Journal of Theoretical and Applied Information Technology
15th December 2012. Vol. 46 No.1

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

AN EVALUATION OPTIMIZATION APPROACH OF IAAS RESOURCE DISTRIBUTION BASED ON GENETIC ALGORITHM

[1]FU HONGXIA, [1, 2]ZHOU XIANGBING,

[1] Department of Computer Science, Aba Teachers College, Weichuan Sichuan, China 623002

[2] School of information Science and Technology, Chengdu University of Technology, Chengdu Sichuan, China 610059

ABSTRACT

In cloud computing, IaaS (Infrastructure as a Service) is the foundation of cloud service applications. It enables problems like resource allocation scheduling and bearing capacity to be further optimally solved. Therefore, a genetic optimized IaaS resource optimization evaluation approach is presented. This approach abstractly defines its 7 factors (properties) including usability of server, network performances, load balancing, anomaly notification mechanism, response time support, trusted security and measure payment, and explains quantitative calculation and expressions of each factor. And finally, tested by CloudSim, this approach is indicated to be feasible and effective for resource allocation.

Keywords: *IaaS Resources, IaaS Evaluation, Genetic Algorithm*

1. INTRODUCTION

Cloud computing is a new distributed computing mode in recent years. It is the further expansion of utility computing, network computing and service computing [1]. Thus users can flexibly and elastically use resources and services by using cloud computing. In the process of providing services, cloud computing achieve it mainly by virtualizing all kinds of resources. Usually, the service level provided by the cloud computing can be mainly divided into SaaS, PaaS and IaaS. Thereinto, SaaS provides specific application service, and that refers to the service quality and the investment return; PaaS provide application support for SaaS, so it provides a platform for SaaS's construction and deployment; IaaS do service package to the network hardware resources through the virtualization methods. And achieve the communication of all levels through SOAP and REST (Representational State Transfer). But the mainly cloud computing deployments are public cloud, private clouds and mixed cloud, etc.; Thus IaaS would be the basis of cloud computing application and deployment, and it provides cloud service [2]. Usually it offers service for computer, storage, network and other related IT facilities [3], and provides these facilities for different cloud service providers and consumers through virtualization methods. Therefore, many researchers studied related aspects such as: resources scheduling [4], trusted security [3], charge mode [5] [6], resource allocation [7], resource sharing and virtualization [8], and cost benefit [9], and so on. Combining these research foundations and literatures [10], this paper tries to find out an optimization evaluation method base on IaaS's 7 characteristics: usability of server, network performance security, load balancing, anomaly notification mechanism, and response time support, safe and reliability, and measuring payment (price). It provides the quantitative calculation formula and expression of each characteristic, and then establishes an optimization evaluation model base on these quantitative calculation formulas and expressions, and use genetic algorithm to do optimization analysis. Finally, it used CloudSim to test and show that this evaluation optimization method is effective.

2. EVALUATION FACTOR CHARACTERISTICS OF THE IAAS RESOURCE ASSESSMENT

IaaS is the basis for the operation of PaaS and SaaS. IaaS provides virtual server, data storage, database and other basic services to PaaS so as to deploy and run applications. In addition, it provides abilities for processing, storage, network and other computational resources, and it can access through a SOAP (Web services) or REST information. At the same time, IaaS cloud is an environment of full

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

script; it's quite suitable for creating a framework to complete assembly rapidly and it can expand or contract environment application according to resource demand. This kind of ability is the important value proposition of IT department nowadays, in addition to flexibility; another characteristic of this service is that it is pay per use, just like SaaS and PaaS.

Therefore, we can start the IaaS resources optimization evaluation from its 7 characteristics: usability of server, network performance security, load balance, abnormal notification mechanism, response time support, trusted security and measure payment [3-10], and they can be described as:

$IaaS_f = (IaaS_{ava}, IaaS_{ncg}, IaaS_{lb}, IaaS_{enm}, IaaS_{srt}, IaaS_{sc}, IaaS_{pf})$ (1)

(1) Usability of Server

Usability refers to the virtual machine of cloud computing (such as virtual machine monitoring equipment) and storage system provided by the cloud environment can be operated normally.

In a particular cloud environment, we say the available maximum virtual machine be provided is $\rho_{\max}$, and the minimum is $\rho_{\min}$. In a certain time T, the rented virtual machine number is $\rho_{I,}$ so the usability of server can be defined as:

$$IaaS_{ava} = 1 - \frac{|\rho_{\min} - \rho_i|}{\rho_{\max} - \rho_i} \quad (2)$$

(2) Network Performance Guarantee

This network performance guarantee indicates the packet loss rate, network delay and network jitter guarantee in the cloud environment. If a packet dropping rate is a% (less than 10%), data packet is b, and the used memory space is G, so the network performance can be defined as:

$$IaaS_{ncg} = 1 - \frac{G \cdot a\%}{b} \quad (3)$$

(3) The load balancing

Load balancing indicates a mechanism of the processing delay, throughput rate and access concurrent rate in the cloud environment. Usually, when a node's load exceeds its threshold, then the load balancing would help to spread it evenly to the free nodes.

If a load balancing tank node $i(\in \mathbf{N})$ does packet sampling to available service examples in a certain period of time. Then we say $q_i(k)$ indicates the k^{th} sampling value, sampling number is N, the total length of the queue is $q_{\max}(i)$, threshold value is $t_i(k)$; If encountering unusable examples, turn to other groups and continue sampling (turning ratio is p_i). We return and continue when the unusable experiment is repaired, so the node load balancing $IaaS_{lb}$ would be defined as:

$$IaaS_{lb} = \begin{cases} \dfrac{\sum_{i=1}^{N} q_i(k)}{N \cdot q_{\max}(i)} & p_i \le t_i(k) \\ \dfrac{\sum_{i=1}^{N} q_i(k) \cdot p_i}{|q_{\max}(i) - N|} & p_i > t_i(k) \end{cases} \quad (4)$$

(4) Anomaly notification mechanism

Anomaly notification mechanism refers to the notification time when infrastructure anomaly happens (such as virtual machine failure). Suppose anomaly number of a certain time (time unit can be set to be minute according to specific circumstances) is N_N.

(5) Supported response time

This response time refers to the response time that the service providers provided in the cloud computing environment, when the cloud service is applied. We use the frequency of time T (the same as the time in (1)) to indicate the supported response time.

(6)Safe and credible mechanism

In order to ensure that when using cloud service, customers' data and privacy can be effectively protected, and the whole cloud environment is not affected, we shall make sure the provided cloud environment to be safe and credible. Usually, we use 1 to indicate that it's safe, 0 as well. But in this paper, safety is not prominent, so we set 0, 1 according to the test.

(7) Payment measuring

Payment measuring means we can choose the cost of the corresponding IaaS configuration (such as the calculating unit, memory, storage space of

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

examples, platform types, etc.) of the virtual machine according to users' demands. This paper is tested by CloudSim, so we would choose the cost according to the test.

Now suppose there are n cloud services (CWS) satisfy users' demand in time T. We say Task indicates the cloud task, Op indicates the cloud task operation, and the common storage space is M, so the characteristics definitions are as shown in **Table 1**.

Table 1 IAAS Characteristics Assessment

Characteristic name	Quantitative definition	Requirements on evaluation characteristics	Weight	Adopt the (2.6) and (2.7) normalized processing
Server usability	$\sum_{i=1}^{T}\sum_{j=1}^{n}(IaaS_{ava})_{ij}$	good	w_1	$f(x_1)$
Network performance guarantee	$\sum_{i=1}^{T}\sum_{j=1}^{n}(IaaS_{ncg})_{ij}$	high	w_2	$f(x_2)$
Load balance	$\sum_{i=1}^{T}\sum_{j=1}^{n}(IaaS_{lb})_{ij}$	excellent	w_3	$f(x_3)$
Anomaly notification mechanism	$\sum_{j=1}^{n}\left(\frac{N_N}{T}\right)_j$	few	w_4	$f(x_4)$
response time support	$\sum_{j=1}^{n}\left(\frac{1}{T}\right)_j$	few	w_5	$f(x_5)$
Safety credible	$\sum_{i=1}^{T}\sum_{j=1}^{n}(IaaS_{sc})_{ij}$	high	w_6	$f(x_6)$
Measure payment	$\sum_{i=1}^{T}\sum_{j=1}^{n}(IaaS_{pf})_{ij}$	few	w_7	$f(x_7)$

In **Table 1**, $\sum_{i=1}^{7} w_i = 1$.

In order to unify the result values in **Table 1**, we use the following formula to do normalization.

$$IaaS.value = \begin{cases} \dfrac{IaaS.value - IaaS.\min}{IaaS.\max - IaaS.\min} & \text{if } IaaS.\text{max-}IaaS.\text{min} \neq 0 \\ 1 & \text{if } IaaS.\text{max-}IaaS.\text{min} = 0 \end{cases} \quad (5)$$

$$IaaS.value = \begin{cases} \dfrac{IaaS.\max - IaaS.value}{IaaS.\max - IaaS.\min} & \text{if } IaaS.\text{max-}IaaS.\text{min} \neq 0 \\ 1 & \text{if } IaaS.\text{max-}IaaS.\text{min} = 0 \end{cases} \quad (6)$$

3. IAAS RESOURCE EVALUATION MODE OPTIMIZATION BASED ON GENETIC ALGOR ITHM

In this section, we optimize the IaaS resource evaluation factors in the second section; consequently, the IaaS resources can be more reasonably applied. Namely, we will make each IaaS quantification factor in **Table 1** effective and mostly satisfy users' requirement for cloud services, and optimize the genetic algorithm [12]. Thus, we set up the IaaS resource evaluation optimization model according to **Table 1** as follow:

$$F(x) = \max \sum_{i=1}^{7}\sum_{j=1}^{7} w_i f(x_j) \Big/ 7 \quad (7)$$

s.t

$f(x_1)$, $f(x_2)$, $f(x_3) \in$ [0.92n,0.98n]; $f(x_4) \in$ [0.01n,0.1n]; $f(x_5) \in$ [100,10000];

$f(x_6) \in$ (the sum of 1)/(the sum of 0,1) $\in$[0.92,0.98];

$f(x_7)$ can be set randomly according to the test requirements .

Now, we optimize the $F(x)$ value according to the steps of the genetic algorithm. And according to the requirements of the genetic algorithm, we shall conduct IaaS evaluation factors to satisfy the rule of 0 and 1. Namely, according to the s.t in **Table 1**, if satisfy the given value, choose 1, otherwise choose 0. Thus, it meets the requirements of the genetic algorithm coding.

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

(1) Gene

Gene is defined as a group of integer values, and these values are composed by n cloud services. And then encode the IaaS resources evaluation factors obtained by these cloud services, as shown in Figure 1.

Op	Task	$IaaS_{ava}$	$IaaS_{ncg}$	$IaaS_{lb}$	$IaaS_{enm}$	$IaaS_{srt}$	$IaaS_{sc}$	$IaaS_{pf}$	CWS

Op	Task	$IaaS_{ava}$	$IaaS_{ncg}$	$IaaS_{lb}$	$IaaS_{enm}$	$IaaS_{srt}$	$IaaS_{sc}$	$IaaS_{pf}$	CWS

……

Op	Task	$IaaS_{ava}$	$IaaS_{ncg}$	$IaaS_{lb}$	$IaaS_{enm}$	$IaaS_{srt}$	$IaaS_{sc}$	$IaaS_{pf}$	CWS

Figure 1: Genetic Encoding Of Genetic Optimization

(2)Population initialization

Initialize the population according to (1), and pair them with the genes in Figure 1. Furthermore, replenish gene matching positions with random 0, 1.

(3) Regulate the constraint conditions

Set genetic iterative termination and constraint condition according to the s.t constraint conditions in **Table 1**. Local IaaS evaluation factor optimal value might appear when the optimal F(x) value emerges, so we shall balance the minimum evaluation factor weight.

(4) Fitness function

Fitness function is $F(x)$. It makes the genetic iteration to meet the constraint conditions, and adjusts the weight of minimum assessment factors.

(5)Genetic operation

Usually genetic operations including three parts: mutation, crossover and selection. Thereinto, mutation is an encoding way that can randomly change the genome according to the constraints, and we say the mutation rate is P_1; Crossover is to interchange the mutated genetic coding group and the current genomes, we say the crossover rate is P_2, and usually we say $P_1 > P_2$; When the value of fitness function F(x) can satisfy the requirements of the constraint conditions, selection will choose a relatively optimal value from the previous iterative to contrast the current iteration.

(6) Optimization termination

When genetic iterative meets the constraint conditions in (3), the entire genetic iterative would stop and enter into the next iteration operation.

4. EXPERIMENTAL TEST

This experiment is carried out based on CloudSim[2] (http://www.cloudbus.org/cloudsim/), and the genetic algorithm in this paper is realized by source code JGAP (http://jgap.sourceforge.net). Set up a cloud task counter, a timer, a data monitor, an anomaly monitor to respectively record the mission number, time, data interactive state and record the virtual machine anomaly frequencies when evaluating.

There are 4 reasons for us to adopt CloudSim[13]:

(a) It provides virtualized engine, namely it can help to establish and manage multiple, independent, synergetic virtualization services for data center nodes;

(b) When distribute and process the virtualization service, it can flexibly switch between time sharing and space sharing;

(c)It can do modeling for data center, service agent, scheduling and allocation strategy;

(d)It supports modeling and installation of large-scale cloud computing infrastructures, including the data center of single physical calculation node and Java virtual machine.

At the same time, in order to satisfy this test, we expand GridSim and SimJava simulation library, and bring JGAP library to CloudSim.

Genetic algorithm parameters are set as: iteration times are 500, genetic generation gap is 0.9, and genetic code number is 32. The initial populations are generated according to cloud service n, P_1=0.95， P_2=0.15; other parameters are randomly generated according to the test.

Now we test a group of 50, 100 cloud services in CloudSim. In Figure 2, it shows the IaaS resource utilization fitness value change curve when there are numbers of cloud services change as the iteration times. In Figure 3, it shows the IaaS resource utilization change curve of the method in this paper and exhaustive method under different iterations. The utilization ratio here refers to the rented IaaS utilization, and it manages the whole test environment of IaaS resources according to the CloudSim. It can be defined as:

$$\text{Utilization rate} = \frac{\text{IaaS resource state when tested}}{\text{IaaS resource state}} \quad (8)$$

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

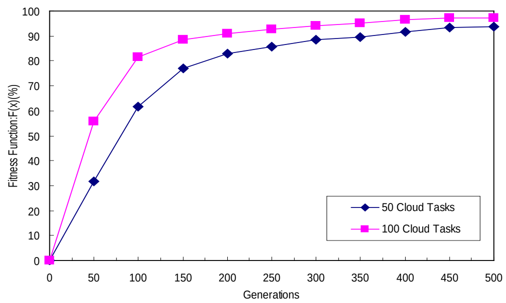

Figure 2: Iaas Resources Utilization Fitness Value Change When Different Cloud Service Number Change With The Iteration Number

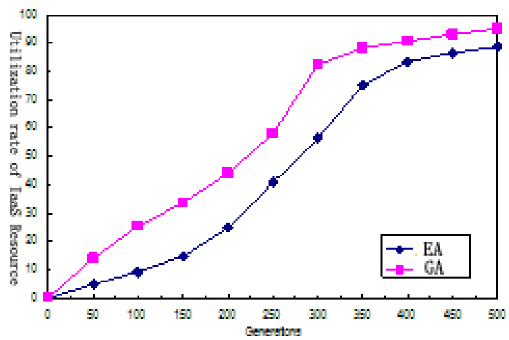

Figure 3: Iaas Resource Utilization Rate Curve Of Different Methods

5. CONCLUSION

From the IaaS resource state, this paper analyzed the 7 characteristics that influence the rational use of IaaS. In addition, it defined the specific quantitative method and expression according to the characteristics of each factor; then, in order to make better use of the rented IaaS resource, it also set up a genetic optimization model based on these 7 characteristics; at last, tested by CloudSim cloud computing mode, this IaaS resource optimization approach proposed in this paper is proved to be feasible and effective.

ACKNOWLEDGMENT

This work was supported by the Natural Sciences of Education and Science Office Bureau of Sichuan Province of China under Grant No. 2010JY0J41, 11ZB152.

REFERENCES:

[1] Thomas Rings, Jens Grabowski, Stephan Schulz, "Grid and Cloud Computing: Opportunities for Integration with the Next Generation Network", *Journal of Grid Computing*, Vol.7, No.3, 2009, pp. 375–393.

[2] Xin Siyuan, Zhao Yong, Lin Li, Wang Xiaohai, "Research on Trusted Attestation Method in IaaS Environment", *Computer Engineering*, Vol.38, No.5, 2012, pp.117-119.

[3] He Guofeng, Chen Gongchao, Li Jianhua, "Construct Safe and Reliable IaaS Platform", *Telecom science*, Vol.30, No.5, 2011, pp. 121-125.

[4] Yue Dongli, Liu Haitao, Sun Aobing, "Research on IaaS public cloud computing platform scheduling model", *Computer Engineering and Design*, Vol.32, No.6, 2011, pp.1889-1892.

[5] Baomin Xu, et al., "Job scheduling algorithm based on Berger model in cloud environment", *Advances in Engineering Software*, Vol.42, No.7, 2011, pp.419-425.

[6] Chee Shin Yeo, et al., "Autonomic metered pricing for a utility computing service", *Future Generation Computer Systems*, Vol.26, No.6, 2010, pp.1368-1380.

[7] Zhao Gansen, et al., "The Self-Adoption Resource Supply of Cloud Computing Platform", *Telecom science*, Vol.31, No.1, 2012, pp.31-36.

[8] Ravi Iyer, et al., "VM3: Measuring, modeling and managing VM shared resources", *Computer Networks*,Vol.53, No.17, 2009, pp.2873–2887.

[9] Tien Van Doa, Csaba Rotter, "Comparison of scheduling schemes for on-demand IaaS requests", *Journal of Systems and Software*, Vol.85, No.6, 2012, pp.1400-1408.

[10] Luo Junzhou, et al., "Cloud Computing, System Architecture and Key Technologies", *Journal on Communications*, Vol.32, No.7, 2011, pp.3-21

[11] Wu Jun, Xu Ming, "Comparative Study on Charging Mode of Public Cloud Services", *Telecom Science*, Vol.31, No.1, 2012, pp.127-131

[12] Danilo Ardagna and Barbare Pernicl, "Adaptive Service Composition in Flecible Processed", *IEEE Transactions on Software Engineering*, Vol.33, No.6, 2007, pp.369-384

[13] Saurabh Kumar Garg, Rajkumar Buyya, "Network CloudSim: Modeling Parallel Applications in Cloud Simulations", *Proceedings of the 4th IEEE/ACM International Conference on Utility and Cloud Computing*, December 5-7, 2011. pp.105-113.