IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

VM Selection using Index Approach for Deploying Cloud Computing Application and Approach to Obtain Equal Utilization of Virtual Machine

Garima Dubey[1], Yogendra Kumar Jain[2]

Research Scholar, Computer Science and Engineering, Samrat Ashok Technological Institute, Vidisha (M.P.), India[1]

Head of Department, Computer Science and Engineering, Samrat Ashok Technological Institute, Vidisha (M.P.), India[2]

Abstract: Cloud computing is an upcoming field having lots of scope and research areas. In this age of technology everything from our home to office is done in the cloud and it is the future of technology. There are so many research has been done in this field to improve the quality of service. In this paper, we propose a model for deploying cloud computing applications on indexed cluster. The proposed model is a hybrid cluster approach which controls the cloud operations. The cloud manager will allocate a VM within the CLUSTER in such a way that it will save the power to a greater extent while maintain the latency and also maintains the efficient utilization of the CLOUD resources.

Keywords: Cloud Computing, Clustering, VM Allocation, Indexing, and Topology aware.

I. INTRODUCTION

Cloud computing is the latest effort in delivering computing resources as a service. User can hire computing resources online as a product these computing resources will be used by user online [1], this helps in reducing costs and providers will take charge of it and provide on demand service that is delivered to consumers over the internet from large-scale data centres – or "clouds". Whilst cloud computing is gaining growing popularity in the IT industry.

Cloud Computing consist of collection of distributed servers known as masters which provides demanded services and resources to the clients known as central controller or cloud manager in a network with the scalability and reliability of servers. On-demand service will be provided by distributed servers. Services may be of software resources (e.g. Software as a Service, SaaS) or physical resources (e.g. Platform as a Service, PaaS) or hardware/infrastructure (e.g. Hardware as a Service, HaaS or Infrastructure as a Service, IaaS). Amazon EC2 (Amazon Elastic Compute Cloud) is an example of cloud computing services.

Cloud services are provisioned to use by service providers, for example, Amazon, Google on the Internet. Usually, the resources available to the user of the cloud are virtualized that is (Paas, Iaas, and Saas) services are virtual service. User gets required service without any dependencies or constraints in return companies will take some charge for using their services which are nominal as compared to the actual cost of that particular service, because of this cloud is becoming popular.

Cloud computing technology uses internet and central remote servers to maintain data and applications [2]. Cloud computing allows any user from anywhere to use the updated version of services and application. We need not to purchase software with the license because updating and maintaining software are the server's responsibility, only we need to have internet connection, with that we can use applications without installing it on our system. Gmail, yahoo mail or Hotmail, etc. are the common and widely used cloud example, for using any kind of cloud service, you must have an internet connection.

Table 1: Abbreviations used

VM	Virtual Machine
PDF	Program Descriptor File
Paas	Platform as a Service
Iaas	Infrastructure as a service
Saas	Software as a Service

Scientific application needs a large amount of calculation and storage for which you needs a large computation storage and power. Initially, all the scientific applications are deployed on Grid [3]. However, Grid computing is costly and not available all over the word. Therefore the scientific applications are moving toward the cloud, Cloud provides an alternative to grid and supercomputers for a scientist in a lower cost. Cloud is an emerging area and perfect for this kind of application. For the improvement of deploying an application of cloud there are so many strategies developed, for example load balancing, scheduling algorithm for VM allocation in cloud.

Virtualization: Virtualization is a layer between Hardware and operating system. Initially mainframe is used to support many users using the virtual machine terminal. This terminal shows the simulated behaviour of an operating system for each user. VMware Workstation is a similar product started in 1999 and it facilitates to run

multiple operating systems in personal computers. Virtualization is the foundation of cloud technology. Users can access resources for computing or for storage using virtualization without knowing background detail [2].

Virtual Machine Allocation: VM allocation allows efficient sharing of virtual machines to available data centers and these allocation methods help to evaluate and enhance the performance of cloud [4]. Different allocation policies are available and they have their own advantages and limitations. The major objective of every VM allocation method is to minimize time. Throttled Load Balancer (TLB) is also the similar research done before.

In addition to VM allocation, proposed work is also handling the following issues:

Delay Minimize: In the proposed approach, all the nodes of the cluster get booted initially and it shuts down when not used over time equal to the thresholds value, however, all the systems get booted and activated for allocation initially and therefore the time taken in booting is minimized during the first process to allocate and execute.

Wear and Tear maintained: we analyze and store the working hour of each virtual machine. According to that we select least used VM so that wear and tear of each machine could be managed.

Power save: Suppose any VM is free for a time equal to or greater than the threshold value, the cloud manager turn off that VM and controller will look if all the VMs at same CLUSTER are also in turn off state. Then it will turn off the whole CLUSTER.

Load balancing: In the proposed approach, least used VM will be selected every time when the request arrive this approach distribute equal load to every Virtual Machine.

II. LITERATURE REVIEW

Cloud computing is the fastest growing area in computing research, industry and business today. Many studies have been done in the cloud computing to increase the QOS. Load balancing and Scheduling algorithm are mainly two research areas growing for improving the performance of the cloud computing algorithm. In the existing similar approaches for Cloud Computing which is covered in this section. Daniel [5] in 2009 has explained the concept of cloud computing, its advantages and disadvantages and also describes several existing cloud computing platforms. Researcher discussed the results of quantitative experiments carried out using Planet Lab, a cloud computing platform.

Calheiros et al. [6] has proposed VM allocation according to QOS performance comparison and also detected the change in intensity of the workload and allocated multiple resources. The proposed model automatically adopts to workload changes model behaviour and performance of applications and cloud based IT resources to adaptively serve end user.

Michael [7] in 2011 has presented fast and accurate k-means clustering for large dataset. Cloud computing deals with large data centres, so huge amount of data is accessed by many users. In this way, fast and accurate clustering algorithm seems to be a better option to adopt in clouds. The results of presented work have shown that K-Means clustering is much faster than other Divide and Conquer Algorithms.

In 2012 Ahmed et al. [10] has explained the research areas of cloud computing, mainly researcher discussed and simulate the elasticity problem in cloud, for handling elasticity problem he simulated a feedback elastic controller. He designed it for VM allocation according to the load of request when load is increased the VM could be added in it and when load is decreased VM can be removed. Author applied the load balancing concept in it as well. The result of research shows better resource oscillation.

Soumya [8] explained the basics of cloud computing component and types of cloud computing, load balancing in cloud environment, benefits and barriers. Load balancing is a process of distributing workload to the multiple computers over the network or many other resources as well to achieve maximum resource utilization, maximize throughput, minimum response time, and avoid overload. With the new era of technology, resource control or load balancing in cloud computing is a main challenging issue the most important part of her research is he explained the concept of cloud simulators "cloudsim".

Nidhi [13] has presented a comparative study of existing load balancing technique according to some specified parameter reducing overhead, service response time and improving performance and focused on the energy efficient techniques for the green computing.

Abhishek [14] have designed a scheduler for Optimizing VM Placement, for High computing cloud (HPC) and implement it on the top of the Open Stack scheduler. Author considered the topology awareness and homogeneity while allocating VMs. Author also demonstrated the benefits of these techniques by evaluating them on a cloud setup on Open Cirrus test-bed. He explained the opportunities and limitation of high computing cloud. Scheduling algorithm increases the performance by 45%.

In 2013 author Kliazovich and friends [15] has proposed energy efficient scheduling for cloud computing application with load balancing technique. Author designed a scheduler which works in two simple steps first author selected group of server connected to data centres with highest available bandwidth. Second among those selected server select computing server with smallest computing capacity to do scheduling task. The energy consumption of each server at minimum and maximum load is calculated and according to that, load is distributed to the servers.

The Fan [3] in 2014 has explained the method of topology aware deployment of scientific application in a cloud computing it takes the information from the cloud manager about the communication between the process and compares the physical topology (which already available to the cloud manager) and logical topology and deploy the application accordingly for example if we have

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

5 process to complete a job and 4 VM in a cluster then we select four process which are communicating to each other frequently, it reduces the communication overhead and.
Madhukara [15] has implemented a scheduling algorithm in scheduling environment for the cloud. He stored the average execution time of each virtual machine according to the previous patterns and then sort it with lowest execution time in the first place with this array when a request arrives it checks if it already have jobs assigned greater than the threshold value then it will go to the next place. The proposed algorithm is a different approach which deals with the indexing in cloud, to achieve higher speed in the cloud. This is the first attempt to work in this field for increasing the searching ability.

III. PROPOSED METHODOLOGY

The proposed method is an extension of the topology awareness deployment of cloud application in the cloud. In this paper we proposed the VM selection method, any application to run in cloud it should have virtual resources which it can take from the free available resource pool. For allocating virtual machine as a resource to execute job it will need to search the available VM in the resource pool, and by selecting VM it allocate job to that VM. Existing technology has worked on the topology awareness that is, the prior knowledge of topology which can improve in allocation of virtual machines in a clustering environment. This is done by calculating the physical and logical topologies, physical topology is the network topology and the logical topology is the topology calculated according to the communication done between the processes it can directly be obtained from the program file descriptor submitted by client in case client had not submitted it can easily be calculated by cloud manager. By comparing both the values cloud manager will allocate the process. This technique overcomes the extra communication overhead due to different cluster allocated to the process communicating frequently. After applying the above approach cloud manager will search the index table, it also searches for the least used VM and its status and allocate the VM to the request.

1. PROPOSED SYSTEM MODEL

Framework for the proposed method is shown in the figure 1.In this framework users are sending jobs with their PDF (program descriptor file) to the cloud manager. Cloud manager will match and compare the logical topology to the physical topology [3] and select the cluster which reduces the communication overhead for example if we have five processes and four VM available in a cluster then the cloud manager will allocate four processes which are communicating frequently to one cluster and allocate one that is less communicating to another cluster. In the second stage cloud manager will search existing active VM from the stored indexes and select the VM which have least working time and deploy the application.
Initially we have stored the indexes of all the clusters and their VM's in a table along with the status that a system is in ON condition or is in OFF condition.
CLUSTER Activity Status Monitoring Table 1 shows the simple example, the indexes of each cluster and VMs within each cluster. By using the index table a cloud manager could allocate VM just in a single click. The table is stored at the cloud manager and every time when a new VM is activated or initiated it first added in to the status monitoring table.

Table 2: CLUSTER Activity Status Monitoring Table

No of Cluster	VM 1	VM 2	VM 3	VM 4	BUSY
CLUSTER 1	0	0	0	0	0
CLUSTER 2	0	0	0	0	0
CLUSTER 3	0	0	0	0	0
CLUSTER 4	0	0	1	0	1

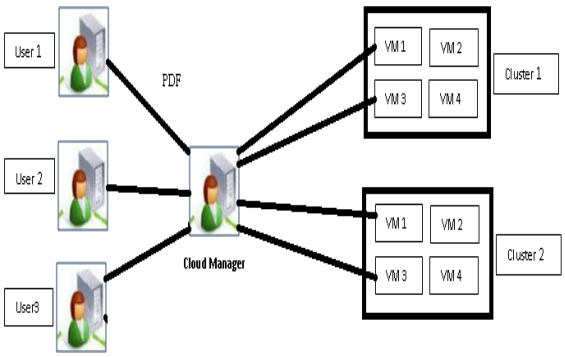

Figure 1: Proposed Framework system model.

Cloud manager first checks in the cluster if the cluster is in active mode and available, then it checks for the Virtual Machine under that cluster, if a Virtual Machine is in active mode and is available and then it select that and allocate it to the client.

2. PROPOSED ALGORITHM

When cloud received a request, it first needs to select the appropriate cluster for the deployment of cloud application, by selecting a Cluster, which consists of two or more VM and other resources; it will select the freely available VM, and will allocate the job to the selected VM. The proposed Algorithm uses an index approach for VM selection. In this approach the indexing is done to all the resources with their status as it is available or not in this way it can be easily selected in a fast manner. It is explained in the following steps.

STEPS 1 - The cloud manager initially forms the Indexes of each CLUSTER, Indexes of each VM within each CLUSTER, and Indexes of each resource available within each VM
STEP 2 - The cloud manager collects the Indexes of current VMs in use and indexes of CLUSTER in which currently active VM is present
STEP 3 – Cloud manager generate a reference table to store the status of different VMs and CLUSTERs for example, let the selected CLUSTER is number 4, selected VM within CLUSTER is 3 and all other CLUSTER and VM remains idle, then it will be represented as shown in Table 1.
STEP 4 - One extra column is added in reference table which shows the status of that CLUSTER whether it is

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

started or not in case of CLUSTER.
'0' represents CLUSTER is off.
'1' represents CLUSTER is started.
In case of VMs,
'0' represents that VM is free.
'1' represents that VM already has data.
Whenever any CLUSTER is selected for the first time, its entry is made in the reference table along with its CLUSTER and all the VM within it, and the table is immediately updated after any action is performed on VM or CLUSTER level either assigning workload (wakeup) or turning off (power down) it.
Initially the entire CLUSTER is in off state and all the VM within them remain started.
This is done in order to assure that whenever a new cluster is selected that was in off state, only one wakeup signal will be sufficient to wakeup whole CLUSTER for reducing further latency to wake up each VM.

2.1 VM Allocation
STEPS 1 - Cloud manager will search for the free VM in the currently active VM within the currently pointed active CLUSTER. If free VM is Available then it will select that VM.
ELSE
STEP 2 - Cloud manager will search the table for the already active CLUSTER with free VMs from the scratch. If the search is successfully done then, it will select that VM.
ELSE
STEP 3 - Scheduler will select next lowest indexed CLUSTER to start it and the lowest indexed VM will be selected finally its entry will be made in the table immediately.
Suppose a CLUSTER's entry is present in the table showing that it was previously used but presently it is in power off state ,then that CLUSTER will not be selected (unless and until that is the last free available CLUSTER) , for efficient utilization of all memory blocks.

2.2 Turn off CLUSTER and VM
Suppose any VM is free for time equal to greater than the threshold value, then the cloud manager will turn off, that VM, however, controller will look if all the VMs at same CLUSTER are also in turn off state therefore, it will turn off the whole CLUSTER

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experiment is performed in simulation environment; tool used for the implementation is MATLAB. The algorithm is implemented with the help of a function which takes input as number of virtual machine available for deployment of cloud application, capacity of the each system, total time and average workload of the cloud system. The main comparison parameter is average workload. The average workload is shown for both the existing and proposed system in Table 2. Experiment is performed on the following values.
No. of virtual machine = 16
Total capacity of system = 1
Average workload = 10
Total time = 100
Cloud manager will search active Cluster and Active VM within cluster from the index he already have and then select the VM and allocate the resource. If cloud manager fails in searching active resource it will start/activate a cluster and all the VM within it 'having the lowest' index value to assign jobs. After a threshold value that is equal to the half of "Average of cores idle time" check if system is idle for the time equal to threshold value then turn it off.

$$\text{Threashhold value} = \frac{\sum \text{cores Idle time}}{2 * (\text{No. Of Cluster})}$$

The configuration used for performing the experiment is windows 7, MATLAB R2012b, and virtual clone drive.

4.1 RESULT COMPARISION
Indexing in searching is done where we need minimum time complexity. With the help of Index table which is stored at the cloud manager we can access and allocate virtual machine in just a second. The result comparison is done between the four cloud deployment techniques these are listed below. Random selection, Sequential selection, Clustered selection and proposed. Random selection and sequential selection have same average active time. Most of the application in cloud is by default allocated randomly which gives worst case complexity. In the proposed technique we can have least Average Active time, Power is costly so we can reduce the cost by applying approach. Average boot up latency and the control required is also better as compared to Random and Sequential selection. Comparative study has been done which is described in the Table 3.
Average Boot up latency: When system is booted foe initializing and deploying a cloud application every time or one can say that when a system get activated from OFF to ON state it is counted in a variable and calculated the Average boot up latency. The value should be less for achieving higher rate of execution.
Average active time: A system is in ON state that is it is in active state therefore it is using the power which is the matter to deal with; every approach has calculated the average value of active time of clusters which should be less.
Control: The variable Control required is the monitoring needed to execute a job in that clustered approach the value should be less.

Table 3: Comparison between different cloud deployment techniques

	Seq.VM selection	Random VM selection	Cluster	Cluster Propose
Average Boot up latency	0.43	0.9	0.02	0.2375
Average active time	10.75	10.75	3.04	2.96
Control Required	16	16	4	8

DOI 10.17148/IJARCCE.2015.41184

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

4.2 LOAD DISTRIBUTION

Load distribution among the available cluster and used cluster is an important concern to deal with; in the 3-d load distribution graph, which plotted, x coordinates as VM number, Y coordinates as average load and the z coordinate as VM load distribution, in figure 3 we can show the load distribution in our approach uses the entire four clusters available in the cloud equally. However, in the existing clustered approach in figure 2, every time it uses only first cluster that is CLUSTER1 in that it uses first VM 1 for every request therefore CLUSTER1and VM 1 is heavily loaded all the time. Remaining clusters are in active state but not used for deployment whereas VM 2, VM 3 and VM 4 are used. When a request arrives from the user it will be allocated to the first available cluster so all other cluster remains idle for a long time this is the main drawback of clustered approach it can be removed by adding a load balancing technique. In our approach there is no need to add load balancing technique because in our approach all the clusters are equally used as shown in figure 3.

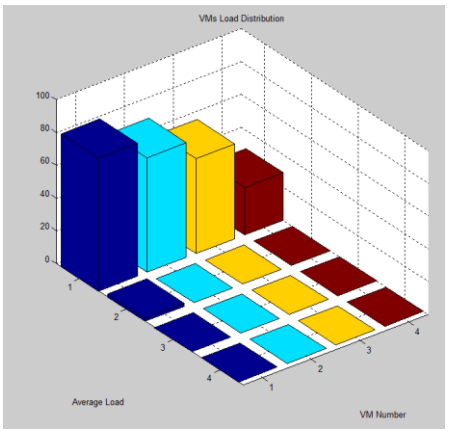

Figure 2: Load Distribution graph in clustered approach

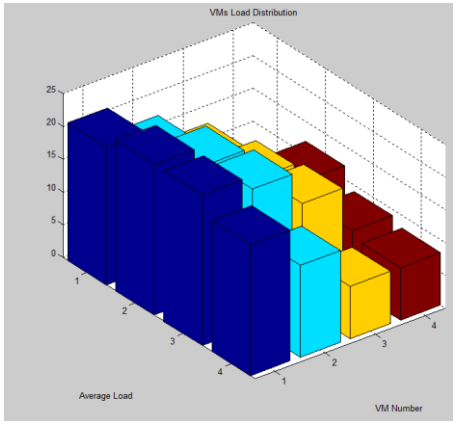

Figure 3: Load distribution graph in proposed approach

4.3 VM ALLOCATION

Among the four techniques available, we have plotted the VM number and average load in the graph figure shown below, it is showing the used VM and load among each VM.

Case 1: Sequential selection in figure 4 shows the load distribution graph shows that the load is distributed among initials 1 to 8 virtual machines and VM 8 to VM 16 remain idle and not used. Average active time and control required is similar to random selection approach.

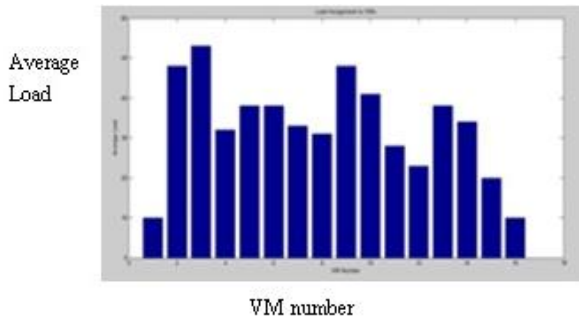

Figure 4: Random VM selection

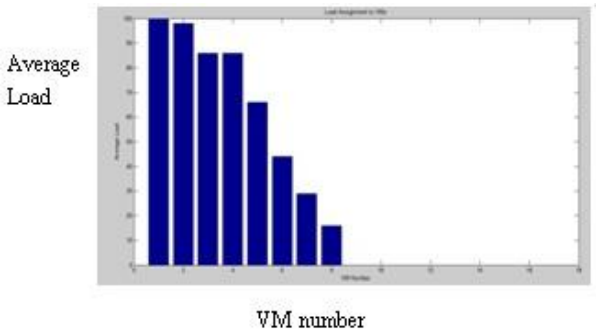

Figure 5: Sequential VM selection

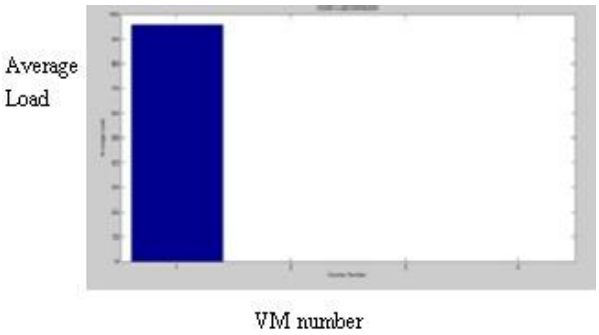

Figure 6: Cluster VM selection

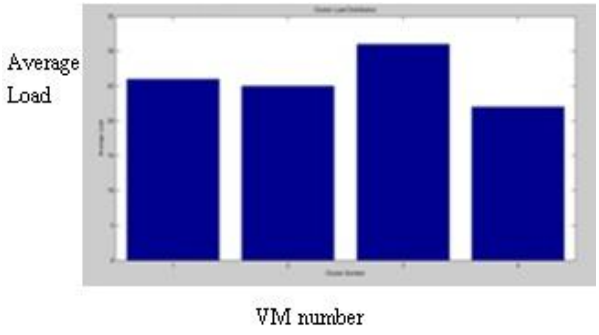

Figure 7: proposed cluster VM selection Average load

Case 2: Random selection in figure 5 shows the use of every virtual machine because VM allocation is done randomly any VM can be selected there is no criteria. This technique gives us bad utilization factor. The average active time is similar to the sequential selection. Random approach shows better result when comparing with

DOI 10.17148/IJARCCE.2015.41184

sequential selection in terms of load distribution and average active time.

Case 3: Clustered approach in figure 6 shows all the load is on the first cluster every time when request arrives, it is allocated to the first active cluster and so all other clusters in cloud remain unused and idle using power but not being used. In case of Average active time it shows comparatively better result here control required is also smaller than Random as well as sequential approach.

Case 4: The graph figure 6 shows result of the proposed approach in which x coordinates show VM number and y coordinate shows average load. All the clusters available are used equally; load is distributed equally in all the four clusters. This approach will give minimum Average active time, compared in table 3. When we increase the average load of the system, then the graph shows better results. Time of execution will be minimized. Control used in this approach is 8 which are much smaller than random approach. Overall performance compared to other approaches proposed cluster gives us the best result.

V. CONCLUSION

The Indexed clustering approach is implemented in simulated environment. Topology awareness deployment method is added with the indexing and compared it without indexing as Indexed approach shows better result in terms of load distribution and power saving because average active time of VM's and Cluster's are smaller in proposed approach than the existing techniques. In scientific application where time constraints are the most important value we can use this approach. Different research work has been done in this field but the proposed methodology gives fastest among existing technique this is only because of considering the two technology one is topology awareness technology and other is indexing as the name suggest indexing gives us the selection of virtual machine in just one click. For each application to deploy on cloud, cloud manager has to search the index table which could be costly for a large system such as cloud. One more problem of storage cloud manager needs enough space to store the index table of entire cloud system. In future the speed of this methodology could be increased by replacing the indexing technique with other searching method such as Hashing which is the fastest searching method with complexity of omega of one.

REFERENCES

[1] http://www.wikinvest.com/concept/Cloud_Computing
[2] Mladen A. Vouk, Cloud Computing – Issues, Research and Implementations, Information technology Interfaces 30th International C 1onference, Vol. 4, pp. 235–246, June 2008.
[3] Pei Fan and Zhenbang Chen and JiWang, Topology-Aware Deployment of Scientific Applications in cloud computing, Int. J. Web and Grid Services, Vol. 36, PP. 319 - 326, July 2014.
[4] Bhupendra Panchal and Prof. R. K. Kapoor, "Dynamic VM Allocation Algorithm using Clustering in Cloud Computing", International Journal of advanced research in computer science and software engineering, Volume 3, Issue 9, PP. 143-150, September 2011.
[5] Daniel A. Menasc´e Paul Ngo, "Understanding Cloud Computing: Experimentation and Capacity Planning", Proc. Computer Measurement Group Conf, Dallas, TX, Dec. 7-11, 2009.
[6] Calheiros R.N., "Virtual Machine Provisioning Based on Analytical Performance and QoS in Cloud Computing Environments", International Conference on Parallel Processing, page no 295-304,Sept 2011.
[7] Michael Shindler, Alex Wong, "Fast and Accurate k-means For Large Datasets", PP. 1-9, 2011.
[8] Soumya Ray and Ajanta De Sarkar, "Execution analysis of load balancing algorithm in cloud computing environment", International Journal on Cloud Computing: Services and Architecture (IJCCSA), Vol.2, No.5, PP. 1-13, October 2012.
[9] http://explainingcomputers.com/cloud.html
[10] Ahmed ali eldin, Maria Kihl and Johan Tordsson, "Efficient provisioning of bursty scientific workload on the Cloud Using Adaptive Elasticity Contro"l, journal of Computer Science Issues, Vol. 9, Issue 1, No 1, PP. 31-40, January 2012.
[11] http://www.whatiscloud.com
[12] Nidhi Jain Kansal, Inderveer Chana, "Cloud Load Balancing Techniques : A Step Towards Green Computing, International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, PP. 1694-0814, January 2012.
[13] Rajleen Kaur, Amanpreet Kaur, "A Review Paper on Evolution of Cloud Computing, its Approaches and Comparison with Grid Computing", International Journal of Computer Science and Information Technologies, Vol. 5, PP. 6060-6063, 2014.
[14] Abhishek Gupta, Osman Sarood, Laxmikant V Kale, "Optimizing VM Placement for HPC in the Cloud", International Letters of Social and Humanistic Sciences,Vol. 16, PP. 1-6, 2014.
[15] Dzmitry Kliazovich, Sisay T. Arzo, Fabrizio Granelli, "e-STAB: Energy-Efficient Scheduling for Cloud Computing Applications with Traffic Load Balancing", IEEE International Conference on Green Computing and Communications, pp 7-13, 2013.
[16] Ellendula Madhukara, Thirumalaisamy Ragunathan, "Efficient Scheduling Algorithm for Cloud", 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15), 50, PP. 353 – 356, 2015.