

An Optimizing Research about the Cloud Service Resource Based on the Requirements of the Users

Yanfang Kang^{*}, Dapeng Li, Xuyang Li and Mingna Guo

State Grid Henan Economic Research, Zhengzhou, 45000, China

*Corresponding author e-mail: 154626491@qq.com

Abstract. Oriented the amounts of the service requests on the user's demand, it is possible to provide the service to the users with the lowest price, the most economical resource and the type of service closest to the users' requests. In order to provide the different resources to the users effectively, the optimization model of the management about the cloud services resources is set up. In the first place, under the guidance of the resources management center, it is best to match the most suitable cloud service resources for the user needs with the grey relational comprehensive evaluation, then according to the time, the price, the data workflow and the service type attribute value. Optimal resource deployment algorithm is established with the target of the best cloud service resources offered to the users. Finally, it is verified the validity and rationality of the method by the simulation.

Keywords. Requirements of the user, Cloud computing, Service resources, Grey correlation, Optimizing model, Management.

1 Introduction

With the rapid development of electronic technology and the Internet, there is a new innovative services based on the Internet environment - cloud computing [1]. This new approach provides a convenient, on-demand business service model [2]. In order to meet user's growing requirements on huge amounts of data and the service processing, this model of the new information resource service provides a broad prospects in application. The cloud computing, with its five basic characteristics[3] the ability of rapid elasticity, the self-service on-demand, the broad network access, the shared resource pool, the measurable service, have received the unprecedented focus from all over the world. The problem of optimizing cloud resource is a core problem in cloud computing, which determines the efficiency of the tasks undertaken and the overall performance of the systems. The service users, according to their actual needs, orders the cloud services resources through the Internet from the cloud service providers. The cloud service providers, according to their users' requirements, schedule and optimize the cloud services resource distributed in different providers. So how to optimize the management service resources that satisfy the users' requirements is the main problem in this paper.

The rest of the paper is organized as follows. In the Section 3, we elaborate the cloud service model of the resources optimization management. In the Section 3, we first described the cloud service resources as quad group. We construct the associated correlation matrix about the service providers and data centers, and the associated correlation matrix about the demand and request service nodes. The users' requirements are matched to the cloud resources by the grey correlation theory in the Section 3, according to the four attributes about the time (T), the price (P), the data workflows(D) and the data service type value (Ts), we determine the optimal goal. The optimal constraints are considered in the Section 3. We introduce the optimal allocation algorithm after the mapping resources matched. In the Section 4, The effectiveness and superiority of the algorithm are analysis by the simulation experiments. We conclude in Section 5.

2 Related work

There is great difference of the information service model in the cloud computing between the traditional networks. The relationship between the users' requirements and the service resources will become more and more closely if they can access the service resources depend on the users' needs. The users' requirements gradually is various. In order to meet the users' needs, we should improve the complete efficiency of the tasks. In the paper, the users' requirements match the best cloud service resources by the method of the grey relational comprehensive evaluation, and then the optimal the service is provided to the users that according to the required target function. The cloud service resource management center as a medium, we set up the optimal management model to configure the cloud service resources.

3 Build optimal management model of the cloud service resources

3.1 Problem description

In cloud computing environment, the requirements of the users is complex and diverse. The demand of the cloud users for now mainly is the demand of information resources. The cloud users can purchase the necessary information resource according to their actual needs via the Internet. The cloud computing can offer the users various servers, personal computers, mobile phones. In order to meet the needs of the users' task better, to match the service resources conveniently, it is essential to classify the task requirements. The different fitness of the task will affect the optimal results of the allocated resources. The matching cost of the service resources would be increased if the divided fitness was too mint. The matching process of the resources would be complex if the fitness is too rough [4]. In the paper, we select the service time T_R , the service price P_R , the data workflow D_R , the attribute value of the service type (TS_R) to describe the cloud service resources. The task can be described as $S_R = \langle T_R, P_R, D_R, (TS_R) \rangle$.

The service provided is not only quantity but also diversity. In the IT industry and the academia society, the cloud services can be divided into three categories, denoted [5, 6] the infrastructure as a service, the platform as a service [7, 8] [9] and the software as a service. The information resources allocated in different parts will be scheduled again and redeployed to build the cloud information resources by utilizing the cloud computing. The cloud service will be registered in the resource management center. In order to manage and match the cloud resource, the resources in the management center will be catalogued. Above analysis, according to the diversity characteristics of the cloud services, we should build the dynamic optimal model to manage the cloud resource and solve the optimal problem of the cloud resources, with the coordination of the cloud resource management center.

The optimal model can deploy the cloud service resource dynamically based on the users' requests. The request tasks are described into quad group as $S_{Ri} = \langle T_{Ri}, P_{Ri}, D_{Ri}, TS_{Ri} \rangle i = 1, 2, 3, \dots M$, in the center of the cloud service resource management. In order to match the cloud services resources, the requests tasks of $U_1, U_2, \dots U_m$ is decomposed to different service node N_1, N_2, \dots and N_n , according to the service properties of the requests for time, the price, the data workflow and the attribute values of the service type. We build the matrix of the users - service nodes depend on the service properties. At the same time, the cloud service resources registered by the providers, $A_1, A_2, \dots A_m$ in the resource management center are described as the quad group $Re_i = \langle T_{Si}, P_{Si}, D_{Si}, TS_{Si} \rangle$. and then the resource is decomposed into the different data centers D_1, D_2, \dots, D_n , according to the service properties of the time, the price, the service type, the data workflow and the attribute values of the service type. We build the matrix the services - data center. The users' request task is assigned to the management center based on the method of the grey relational comprehensive evaluation. The cloud services are matched by the map. The services is deployed optimally according to the intelligent optimal algorithm. Finally the optimal cloud services are allocated to the users. The specific process is shown in figure 1:

Above the figure 1, we can conclude the detailed explanation: matrix of the users - service nodes and matrix services - data center. They are defined as follows:

Definition 1: Matrix of service provider –property –data center

In matrix, X_{ij} denotes that the service resources in management center will be placed in the data center, by the time T_{ij} , the price P_{ij} , the data workflow D_{ij} , the attributes value of the service types Ts_{ij} . The i, j denotes the service resources provided by the A_i are placed in the data center DC_j .

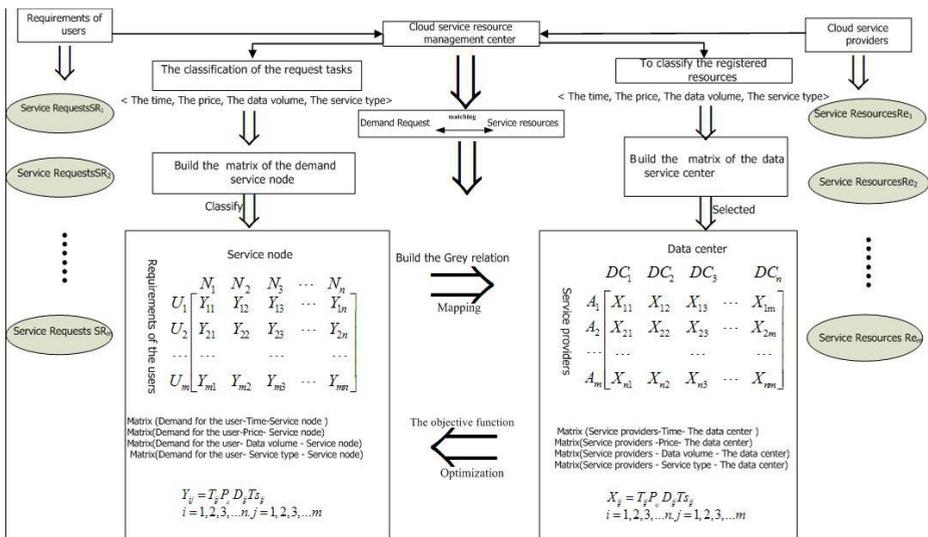


Fig. 1. The resources optimal management for the cloud service model.

Definition 2: Matrix of users –property –service node

In matrix, y_{ij} denotes that the users' requirements will be placed in the nodes by the time T_{ij} , the price P_{ij} , the data workflow D_{ij} , the attributes value of the service types Ts_{ij} . The i, j denotes the service requests of the users U_i are placed in the service node N_j .

3.2 Analysis about the optimal deployment of cloud service resources

Cloud service resources matched the cloud service requests will be deployed optimally. During the process, we should first determine the optimal goal. The objective functions represent the goal. They are established according to the time (T), the price (P), the data (D) workflow and the attributes value of the service types Ts (Ts).

The objective function 1: Execution time completed to satisfy the users' tasks service is minimum: MinT:

$$\text{Minimize obj-1} = \sum_{i=1}^n \sum_{j=1}^m T_{ij} \theta_{ij} + \sum_{i,i^*=1}^n \sum_{j,j^*=1}^m T(ij, i^*j^*) \theta_{ij} \theta_{i^*j^*} \tag{1}$$

The objective function 2: The price completed to satisfy the users' tasks service is lowest: MinP:

$$\text{Minimize obj-2} = \sum_{i=1}^n \sum_{j=1}^m P_{ij} \theta_{ij} + \sum_{i,i^*=1}^n \sum_{j,j^*=1}^m p(ij, i^*j^*) \theta_{ij} \theta_{i^*j^*} \tag{2}$$

The objective function 3: The service data workflow completed to satisfy the users' tasks service is least: MinD:

$$\text{Minimize obj-3} = \sum_{i=1}^n \sum_{j=1}^m D_{ij} \theta_{ij} + \sum_{i,i^*=1}^n \sum_{j,j^*=1}^m D(ij, i^*j^*) \theta_{ij} \theta_{i^*j^*} \tag{3}$$

The Objective function 4: The difference between the service attribute type values completed to satisfy the users' tasks service in a certain period of time is minimum: $MinSim(Ts_i, Ts_j)$

$$\text{Minimize obj-4} = \sum_{i=1}^n \sum_{j=1}^m Sim(Ts_i, Ts_j) \theta_{ij} + \sum_{i,i^*=1}^n \sum_{j,j^*=1}^m Sim[Sim(Ts_i, Ts_j) Sim(Ts_{i^*}, Ts_{j^*})] \theta_{ij} \theta_{i^*j^*} \tag{4}$$

The resources in the data center DC_j or DC_{j^*} that provided by the provider A_i or A_{i^*} is selected in the process of the optimal deployment

Otherwise,

$$\theta_{ij}, \theta_{i^*j^*} = \begin{cases} 1 \\ 0 \end{cases} \sum_{i=1}^n \sum_{j=1}^m T_{ij} \theta_{ij}$$

In the formula (1) denotes the accumulated time generated by the map matching service resources,

$\sum_{i,i^*=1}^n \sum_{j,j^*=1}^m T(ij, i^*j^*) \theta_{ij} \theta_{i^*j^*}$ denotes that the accumulated time generated is between the map matching service resources

offered by the provider A_i from the data center DC_j and the provider A_{i^*} from the data center DC_{j^*} . In the formula (2) and

(3) $\sum_{i=1}^n \sum_{j=1}^m P_{ij} \theta_{ij}$, $\sum_{i=1}^n \sum_{j=1}^m D_{ij} \theta_{ij}$ and $\sum_{i,i^*=1}^n \sum_{j,j^*=1}^m p(ij, i^*j^*) \theta_{ij} \theta_{i^*j^*}$, $\sum_{i,i^*=1}^n \sum_{j,j^*=1}^m D(ij, i^*j^*) \theta_{ij} \theta_{i^*j^*}$ have the same

meaning as the above. In the formula (4) $\sum_{i=1}^n \sum_{j=1}^m Sim(Ts_i, Ts_j) \theta_{ij}$ denotes the accumulated attribute value required for

$$\sum_{i,i^*=1}^n \sum_{j,j^*=1}^m Sim[Sim(Ts_i, Ts_j) Sim(Ts_{i^*}, Ts_{j^*})] \theta_{ij} \theta_{i^*j^*}$$

the similar service types in the map matching process. denotes that the accumulated attribute value required for the map matching service resources is generated between the service resources offered by the provider A_i from the data center DC_j and the provider A_{i^*} from the data center DC_{j^*} .

Based on the above considerations, the target functions in the optimal process are always interconnected and conflicted. Sometimes maybe a target function has reached the optimal state, the other objective function has been changed to the other optimal state. So the relative weight of each target function has to be distributed, the multi-objective optimal problem is converted into the single objective optimal problem [10], where

$$MinZ = w_1 MinT + w_2 MinP + w_3 MinD + w_4 MinSim(Ts_i, Ts_j) \tag{5}$$

In the formula (5) W_1 , W_2 , W_3 and W_4 denotes the weight coefficient that belong to the time, the price, the data workflow and the attributes value of the service types.

3.3 Optimal constraints

In the optimal process, we should consider the constraints from the following four aspects.

(1) In the whole executed task process, we should set the deadline; otherwise, the users can't wait for a long time.

$$\sum T \leq T_d \tag{6}$$

(2) In the whole executed task process, we should set the top price; otherwise, the users will give up the service, because of the price is too high.

$$\sum P \leq P_h \tag{7}$$

(3) In the whole executed task process, we should set the top line about the data flow; otherwise, the service process is overload, because of the excessive amount data.

$$\sum D \leq D_h \tag{8}$$

(4) In the whole executed task process, we should set the maximum of the service type attribute value. The sum of the service attribute value should no more than the maximum; otherwise, the unnecessary costs will be increased, because of the different gap of the service type is too large.

$$\sum Sim(Ts_i, Ts_j) \leq MaxSIM \tag{9}$$

3.4 Optimal algorithm of cloud service resources in deployment

In the cloud service resource management center, according to the correlation matrix established (the demand users - service node), we deploy the cloud service resources purposeful in the data center. First we encode the service nodes of the users' requirements. This paper adopts the indirect coding method. Each service node N_i denotes different users' requirement attributes. $N_i = \langle T_{ij}, P_{xy}, D_{mn}, TS_{kl} \rangle$, where T_{ij} is the time, P_{xy} is the price, D_{mn} is the data workflow, TS_{kl} is the service attribute type value. Number of chromosome is the requirements. We should code the service resource that is taken by the tasks; the specific graph is as follows in figure 2:

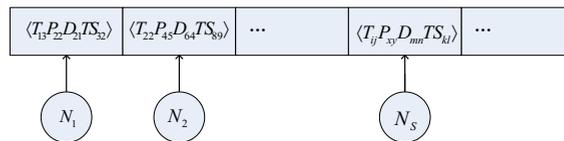


Fig. 2. Chromosome coding correspond the task node – the service provider- the data center.

In the figure 2, where T_{ij} is the execution time, P_{xy} is the pay price, D_{mn} is the service data workflow, TS_{kl} is the service attribute type value. They are all designed to complete the tasks in the service nodes N_i . In order to satisfy the requirements of the users, we set the parameters of the corresponding service nodes $\langle T_{ij}, P_{xy}, D_{mn}, TS_{kl} \rangle$, where the execution time T_{ij} is stored in data center DC_j provided by the service provider A_i , where the pay price P_{xy} is stored in data center DC_y provided by the service provider A_x , where the service data workflow D_{mn} is stored in data center DC_n provided by the service provider A_m , where the service attribute type value TS_{kl} is stored in data center DC_l provided by the service provider A_k . The main task of this paper is to deploy the cloud service resources optimally.

The optimal deployment problem of the cloud resources is to set the parameter problems $\langle T_{ij}, P_{xy}, D_{mn}, TS_{kl} \rangle$ in the service node. The parameters of the service nodes are processed by the cloud service resource management center. After the service registered by the cloud service providers in the cloud service resource management center, in order to schedule the resource conveniently, the cloud resource will be allocated in the different data centers with the form of the execution time, the paid price, the data workflow, the service attribute type value. This problem is converted to the optimal deployment about the cloud resource stored in the different data centers, registered by the different cloud service providers. In order to optimize the service resource allocation, it is necessary to use optimization algorithm to optimize the task.

4 Simulation

In this section, according to the simulation example, the effective of the optimal management model of the cloud services resources is analyzed. Hypothesis that at any time, the five users' requirements are received in the cloud service resource management center. They are the service requirement SR_1 , the service requirement SR_2 , the service requirement SR_3 , the service requirement SR_4 , the service requirement SR_5 . First the five users' requirements will be set in the different service nodes, according to their different properties of the different users. Then according to the grey relation map, the highest correlation between the service requirements and the cloud service resources is to be found.

We hypothesis that there are three cloud service providers who offered the resources to meet the requirements of users. According to the above mentioned, the five kinds resources is selected to form the initial population respectively, according to the sub-objective functions. Hypothesis that the size of the initial population is 40, the length of the

chromosome is 10. The optimal matched service resources offered to the users is selected by the optimal deployment algorithm, The information can be get in the table1 and table2.

When the resource is to be optimized, the coefficient of the weights are respectively that $w_1 = 0.42$, $w_2 = 0.21$, $w_3 = 0.19$, $w_4 = 0.18$, $TD=45S$, $Ph=120 \text{ ¥/h}$, $Dh=230\text{Mb/h}$, $MaxSim=0.55$. We adopt the general genetic algorithm and the improved genetic algorithm as the above 2.5. Each algorithm is carried out experiments for 60 times. The results is $MinZ = 21.86$. The corresponding optimal resource group is seen as the table 3.

From the Figure 3, the global optimal solution can be obtained by the general genetic algorithm or the improved genetic algorithm in the paper. But in the searching process, the optimal solution fluctuates most by the general genetic algorithm (GA), the stability is poorer. However the optimal solution fluctuates slight by the improved genetic algorithm within 35 times. The wave number is 8. So the stability of the improved genetic algorithm is better than the general genetic algorithm. The improved GA during the global optimal process is more stability. Figure 4. shows the number of iterations, that the solution has reached the optimal results by the two kinds of genetic algorithm. The figure 4 illustrates that the average iterations number of the improved genetic algorithm is 51, the average iterations number is 157 by the general genetic algorithm. The average iterations number of the improved genetic algorithm is lower than that of the general genetic algorithm. Wave amplitude is less than that of the genetic algorithm. General genetic algorithm will fall into local optimal solution. Because their fitness functions set have the consistent standards. Search speed of the optimal solution is slower. However, the improved genetic algorithm has the unique method to set the fitness function, which can reach the optimal solution at a rapid speed.

5 Conclusion

The optimal deployment problem of the cloud service resources is the core of the cloud computing. At the same time the deployment problem is the basis that realizes the flexibility, the scalability, the resource sharing. It is the key that we choose the best cloud services provider to the users. It is related the basic interests about the cloud services providers and the cloud users. This paper constructs the optimal allocated management model about the cloud service resources. The model distributes the cloud service resources to the different service nodes by the cloud service resource management center, based on the diversity of the cloud users' requirements. The model takes advantage of the grey relational comprehensive evaluation to match the cloud service resources. It uses the improved genetic algorithm to optimize the matching cloud services, with the optimal goals of the time (T), the price (P), the data workflow (D) and the service type value (Ts). The optimal resource selected is offered to the users. Cloud service resource optimal deployment model and the solution algorithm proposed in the paper can provide the theoretical support and the research method for the optimized deployment management of the cloud computing resources.

Table 1. Distribution list of the requirments task nodes.

The task for Requirements	SR ₁	SR ₂	SR ₃	SR ₄	SR ₅
	N ₁₁	N ₂₁	N ₃₁	N ₄₁	N ₅₁
	N ₁₂	N ₂₂	N ₃₂	---	N ₅₂
The service nodes of the assigned tasks	N ₁₃	---	N ₃₃	N ₄₄	---
	N ₁₄	N ₂₄	---	---	N ₅₄
	---	N ₂₅	N ₃₅	N ₄₅	N ₅₅

Table 2. Service resource description matched for the user's requirements.

	Data Center	Time (S)	Link time(S)	price(¥/h)	Link price(¥/h)	Data traffic(Mb/h)	Link flow(Mb/h)	Type attribute value	Link attribute values
	DC ₁	18	3	16	8	41	23	0.78	0.12
	DC ₂	6	7	24	5	71	18	0.83	0.23
Service providers	DC ₃	16	2	12	7	93	10	0.52	0.30
A ₁	DC ₄	20	5	19	4	46	15	0.90	0.25
	DC ₅	7	4	31	10	52	19	0.64	0.28
	DC ₁	13	1	30	12	80	12	0.85	0.31
	DC ₂	8	6	18	9	66	17	0.82	0.15
Service providers	DC ₃	10	4	24	6	58	16	0.7	0.24
A ₂	DC ₄	15	3	15	20	35	25	0.73	0.19
	DC ₅	17	7	18	15	48	27	0.59	0.27
	DC ₁	11	8	34	17	78	11	0.75	0.18

	DC ₂	13	6	20	11	67	18	0.68	0.16
Service providers	DC ₃	14	3	22	13	50	14	0.55	0.34
A ₃	DC ₄	9	5	19	10	63	20	0.72	0.21
	DC ₅	5	4	12	9	82	19	0.67	0.26

Table 3. Combination table of the optimal resource.

N ₁₁ : <(3,5)(1,5)(1,1)(2,3)>	N ₃₃ : <(1,3)(3,2)(2,3)(3,4)>
N ₁₂ : <(2,4)(3,5)(3,1)(3,2)>	N ₃₅ : <(2,4)(3,1)(3,4)(3,2)>
N ₁₃ : <(3,2)(3,4)(3,4)(3,5)>	N ₄₁ : <(2,1)(1,4)(3,5)(2,2)>
N ₁₄ : <(2,4)(3,2)(3,5)(1,5)>	N ₄₄ : <(2,2)(2,4)(3,5)(3,3)>
N ₂₁ : <(3,5)(3,2)(3,4)(1,5)>	N ₄₅ : <(1,4)(2,5)(3,4)(1,1)>
N ₂₂ : <(3,3)(2,5)(2,4)(3,4)>	N ₅₁ : <(3,2)(3,5)(3,4)(1,1)>
N ₂₄ : <(2,2)(3,2)(3,1)(1,2)>	N ₅₂ : <(1,5)(2,4)(2,3)(3,4)>
N ₂₅ : <(3,2)(1,5)(3,1)(2,2)>	N ₅₄ : <(2,2)(3,4)(1,3)(3,4)>
N ₃₁ : <(2,5)(1,4)(3,2)(3,1)>	N ₅₅ : <(2,4)(3,3)(2,1)(2,4)>
N ₃₂ : <(2,4)(3,2)(2,1)(3,1)>	

(note: N_i: <(a₁, b₁) (a₂, b₂) (a₃, b₃) (a₄, b₄) >. In the table, the (a_i, b_i), respectively, denote the task of the service nodes N_i is performed with the time, the price, the data traffic of the resource, the property values of the service types. The task node require to schedule the resource b_i in the data center provided by the service provider a_i.)

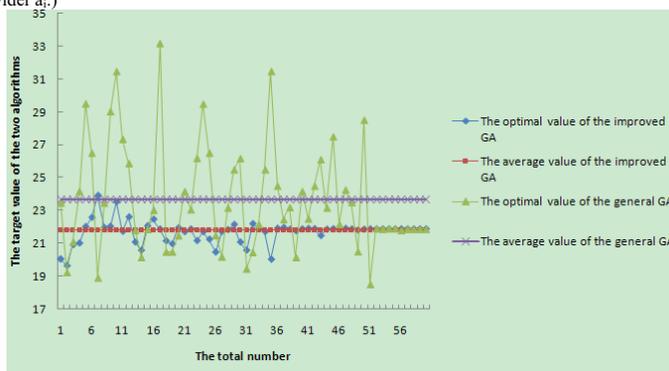


Fig. 3. Optimal target value of the two algorithm.

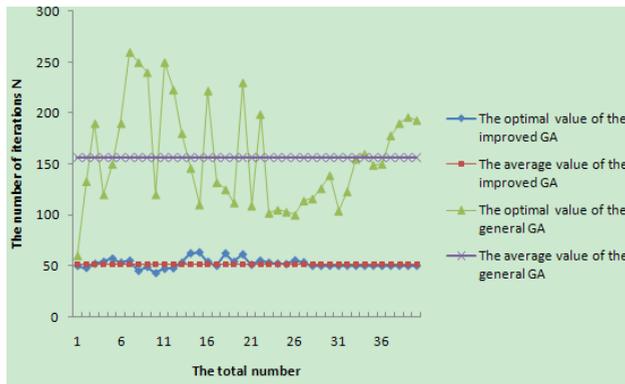


Fig. 4. Number of iterations on the optimal value of the two algorithms.

Acknowledgment

This work was financially supported by project that Research evaluation system on social benefits of major power transmission engineering

References

1. Wang J.J., Lv Z.H., Jie W., et al., Analysis the development and application about the cloud computing technology, Journal of computer applications and design, vol.31,no.20,pp.4404-4409, (2014).
2. Han. W. The influence factors of the cloud computing services market competitive analysis, Journal of the managerial modernization, no. 2, pp. 25-27, (2015).
3. Lin W.W., Qi D.Y. The research review on the cloud computing resource scheduling, Journal of the computer science, vol. 39 ,no.10 ,pp.1-6, (2014).
4. Li B.M., Liu Q. Try on the match of the grid resources based on ontology, Journal of northwestern polytechnic university.vol.28,no.6, pp.608-611, (2015).
5. Ju J.H., Wang Y., Fu J.Q. Research on Key Technology in Saas, Intelligent Computing and Cognitive Information(ICICCI),International Conference on. pp.384-387, (2016).
6. Banerjee, Srikanth, Cukic. Log-Based Reliability Analysis of Software as Service(Saas), Software Reliability Engineering(ISSRE),IEEE 21st International Symposium on.pp.239-248, (2015).
7. Bernstein, Vidovic, Modi. A Cloud Paas for High Scale, Function, and Velocity Mobile Applications-With Reference Application as the Fully Connected Car, Systems and Networks Communications(ICSNC),Fifth International Conference on. pp.117-123, (2014).
8. Lv C.T., Li Q., Lei Z. Paas: A revolution for information technology platforms, Educational and Network Technology (ICENT),International Conference on.pp.346-349, (2014).
9. B.S. Lee, S.X.Yan, D.Ma etal. Aggregating IaaS Service.SRII Global Conference (SRII Global Conference).Annual. pp.335-338, (2015).
10. L.L. Hall. Evolution of telco-services utilizing Infrastructure as a Service(IaaS), Intelligencein Next Generation Networks(ICIN),15th International Conference on. pp.247-252, (2016).