair algorithm is designed thinking as possible to ensure that all jobs are able to obtain the same amount of the shared resources. When there is only one job in executing in system, it will monopolize the entire cluster, and the use of all computing resources.

3) Capacity Scheduler

The computing power scheduling algorithm (Capacity Scheduler) job scheduling algorithm is proposed by Yahoo, which provides similar fair scheduling algorithm. But there are many differences in the way of design and realization between the two algorithms [7].

III. THE JOB SCHEDULING

A. The Design of the system

The Berg model is a theory that individuals obtain their distribution fairness through the comparison with common people's distribution in social system structure, where reference to the common distribution is a kind of common justice (IE, fairness) universally acknowledged. The justice theory of social distribution is mapped to the resource allocation model in the cloud environment. That provides a new idea to job scheduling in the cloud environment. The entities of cloud system are users, resource providers and scheduling systems. Their corresponding main bodies are user tasks, resources and scheduling policy. Therefore, cloud computing needs definition of the following concepts: the meaning of user tasks' fairness, task classification, general expectations, parameterization of tasks and resources, mapping of resources, etc.

The first definition is the fair judging constraints of the resources allocating results. In Berg model, social individual's justice (fairness) is defined as a distribution relationship generating from the comparison with common people's justice distribution. That process is quantified as the ratio of the actual and expected distribution. In cloud

computing, firstly, we should define the meaning of fair task. Cloud computing provides users on-demand services through Internet accompanied by on-demand payments. According to Berg model's principle [8], the fairness of resources used in the cloud system is embodied in the ability that cloud system can provide reasonable and available resources to different users according to their needs and task characteristics. Different users can get resources and satisfy their needs. We give following definitions in the cloud environment:

Definition 1. *Given resource amount actually allocated by user task and is represented by AR.*

Definition 2. *Given resource amount reasonably expected by user task according to the executing characteristics and is represented by ER.*

Definition 3. *Given the maximum consistent between the actual resources and the expected amount. That is called the fair execution of tasks and is represent by J.*

A judging function about task T is set as follows:

$$J_i = \theta \ln\left(AR_i / ER_i\right) \quad \textbf{Error! Reference source not found.} \tag{1}$$

θ is an equilibrium constant, and $0 < \theta < 1$, AR_i is the actual allocated resource amount for task T_i, ER_i is the expected resource amount for task T_i. When AR_i and ER_i are equal, that is, the actual allocated and expected resource amount consistent. The function value is zero, which means justice (fair); The time when $J_i > 0$, which means task T_i get too much allocation, is called upper-injustice; Similarly, the time when $J_i < 0$ represent that task T_i receive little allocation is called under-injustice.

Definition 4. *The Justice of the system (system Justice SJ): Assuming that system task sets are T= {T1, T2... TN}; Corresponding fair judging function sets are J= {J1, J2... Jn}.*

Then the system's fair judging function is defined as:

$$J = \sum_{i=1}^{n} \left| J_i \right| \tag{2}$$

It can be inferred that the justice of system constraint can be implemented as follows:

$$\min J = \min \sum_{i=1}^{n} \left| J_i \right| \tag{3}$$

The system performs best when J obtain the minimum value, which represent the maximum fairest achieved to every system user. Parameter J can be used to optimize the overall justice of the system。

B. Algorithm Description

Algorithm flow is shown in Figure 1. The main program is divided into several major subroutine instructions

(1) Task sorting algorithm

Functions: Using a fast sorting algorithm to prioritize the task list.

(2) Task parameterization and classification

This algorithm expands the Cloudlet class in simulation tools by adding main task element descriptive variables and set their relative setting and access methods. ClassType variable is also added so the tasks' classification parameters can be assigned when users submit their task.

(3) VM parameter normalization algorithm

Functions: VM parameters normalized

(4) Matching algorithm

Functions: According to task classification, the general expectations of various task categories correspond to parametric general resource vector. That matches a binding between task and VM.

(5) Calculating of fair evaluation function

Functions: Fairness judge of resource allocating

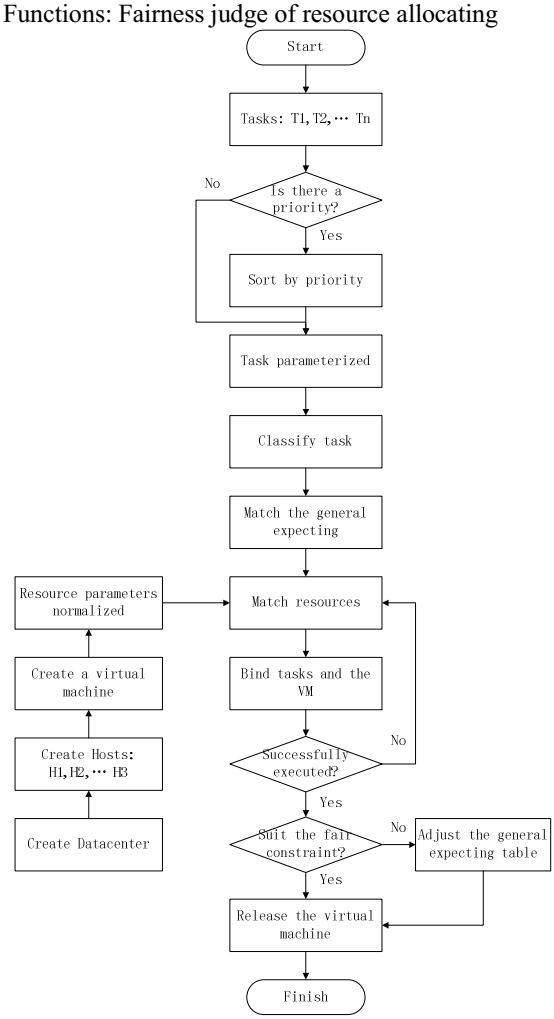

Figure 1. The algorithm description.

IV. EVALUATION METHODOLOGY

A. Experimental Setup

Data processing simulated by CloudSim [9] is shown in Figure 2. When datacenter resources are created, they

register to CIS; the process of information exchange is managed by Datacenter_Broker.

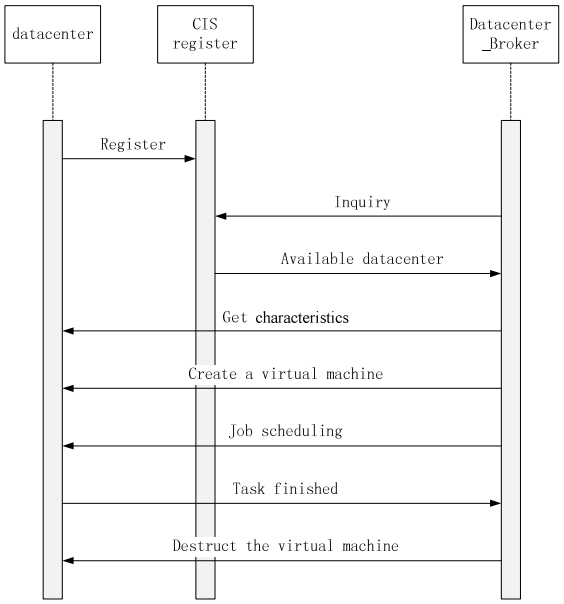

Figure 2. The experiment simulated by CloudSim.

B. Test Result

Due to the limitations of basic classes in CloudSim simulation platform, the failure rate of resources cannot be directly read; Also communication constraints between platform and emulator program makes it complex to calculate the cost of virtual machines, so we use the simulation scheduling of first class (completion time) and second class (bandwidth) tasks to validate the algorithm model.

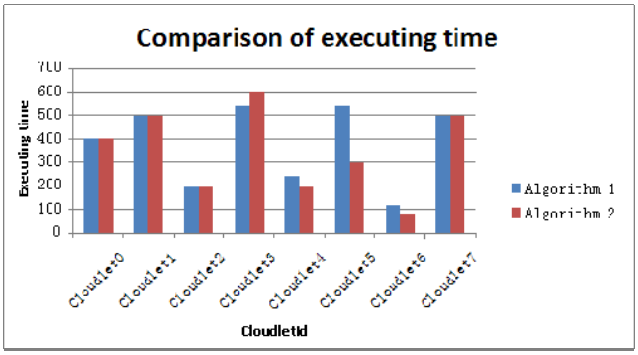

Figure 3. Execution time of different algorithm.

Analysis of the experimental data results, as shown Figure 3. The algorithm 1 is our algorithm based on the Berg model and the algorithm 2 is optimal completion time algorithm. The comparison shows that overall efficiency of our job scheduling algorithm is slightly worse than the optimal completion time algorithm. However task 0 to task3 are high computing power preference tasks which completion time is better than using a job scheduler aiming at an optimal completion time.

V. CONCLUSIONS

This paper researches the Berg model theory of justice distribution in the social field, imports that theory structure to job scheduling in cloud computing, where the dual fair constraint is established. The main element of User tasks is classified according to QoS preference. And justice in allocating sources is constrained according to the general expectance recorded in the main element of the task description. There are lots of works to do in the future which are described in the following directions, such as learning of general expected vector in job scheduling algorithms and Job scheduling algorithm based on the Berg model and so on. There is also a vast space in refinement of the task classification and full support of various tasks classifications.

ACKNOWLEDGMENT

This study was supported in part by the Industry-Academia Cooperation Innovation Fund Projects of Lianyungang (Grant No. CXY1203), by Technology Research and Development Program of Lianyungang (Grant No. SH1223), by the Research plan for Basic and frontier Research of Henan(Grant No.122300410258, 122300410384, 132300410210), by the Huaihai Institute of Technology Natural Science Foundation (Grant No. Z2011036), and The Key Project of Henan(Grant No.112102210408).

REFERENCES

[1] Wikipedia. Cloud computing. Retrieved from http://en.wikipedia.org/wiki/Cloud_computing, 2011.

[2] I. Foster, Y. Zhao, I. Raicu, and S. Lu. Cloud Computing and Grid Computing 360-degree compared [c]. In Grid Computing Environments Workshop, 2008: 1-10.

[3] A.Negi and K.P.Kishore. Applying machine learning techniques to improve linux process scheduling. In TENCON 2005 IEEE Region 10, 2005: 1-6..

[4] Belabbas Yagoubi, Yahya Slimani. “Dynamic Load Balancing Strategy for Grid Computing”, World Academy of Science, Engineering and Technology, 2006.

[5] Luis M.Vaquero, Luis Rodero-Merino, Juan Caceres, et al. A break in the Clouds: Towards a Cloud Definition[J]. ACM SIGCOMM Computer Communication Review. 2009, 39(1): 50-55.

[6] J.Dean.Experiences with MapReduce,an abstraction for large-scale computation[R].In Integernational Conference on Parallel Architecture and Compilation Techiques,2006.

[7] Matei Zaharia,Dhruba Borthakur,Joydeep Sen Sarma,Khaled Elmelegy,Scott Shenker,and Ion Stoica.Job Scheduling for Multi-User MapReduce Clusters[R].Technical Report UCB/EECS-2009-55,EECS Department,University of California,Berkeley,Apr 2009.

[8] Berger, J., M. Zelditch, Jr., B. Anderson, and B. P. Cohen. Structural Aspects of Distributive Justice: A Status Value Formulation [M]. Sociological Theories in Progress (Volume Two). Boston: Houghton-Mifflin. 1972, pp. 119-146.

[9] Rodrigo N. Calheiro, Rajiv Ranjan, Cesar A.F. De Rose, and Rajkumar Buyya,. CloudSim: A Novel Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services. Technical Report, GRIDS-TR-2009.