

Performance Evaluation and Modeling Method Research Based on IaaS Cloud Platform

Jian Wan, Xianghong Yang, Zujie Ren* and Zheng Ye

renzj@hdu.edu.cn

Abstract

With the widespread use of cloud platforms, their performance evaluation tools also have become the research hot spot of academic circle. So far, many performance evaluation tools of the cloud platform have been designed in their corresponding application scenarios, which have brought much convenience on the performance evaluation and management of the cloud platform.

In order to predict the maximum number of virtual machines that can be opened by the cloud platform, this paper integrates the current tools of performance evaluation and proposes a performance evaluation tool based on IaaS cloud platform. The key of the performance evaluation tool is that it not only can evaluate the performance of the cloud platform, but also can predict the maximum number of virtual machines that can be opened by the cloud platform when the configuration of the virtual machine and the workload of each virtual machine have been known. This special performance evaluation tool has not been put forward now. And, the prediction model has been introduced into this tool in this paper that is the most important and core part.

Lastly, to test the effectiveness of cloud platform performance evaluation tool proposed in this paper, some tests have been done on the IaaS cloud platform. According to the contrast results of the forecast error among models, establishing support vector machine and neural network as single forecasting model. The results show combined model can be chosen as the prediction model of cloud platform performance evaluation tool.

Keywords: *cloud platform, performance evaluation tools, virtual machine*

1. Introduction

As a evaluation method of measuring the cloud platform, the benchmark evaluation of performance evaluates the performance of the cloud platform through a series of stress test. Currently, there are a lot of colleges and universities and research institutions have devoted into the research of the benchmark evaluation method of the cloud platform performance. Under the condition of not affecting users' experience or meeting the business' demands, the cloud platform is able to support the maximum workload and become the research objective. The difference of the cloud platform performance evaluation indicator also directly affects the evaluation of the cloud platform performance. Because the cloud platform's structure and operation method are different from the traditional physical machine, its benchmark evaluation is also different from the traditional evaluation methods.

According to the current performance evaluation tool or work, this paper will make a further research. When we know the configuration of the cloud platform, the configuration of the virtual machine and the workload of the virtual machine, is it possible to know the maximum number of the virtual machine which cloud be able to be opened by the cloud platform? In order to obtain this conclusion, firstly, this paper obtained the virtual machine's maximum number which cloud be opened by the cloud platform under

* Corresponding Author

the certain configuration and workload through test, however, the task load of testing the maximum number of the virtual machine which could be opened by the cloud platform under all situations of virtual machines and workload is obviously very large. This paper needs to design a cloud platform performance evaluation tool to extract the cloud platform performance data and the data and configuration of the virtual machine's performance, taking these data as the parameters of the prediction model in the cloud platform performance evaluation, finally, it concludes the maximum number of the virtual machines under the current virtual machine configuration and the current workload. Through this performance evaluation tool, it can plan the capacity of the cloud platform.

The first problem met by this paper is how to obtain some sample data to establish the prediction model of the cloud platform performance evaluation tool through test. When we know the configuration of the cloud platform, the configuration of the virtual machine and the workload, it is able to accurately predict the maximum number of the virtual machines which could be opened through this model.

The second problem faced by this paper is how to design and realize the performance evaluation tool of the cloud platform. This performance evaluation platform needs to simulate the real workload, extract and analyze the performance data of the cloud platform and the virtual machine. The configuration of the virtual machine as a part of the parameters of the prediction model, therefore, the performance evaluation tool needs to extract the configuration of the virtual machine without the human intervention.

Three contributions of our works are listed here:

First, a performance evaluation tool of IaaS cloud platform performance has been designed and implemented to solve the problem of cloud platform's capacity planning.

Second, the difference of this performance evaluation tool with other profiling tools, the tool proposed in this paper adds a prediction model that can predict the maximum number of virtual machines that can be opened by cloud platform.

Third, this tool not only can acquire the maximum number of virtual machines that can be opened by cloud platform, but also can solve some practical problems, such as know how much virtual machines should be opened. Because the configuration of virtual machine is fixed and the workload is also known, we can know the number of cloud platforms with one configuration that needed.

2. Related Work

Currently, there are a lot of performance evaluation tools and benchmark evaluation methods related to the cloud platform. However, these tools are rarely used for the capacity planning of the cloud platform.

In 2009, Binning and other four people [1] put forward a new benchmark method to test cloud characteristic (including expandability, usage cost and fault tolerance), in order to replace the traditional TPC benchmark methods. They pointed out that the cloud benchmark indicator shall reflect the cloud service's adaptability to the flexible load, rather than measure the average performance of the static system under the maximum load.

In 2010, Yahoo cloud service benchmark (YCSB) [2] frame compared the cloud data system performance and simple sharing MySQL realization, and these cloud data system were Cassandra, HBase, YahooPUNTS. YCSB evaluated every operation's throughput and response delay.

CloudCmp [3] uses the following metrics to measure the elastic computing, storage services and network services, including Computation Metrics (including benchmark completion time, cost per benchmark and scaling latency), Storage Metrics (including throughput, response time, the consistency of time and cost per operation) and the Network Metrics (including network throughput and path latency).

In Fair benchmark project [4], the researchers measure the throughput as the only metric for different types of IaaS workload, including memory I/O (MB/s), disk I/O (File/s, KB/s), CPU (GFlops), network (Mbit/s), and application (total running time).

In 2012, Folkerts, E. *et al.*, [5] put forward a meaningful and understandable cloud benchmark measure indicator. They suggested to use the median, average, maximum and minimum running time, meeting SLA throughput, and price/performance under parallel load as the cloud performance indicator during operation.

Later, there are more and more cloud platform benchmark evaluation methods. Smart CloudBench [6] is a cloud benchmark evaluation method proposing taking price as indicator. In the method of Cloud Crawler [7], users can constitute the corresponding key value pairs by declaring measure indicator and SLA demands. Kinnison put forward using genetic algorithm [8] to evaluate the cloud platform. Herbst *et al.*, put forward BUNGEE [9] to evaluate self-adaption IaaS cloud environment.

The above methods defined the indicators of evaluating performance and the methods of benchmark evaluation only under the specific application scenarios.

Rain [10] is a load generation tool. The purpose of Rain is to generate the more realistic workload and more elastic set profile and generate workload. Afterwards, it provides capacity through Oilo testing the service of cloud platform. However, it can not collect the source information of server.

Because the relative high use ratio of Rain in internal storage, Johannes put forward Cloudburst [11] to reduce the usage of the internal storage, and the concurrency is better. Their purpose is to simulate the more realistic workload to support concurrent generation workload. However, they did not consider the generated workload without affecting the maximum number of virtual machines which are opened by cloud platform performance.

CloudRank-V [12] mainly tests the performance of the desktop cloud platform through some set complex workload. Monitor the average throughput of each virtual machine in the cloud platform through Packets Monitor, and determine the maximum number of the virtual machine which could be supported by the desktop cloud platform through the deterioration rate of the throughput. With regard to the value of the deterioration rate of the throughput presented by the author of this article, the author did not explain why is this value. CloudRank-V also did not make a capacity plan to the desktop cloud platform.

3. Modeling

This section mainly describes the selection of independent variables and modeling model before modeling.

3.1. The Selection of the Independent Variables

Firstly, this paper determined the maximum number of the virtual machines which could be opened through opening the same type's virtual machines of cloud host type on the cloud platform. Later, it starts the virtual machines of different cloud host type. The results are shown in Figure 1.

It can be seen from Figure 1, that with the promotion of the configuration of each virtual machine, the maximum number of virtual machines which could be opened by the cloud platform presents the downward trend. In here, the maximum number of virtual machines which could be opened by the cloud platform is obtained through test under the situation of empty load. From the figure, it can see that the size of internal storage limited the opening maximum number of virtual machines. Based on this test result, this paper takes the configuration of virtual machines in the cloud platform as a factor of influencing cloud platform opening maximum number of virtual machines.

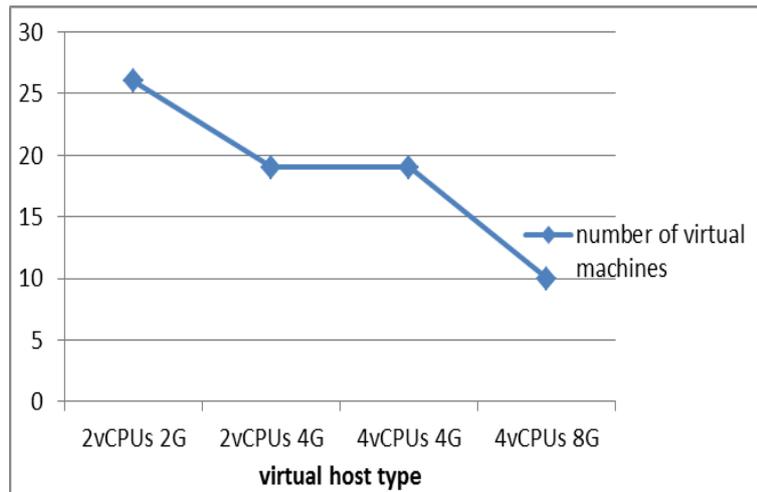


Figure 1. The Maximum Number of Virtual Machines which Could Be Opened by the Cloud Platform under Different Cloud Host Types

After determining the influence of the cloud platform opening different virtual machines' configuration to the cloud platform opening maximum number of virtual machines, controlling the cloud platform opening the virtual machines with the same configuration, it respectively tested the maximum number of virtual machines which could be opened by the cloud platform under different workloads. The test results is showed is Figure 2.

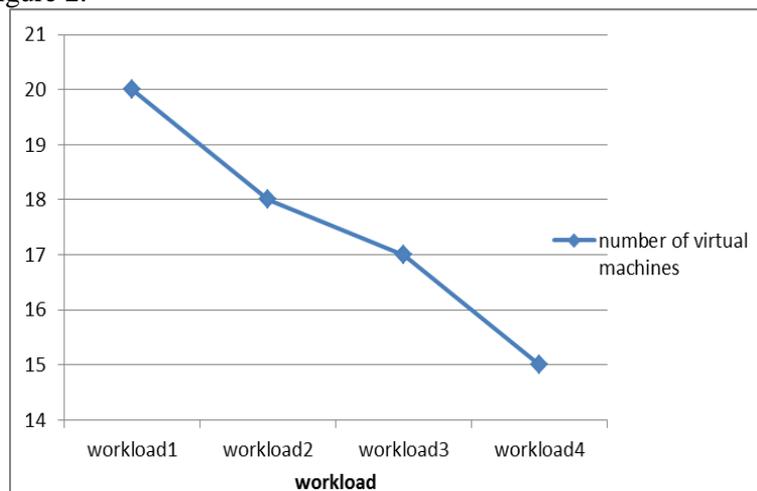


Figure 2. The Maximum Number of Virtual Machines which is Opened by the Cloud Platform under Different Workloads

The result of Figure 2, is measured under the premise of same configuration of virtual machines. The workload of the right side of abscissa in the figure is greater than the workload of the left side. With the constant increase of each virtual machine's workload in the cloud platform, the maximum number of virtual machines which could be opened by the cloud platform presents the downward trend. In here, the maximum number of virtual machines which could be opened by the cloud platform is measured under performance indicator, and the performance evaluation indicators will be illustrated in the experimental part. Through analyzing this test result, this paper takes the virtual machines' workload as the other factor influencing the maximum number of virtual machines which is opened by the cloud platform.

Through experimental measurement, this paper determined that the maximum number of virtual machines which is opened by the cloud platform is influenced by the virtual machines' configuration and workload in the end.

3.2. Establishment of Model

The work is to obtain the maximum number of virtual machines which is opened by this server through knowing the virtual machines' configuration and workload in server. This paper quantifies the workload into the source of consuming virtual machines by consuming the virtual machines' CPU and internal storage. Finally, it takes the size of virtual machine's internal storage, the size of consuming each virtual machine's CPU, the size of consuming each virtual machine's internal storage as the model's independent variable, taking the maximum number of opening virtual machines as dependent variable. During the experiment we found that virtual machine's CPU size does not become a bottleneck to open the largest virtual machine. Therefore, this paper is no virtual machine CPU size as factors. The obtaining of sample data and indicators will be illustrated in the experimental part. Selection 2 vCPUs, 2G RAM and 2 vCPUs, 4G RAM in part of workload that the maximum number of virtual machines can be open as fit data.

With the increase of the virtual machine configuration and the increase of the working load, it can be concluded that the number of the virtual machine is on the downward trend from the sample data. The number of open virtual machine has the characteristic of nonlinearity under different virtual machine configuration and working load. Based on these two characteristics, this paper chooses the support vector machine model [13], BP neural network model to establish the single forecast model, which two models are suitable for small data samples. Because of the hypothetical conditions of a single model and the limitation of scope application, the standard deviation method is used to make the weight distribution, and the combination forecasting model is established.

In order to overcome the defects of single model and reduce the randomness of prediction, the combination forecasting model is established. We use the standard deviation method to determine the weight of combined model. Suppose the standard deviation of the support vector machine and neural network model prediction errors were σ_1, σ_2 , and

$$\sigma = \sum_{i=1}^n \sigma_i \quad (1)$$

(i = 1,2,3), then

$$w_i = \frac{\sigma - \sigma_i}{\sigma} \cdot \frac{1}{n-1} \quad (2)$$

i= 1,2,3, n is the number of model. The weights of each individual model were calculated as w = (0.509, 0.405), according to the combined weights, the combined forecasting model was established as follows:

$$y = 0.505y_1 + 0.495y_2 \quad (3)$$

In the formula, y is the combination forecast value, y_1 is the support vector machine forecast value, y_2 is the neural network forecast value. To obtain a single model and combined model with the sample data through R language, fitting results shown in Figure 3.

3.3. Model Verification

Using 4 vCPUs, 4G RAM and 4 vCPUs , 8G RAM in part of workload that the maximum number of virtual machines can be open as forecast data, the Model predictive indicators are shown in Table 1, predicted effect is shown in Figure 4.

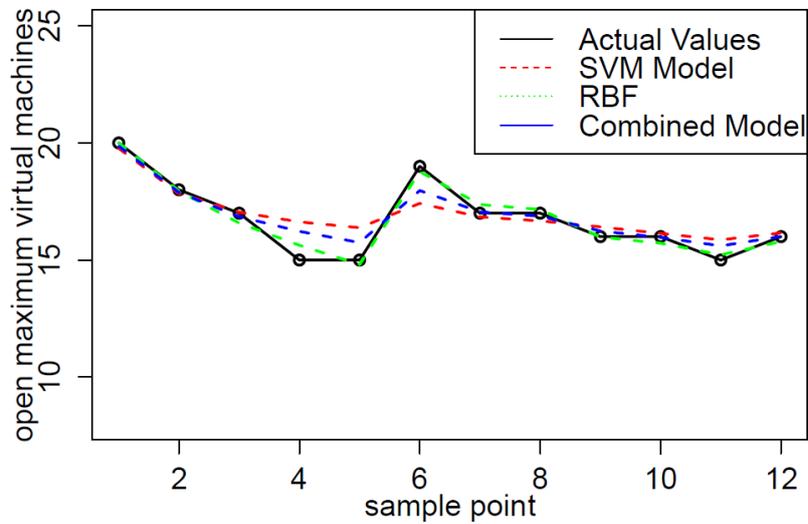


Figure 3. The Fitting Effect of Model and Sample Data Model Verification

Table 1. Model Predictive Indicators

Model	MAE(Mean absolute error)	MRE(Mean relative error)	MSE(Mean square error)	RMSE(Root mean square error)	AvgPD(Average percentage deviation)
SVM Model	1.613	0.146	0.699	2.209	14.613
RBF Model	2.129	0.166	0.779	2.464	16.685
Combined Model	1.782	0.152	0.693	2.191	15.260

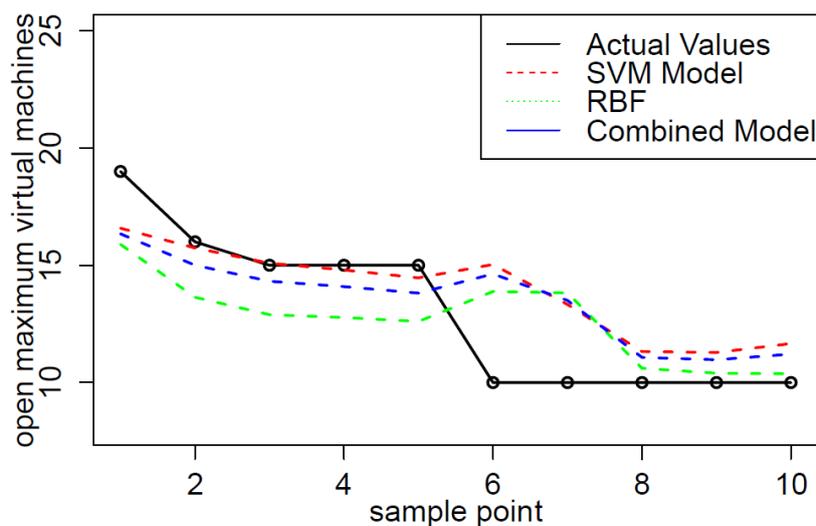


Figure 4. The Prediction Effect Figure of Support Vector Machine

Table I, can be obtained that combined model's root mean square error value up to 2.191%, the other less than 2%, in addition to the average percentage deviation of 15.26%, which is due to the combination of the model there are some errors in the three predicted points. Figure 4, can be drawn from the model with the virtual machine configuration and workload increase, procedures downward trend. Combined Model prediction results were pretty good overall, the error in the sample points 6 and 7 are relatively large, which is due to cloud platform memory limits the number of open virtual machine can be open. The virtual machine memory affect the argument by the sample point, so that these two sample data did not affected by the workload characteristics, resulting in the deviation of the predicted value.

4. Design and Implementation of Tool

In this chapter, it mainly introduces the overall architecture of the cloud platform's performance evaluation tool, which is shown in Figure 5.

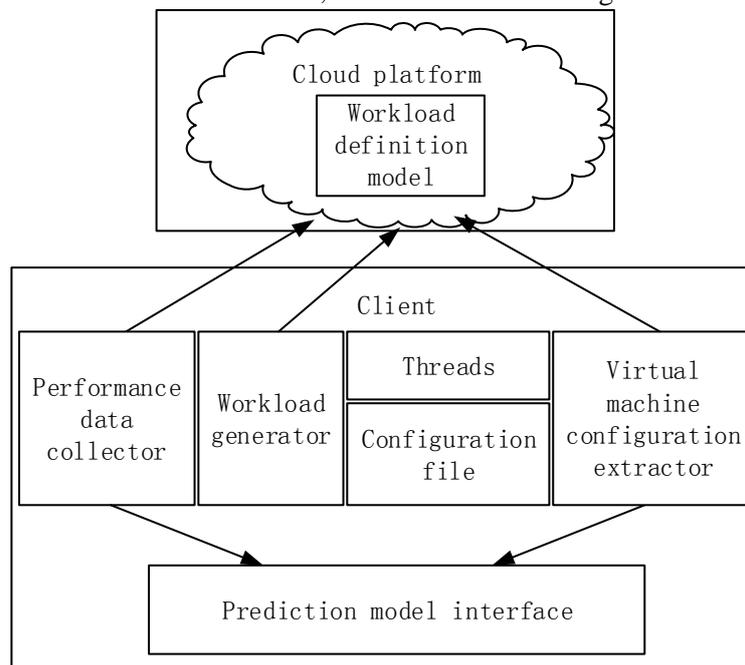


Figure 5. Architecture of Performance Evaluation Tool

4.1. Architecture

4.1.1. Workload Generator: The function of workload generator is to send command to the virtual machines to control workload, which is the cooperative work with configuration file. It can configure the IP address of virtual machine in the configuration file. Workload generator sends command of starting load to the corresponding virtual machines through the IP address of configuration file. It can configure the same IP in the configuration file, multiple IP being corresponding to workload scripts times in starting virtual machines, it changes the virtual machine's workload through this method.

The target system of performance evaluation tool is the workload in the workload definition module. Workload generator needs to a channel to start workload, therefore, the workload generator in this paper also includes the sound monitor in the virtual machine. It is responsible for monitoring the command sent through client-side, and it parse command and run the workload in the workload definition module when there is command. This paper controls and changes the virtual machine's workload through this set of mechanism,

and they communicate through socket. The sound monitor in the virtual machines improves the response speed of the command parsing through multithreading mechanism.

4.1.2. Workload Definition Module: Workload definition module can define a series users' operation, including a series of operations, such as open Word, browse Word, modify Word, open PDF, browse PDF, open PPT, modify PPT, run computer tasks, *etc.*, and the detailed description is shown in Table 2. These operations can make virtual machines generate workload. Workload generates module to control workload definition module. After the sound monitor in the virtual machine receiving the request of starting workload, it will parse request and start the predefined script in the workload definition module, and make virtual machine generate workload through script simulating users' operation.

Table 2. Workload Type and Detail Operation Description

Application	Operations
Adobe Reader	Open->Browse->Close
Microsoft Excel	Open->Fill in data->Save->Close
Internet Explorer	Open->Browse->Close
Windows Media Player	Open->Play->Pause->Resume->Close
Microsoft PowerPoint	Open->Browse->Append slides->Save->Close
Microsoft Word	Open->Browse->Modify->Save->Close
Calculate program	Open->Run->End

4.1.3. Performance Data Collector: The performance data collection includes the cloud platform's performance data collector and the virtual machine's performance collector. The virtual machine's performance collector is through registering service on the virtual machine, and this service can collect the virtual machine's performance data after starting this service. The virtual machine's performance collector mainly collects the use ratio of virtual machine's CPU and internal storage. The cloud platform's performance data collector collects through calling the existing connector of the cloud platform, and its collector will record one time of all performance data of cloud platform per second hand.

4.1.4. Virtual Machine Configuration Extractor and Performance Data Parser: Performance data parser parses the obtained performance data, including the cloud platform's and virtual machine's performance data. It parses this paper's required performance data through parsing script, and then determines the average value of performance data. This paper quantifies workload through the performance data in virtual machine. Finally, it sends the obtained virtual machine's performance data and virtual machine's configuration information to prediction model as the input parameters and the connector of calling prediction model.

4.1.5. Prediction Model: Prediction model obtains the maximum number of virtual machines which could be opened by the cloud platform through running the algorithm of prediction model and after receiving the parameter sent by the virtual machine configuration extractor and performance data parser. Prediction model is the core of cloud platform performance evaluation tool. Through the third part of establishment and verification of model selected combination model, its model formula is $y = 0.505y_1 + 0.495y_2$ (y is a combination of predicted value, y_1 is a support vector machine predictive value, y_2 is a neural network predictive value).

4.2. The Process of Forecast

The workload generator controls cloud platform workload module operation, when virtual machine is running, starting the virtual machine performance collector collects virtual machine and cloud platform performance data. End of collect performance data when set time arrives, and then start the virtual machine data parser begins to analyzing data to obtain the consume of virtual machine's CPU and memory, simultaneously, collecting virtual machine's configuration, these four values passed as a prediction model's parameter. Lastly, run a forecast model algorithms derived maximum number of virtual machines that cloud platform can be open.

5. Experiment

This paper's server equipment CPU is AMD Opteron 6272 2.1GHZ 32 nucleus, with storage being 64GB(DDR3 1600MHZ), hard disk being Toshiba MBF2300RC(300G), network card being Intel gigabit network card. The OpenStack cloud platform is deployed on this server, with the purpose predicting the maximum number of virtual machines which could be opened by the cloud platform. The cloud host platform type in this paper is shown in Table 3.

Table 3. Cloud Host Types

Cloud host name	Cloud host configuration
i5	2vCPUs, RAM 2G
m1.medium	2vCPUs, RAM 4G
big disk	4vCPUs, RAM 4G
m1.large	4vCPUs, RAM 8G

5.1. Sample Data Collection

This paper takes CPU of cloud platform and usage ratio of internal storage as the indicator of performance evaluation, and determines the use situation of CPU by collecting the "Load Average" of CPU. The experimental environment is gigabit local area network. It finds that the network did not become the experimental bottleneck through experiment, and the indicator of this paper did not take the factor of network into consideration. This paper conducted experiment according to these three indicators.

Firstly, the selected cloud host type of virtual machine's configuration is i5, namely 2 vCPUs and RAM 2G. It totally started 24 virtual machines within this configuration. Each experiment and each virtual machine's workload are fixed. This paper collects cloud platform's and virtual machine's performance data under different number of virtual machines. Figure 6 and Figure 7, are obtained by collecting cloud platform CPU and internal storage's performance data. Through parsing the cloud platform performance, it obtained cloud platform CPU and the average utilization ratio of internal storage.

It can know from Figure 6 and Figure 7 that with the increase of the number of virtual machines, cloud platform CPU and the average utilization ratio of internal storage are constantly rising. When there are 20 virtual machines, the utilization of cloud platform CPU is leveling off while the utilization ratio of internal storage is constantly rising. From this, we can see that the bottleneck of cloud platform is on CPU. In order to further determine the maximum number of virtual machines which could be opened under this virtual configuration and a certain workload, this paper collects "Load Average" of CPU, namely, the average load of cloud platform CPU. The obtained result is shown in Figure 8.

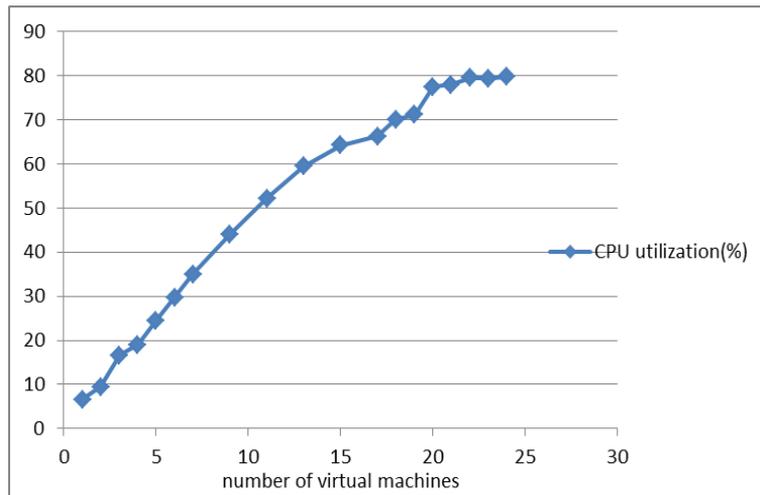


Figure 6. The Changing Situation of the Cloud Platform CPU

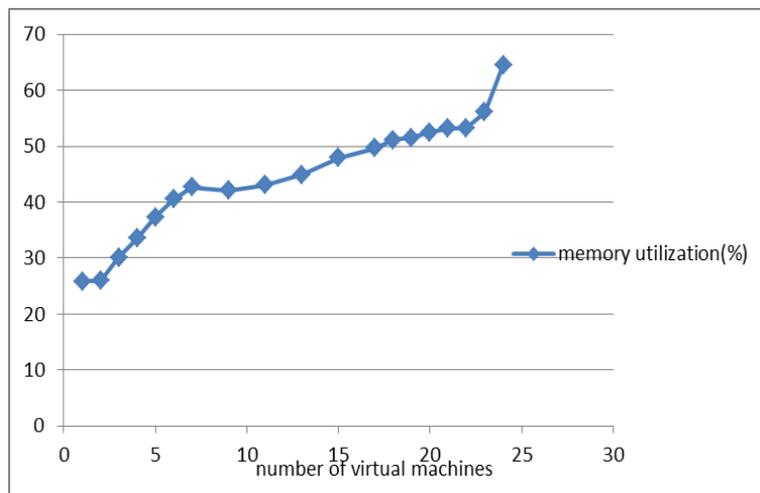


Figure 7. The Changing Situation of Cloud Platform Internal Storage

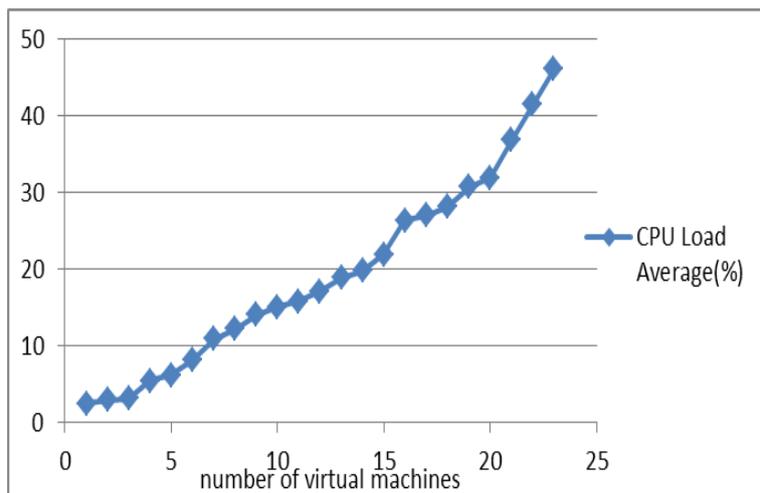


Figure 8. The Changing Situation of Average Load of Cloud Platform CPU

From Figure 8, it can be seen that with the increase of the number of virtual machines, the average load of cloud platform is constantly increasing. In order to guarantee the

running fluency of cloud platform, the average load of cloud platform system shall not be greater than the nucleus number of CPU, namely, this paper's average load shall not be greater than 32. From Figure 8, it can be seen that when there are 20 virtual machines, the average load of cloud platform is 31.97. Ultimately, it obtained the maximum number of virtual machines which could be opened by the cloud platform is 20 through combining Figure 6, and Figure 7.

There is a certain deviation by statistics for the CPU and the internal storage consumed by each virtual machine in the experiment result, however, the deviation is small. According to this method, selecting different cloud host type and controlling different workload, this paper obtained multiple groups of modeling sample data.

6. Conclusion

We researched a large number of performance evaluation tools, concluded the characteristics of these performance evaluation tools and discovered that different performance evaluation tools suit for different application scenarios. This paper designed and realized IaaS cloud platform's performance evaluation tool, and this tool can solve the problems of cloud platform demands and capacity planning. The difference between this cloud platform's performance evaluation tool and other performance evaluation tools is that this paper's performance evaluation added a prediction model, while this model can predict the maximum number of virtual machines which could be opened by the cloud platform. This is the function which a lot of performance evaluation tools do not possess. Through this paper's performance evaluation tool, it not only can obtain the maximum numbers of the virtual machines which can be opened by the current cloud platform, but also can solve the realistic problems. If it needs to start a certain number of virtual machines and the configuration of the virtual machines is fixed, knowing the user's workload, then it can know the requiring number of the cloud platform of this kind of configuration.

Acknowledgment

This paper is supported by NSF of China (No. 6130003) . Jian Wan is partially supported by NSF of China (No. 61572163), and Zheng Ye is supported by NSF of China (No. 61300117)

References

- [1] C. Binnig, D. Kossmann and T. Kraska, "How is the weather tomorrow?: towards a benchmark for the cloud", ACM, (2009), pp. 9.
- [2] B. F. Cooper, A. Silberstein and E. Tam, "Benchmarking cloud serving systems with YCSB", ACM, (2010), pp. 143-154.
- [3] A. Li, X. Yang, S. Kandula and M. Zhang, "Cloudcmp: comparing public cloud providers", In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement, ACM, (2010), pp. 1-14.
- [4] L. Gillam, B. Li, J. O'Loughlin and A. P. S. Tomar, "Fair benchmarking for cloud computing systems", Journal of Cloud Computing: Advances, Systems and Applications, vol. 2, no. 1, (2013), pp. 6.
- [5] E. Folkerts, A. Alexandrov and K. Sachs, "Benchmarking in the cloud: What it should, can, and cannot be", Springer, (2013), pp. 173-188.
- [6] M. B. Chhetri, S. Chichin and Q. B. Vo, "Smart CloudBench -- Automated Performance Benchmarking of the Cloud", (2013), pp. 414-421.
- [7] M. Cunha, N. Mendonca and A. Sampaio, "A declarative environment for automatic performance evaluation in IaaS clouds", IEEE, (2013), pp. 285-292.
- [8] J. Kinnison and S. L. Remy, "Using Genetic Algorithms to Benchmark the Cloud", arXiv preprint arXiv:1508.06705, (2015).
- [9] N. R. Herbst, S. Kounev and A. Weber, "BUNGEE: An Elasticity Benchmark for Self-Adaptive IaaS Cloud Environments", (2015).
- [10] A. Beitch, B. Liu and T. Yung, "Rain: A workload generation toolkit for cloud computing applications", University of California, Tech. Rep. UCB/EECS-2010-14, (2010).

- [11] J. Kross and A. Wolke, "Cloudburst - Simulating Workload for IaaS Clouds", Cloud Computing (CLOUD), 2014 IEEE 7th International Conference on. IEEE, (2014), pp. 841-848.
- [12] L. Cai, Z. Jia and Y. Qi, "CloudRank-V: A Desktop Cloud Benchmark with Complex Workloads", IEEE, (2013), pp. 415-421.
- [13] S. Mukherjee, E. Osuna and F. Girosi, "Nonlinear prediction of chaotic time series using support vector machines", IEEE Workshop on Neural Networks for Signal Processing VII. (1997), pp. 511-520.