IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

A Novel Approach of Task Allocation based on K-means and QoS in Cloud Environment

Sahil Kumar[1], Pawan Luthra[2], Bohar Singh[3]

Research Scholar, Computer Science Engg, Shaheed Bhagat Singh State Technical Campus, Ferozepur [1]

Assistant Professor, Computer Science Engg, Shaheed Bhagat Singh State Technical Campus, Ferozepur [2,3]

Abstract: Cloud computing is a paradigm of distributed computing to provide the customers on-demand, utility based computing services. Cloud users aim at providing more reliable, available and updated services to their clients in turn. Since cloud computing stores the data and widely spread resources in the open environment. So the amount of data storage increases quickly. Cloud itself consists of physical nodes in the data centers of cloud providers and there are number of applications that run on cloud which a user uses on a basis of pay per use. Therefore, the system that incurs a cost for the user should function properly .So as to make the system function properly it should have the proper algorithms implemented. In this paper, a load balancing algorithm has been put forward and implemented in CloudSim toolkit andalso simulation results have been computed andanalysed.

Keywords: Cloudlets; Data Center Broker; Clusters.

I. INTRODUCTION

Cloud computing can be defined as a model for enabling ubiquitous, convenient and on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort from the user side and minimal service provider interaction. Cloud computing is considered the evolution of a variety of technologies that have come together to change an organizations' approach for building their IT infrastructure. Actually, there is nothing new in any of the technologies that are used in the cloud computing where most of these technologies have been known for ages. It is all about making them all accessible to the masses under the name of cloud computing. Cloud is not simply the latest term for the Internet, though the Internet is a necessary foundation for the cloud, the cloud is something more than the Internet. The cloud is where you go to use technology when you need it, for as long as you need it. You do not install anything on your desktop, and you do not pay for thetechnology when you are not using it. The cloud can be both software and infrastructure. It can be an application you access through the Web or a server like Gmail and it can be also an IT infrastructure that can be used as per user's request. Whether a service is software or hardware, the following is a simple test to determine whether that service is a cloud service: If you can walk into any place and sit down at any computer without preference for operating system or browser and access a service, that service is cloud-based. Cloud computing is high utility software having the ability to change the IT software industry and making the software even more attractive.[1]Generally, there are three measures used to decide whether a particular service is a cloud service or not:

- The service is accessible via a web browser or web services API.
- Zero capital expenditure is necessary to get started.
- You pay only for what you use.

Most cloud computing services fall into three broad categories: infrastructure as a service (IaaS), platform as a service (PaaS) and software as a service (SaaS). These are sometimes called the cloud computing stack, because they build on top of one another. Knowing what they are and how they are different makes it easier to accomplish your business goals.

Infrastructure-as-a-service (IaaS)
The capability provided to the consumer is to access all the processing, storage, networks and other many fundamental computing resources. Consumer [2] [3] is able to deploy arbitrary software, which can include operating systemsand applications.

Platform as a service (PaaS)
PaaS [2] provides all the resources that are required for implementation of applications and all services completely from the Internet.PaaS is designed to make it easier for developers to quickly create web or mobile apps, without worrying about setting up or managing the underlying infrastructure of servers, storage, network and databases needed for development.

Software as a service (SaaS)
Software-as-a-service (SaaS) is a method for delivering software applications over the Internet, on demand and typically on a subscription basis. With SaaS, cloud providers host and manage the software application and underlying infrastructure and handle any maintenance, like software upgrades and security patching. Users connect to the application over the Internet, usually with a web browser on their phone, tablet or PC.

II.PROBLEM FORMULATION

Cloud computing involves distributed technologies to satisfy a spread of applications and user desires. The main function of cloud computing is sharing of resources, software and applications which have an objective to reduce the overall cost. Moreover performance in terms of processing time, execution time, turnaround time and the waiting time should be achieved considerably. So there are various measures and technical challenges that has to be addressed like fault tolerance, virtual machine migration, and high accessibility. Moreover the central issue is the load balancing , it's the mechanism of distributing the load among numerous nodes of a distributed system to boost each resource utilization and job time interval whereas additionally avoiding a state of affairs wherever a number of the nodes square measure heavily loaded whereas alternative nodes square measure idle or doing little or no work.Load balancing [4] is completed with the assistance of load balancers wherever every incoming request is redirected and is clear to consumer who makes the request. Supported preset parameters, reminiscent of accessibility or current load, the load balancer uses numerous planning rule to work out that server ought to handle and forwards the request on to the chosen machines. The random arrival of load in such an environment can cause some machines to be heavily loaded while other machine is idle or only lightly loaded. It makes sure that every computing resource is distributed efficiently and fairly [5].The considered characteristics have an impact on cost optimization, which can be obtained by improved response time and processing time.

- In the existing mechanism, the shorter tasks have high priority and always get chance to execute. Due to this low priority tasks (longer jobs) have to wait for a long time. It reduces the overall throughput of the system.
- No suitable criteria has been defined for categorizing the task into High QOS value and low QOS value.
- VM's are categorized only on a single parameter which is MIPS. Multiple parameters like RAM and Bandwidth should also be considered for allocation of cloudlets to VM
- Extra overhead time is involved for processing the non-dominated task set.

III.METHODOLOGY

In cloud computing, the infrastructure, platform and softwarecan be used as services. It is the form of computing, in which customer need not own its infrastructureandpay-as-you-go method in which no resources are wasted , since users only pay for services procured, rather than provisioning for a certain amount of resources that may or may not be used. The computing resources are actually virtual machines. Task scheduling algorithms play an important role in these kind of scenarios whose aim is to schedule the tasks effectively that also helps to reduce the execution time, the turnaround timeandimprove resource utilization. Load balancing is a computer network method for distributing workloads across multiple computing resources that impacts the performance of filesystem. Load balancing techniquesalso improves theefficiency of the file system. The main objective of this paper was to work with load balancing and fault tolerance. In this system to develop such a file system which can execute N number of jobs onprocessors which can take less time and work more. The proposed algorithm adds clustering approach in which VMs with similar capacities are divided into groups. K-means clusteringapproach is used to divideVMs intoclusters. Alist of all the clusters with the minimum and maximum resource specific capacities of each cluster will be maintained by load balancer which is known as range specifierlist.Moreover, load balancer will also maintain the list of VMS for each cluster. The approach considers resource specific demands which helps in reducing the overhead of scanning the entire list of VMs from the beginning. The procedure for the above approach is as follows:-

Step1: Initialize all VMs with their specific resource types, capacities of each resource and status of VMs.
Step2: Cluster the n VMs into k clusters using K-means clustering using the three resource types as parameters i.e. CPU processing speed, Memory and network bandwidth.
Step3: Cloud controller receives a new request
Step4: Cloud controller queries appropriate node controller/load balancer for next allocation.
Step5:Load balancer scans the range specifier list of k clusters to see that which cluster can handle the incoming request.
Step6: Load balancer will then assign the request to the appropriate VM of the chosen cluster by looking into the list of cluster members which will match the specific demands of the task and whose status is available. In case more than one VMs satisfy this, then the first one which is found will get the task.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

Step7: Remaining resource quantities of that VM in the VM list of that cluster is then updated.
Step 8: Status of that VM is changed from A V AILABLE to BUSY.
Step9: When the VM finish processing the request, the status of that VM is changed to AVAILABLE.
Step 10: The load balancer also updates the capacity of that VM in the VMs
capacities list.

Euclidian distance formula is chosen to assign VMs to the clusters. The value of K i.e. the number of clusters is chosen to be the highest prime factor of n where n is the number of VMs.
Eucledian distance is given below to find the distance of VMs:

$$EUD(VM)(Cj)=sqrt[(CPUi-CPUj)^2+(Memi-Memj)^2+(BWi-BWj)^2]$$

A new mean of each cluster can be found when a machine gets assigned to itwhich is given below.
• CPUj=(CPUi+CPUj)/2
• Memj=(Memi+Memj)/2
• BWj=(BWj+BWj)/2

When a user submits a task along with its specific requirements, load balancer matches the task with the clusters of specific ranges so as to assign the task to appropriate cluster. Now as the clusters are having virtual machines of different capacities, the load balancer seeks the suitable available machine to which task is assigned. This approach reduces the scanning time and will assign a better VM according to the requirements.

Fault tolerance can be achieved by creating a backup task of each cloudlet which is going to execute by the VM.If Task T1 executes successfully on VM, then back up of task T1 is no longer required and therefore, backup task is deleted successfully.If VM which is executing the current task has become faulty, then the backup task of T1 will be executed by another VM depending upon its availability.The VM which is failed is added to the blacklisted table so that no other cloudlet is further assigned to the faulty VM.The same procedure will be repeated for all the remaining cloudlets.

IV.SIMULATION IN CLOUD: CLOUDSIM

In CloudSim, cloud computing infrastructures and application services allowing its users to focus on specific system design issues that they want to investigate [8]. Simulation in a CloudSim means implementationof actual environment towards benefit of research. The users or researcher actually analyze the proposed design or existing algorithms through simulation. Resources and software are shared on the basis of client's demand in cloud environment.CloudSim helps the researchers to focus on specific system design issues without getting concerned about the low level details related to cloud-based infrastructures and services [6]. Essentially, dynamic utilization of resources is achieved under different conditions with various previous established policies. Sometime it is very much difficult and time consuming to measure performance of the applications in real cloud environment. In this consequence, simulation is very much helpful to allow users or developers with practical feedback inspite of having real environment. In this research work, simulation is carried out with a specific cloud simulator, CloudSim [7].
Data center: Data center encompasses a number of hosts in homogeneous or heterogeneous configurations . It also creates the bandwidth, memory, and storage devices allocation.

Virtual Machine (VM): VM characteristics comprise of memory, processor, storage, and VM scheduling policy.

Host: This simulation considers VM need to handle a number of cores to be processed and host should have resource allocation policy to distribute them in these VMs. So host can arrange sufficient memory and bandwidth to the process elements to execute them inside VM.

Cloudlet: Cloudlet is an application component which is responsible to deliver the data in the cloud service model. It also contains various ids for data transfer and application hosting policy

V.SIMULATION RESULTS

The simulation results has been carried out by implementing the proposed algorithm and existing algorithm.In this simulation we have focused on turnaround time, waitingtime, execution time and processing cost.VMs considered in this simulation are 20.Time measured in this simulation is in seconds. Simulation parameters have been mentioned in table 1:

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

Table 1: Simulation parameters

Parameter	Value
Number of VMs	20
Number of Clusters	2
Number of cloudlets	5-90000
Host memory	92048MB
Host bandwidth	10000mbps

Table 2: Simulation results

No of Cloudlets	Total Turn Around Time (in seconds)	Total Execution Time(in seconds)	Total Waiting Time(in seconds)	Total Processing Cost(in dollars)
5	1.7	1.7	0	10.2
10	3.47	3.4667	0	20.8
40	26.87	16.6817	10.19	100.09
60	56.29	25.59	30.7	153.54
100	141.3	43.04	98.26	258.24
150	304.99	64.78	240.21	388.68
200	524.5	86.27	438.23	517.62
300	1162.65	130.06	1032.59	780.36
400	2036.37	173.565	1862.81	1041.39
500	3161.65	217.2383	2944.41	1303.43
700	6132.58	303.9849	5828.6	1823.91
1000	12442.38	434.5599	12007.82	2607.36
2000	49451.72	869.9532	48581.77	5219.72
3000	111031.5	1305.366	109726.1	7832.2
5000	307674.5	2175.613	305498.9	13053.68
7000	602598.3	3046.211	599552.1	18277.27
8000	786957.3	3481.623	783475.7	20889.74
9000	995895.3	3917.066	991978.2	23502.4
10000	1229205	4352.243	1224853	26113.46
20000	4913147	8704.952	4904442	52229.71
30000	11053091	13058.3	11040033	78349.79
40000	19645584	17410.56	19628174	104463.4
50000	30693840	21763.29	30672077	130579.8
60000	44199349	26116.63	44173232	156699.8
70000	60155135	30468.89	60124666	182813.3
80000	78567700	34821.62	78532878	208929.7
90000	99438837	39174.99	99399662	235050

Processing Time/ Execution Time:

The time taken by the allocated VM to execute the task is called execution time. Figure 1.depicts the totalprocessing time / execution time for the base algorithm and proposed algorithm for the given number of cloudlets.

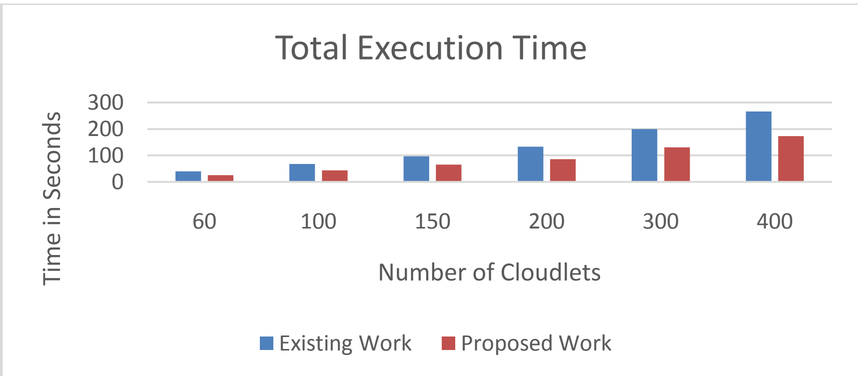

Figure 1: Execution Time for Existing work and proposed algorithm.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

Waiting Time:
Waiting time is the amount of time that is taken by a load balancer to allocable VM to task. Waiting time=Allocation time-Entry time. Figure 2.depicts the waiting time for the base algorithm and proposed algorithm for the given number of cloudlets.

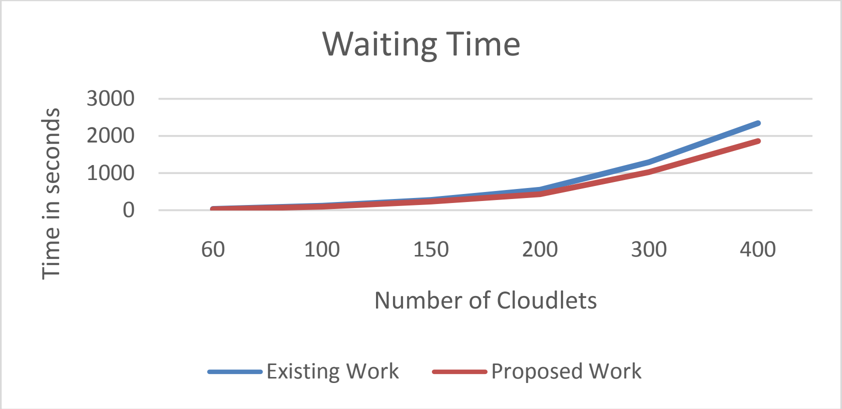

Figure 2: Waiting Time for Existing work and proposed algorithm.

Turnaround Time:
Time taken by the system from the submission of request to the generation of response is called turnaround time. Figure 3.depicts the turnaround time for base algorithm and proposed algorithm.

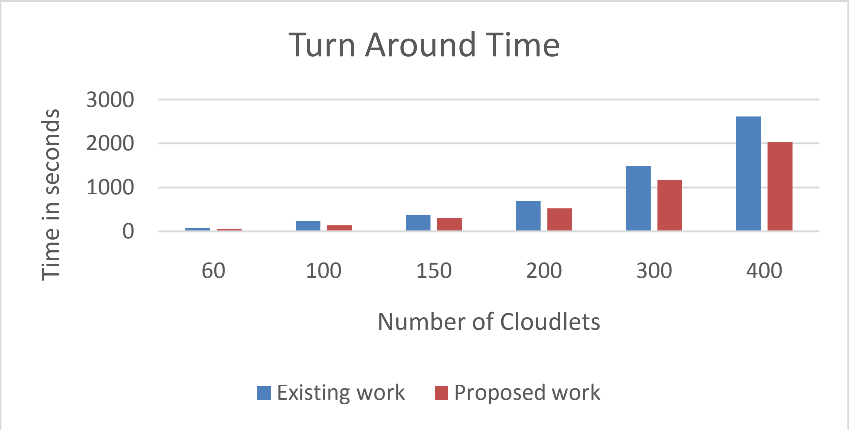

Figure 3. Turnaround time for Existing work and proposed algorithm.

Processing Cost:
The processing cost is calculated on the basis of actualCPU time when the tasks are finished to execute on thecloud resources and the cost of resources per second.

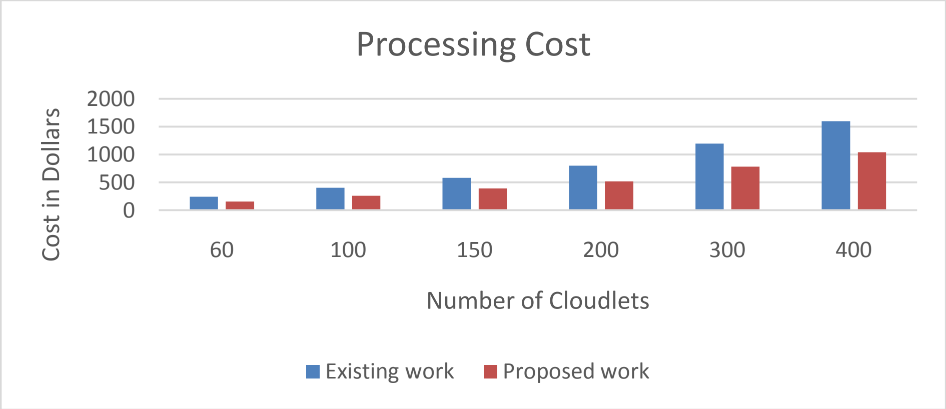

Figure 4: Processing cost for the existing work and proposed algorithm

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

From the figure 4, it is clear that the processing cost incurred to the user is lesser in the proposed algorithm than in the base algorithm. Lesser the processing cost, more will be the benefit to the user.

VI.CONCLUSION

In this paper, we have implemented load balancing algorithm by making clusters in cloudsimtoolkit. The main approach of this algorithm is to make the proper utilization of resources by allocating the tasks to suitable VMs of specified requirements.The research work has involved developing of an efficient VM load balancing algorithm and conducting a comparative analysis of the proposed algorithm with the existing algorithm. By observing the above cited parameters in tables and graphs we can easily observe that the overall turnaround time, processing time, waiting time and processing cost is improved in comparison to the existing scheduling parameters.

REFERENCES

[1] S. Kapoor and D. C. Dabas, "Cluster Based Load Balancing in Cloud Computing," IEEE, 2015.
[2] Haozheng Ren. “The load balancing Algorithm in cloud computing Environment” 2nd International Conference on computer science and network technology 2012.
[3] Tushar Desai. “A survey of various load balancing techniques and challenges in cloud computing” International Journals of scientific and technology research volume 2. Issue11,Nov2013.
[4] Raju, R., Amudhavel, J., Kannan, N., and Monisha, M. (2014). A bio inspired energy- Awaremulti objective chiropteran algorithm for hybrid cloud computing environment.In Green Computing Communication and ElectricalEngineering (ICGCCEE), 2014 InternationalConference on, pages 1-5. IEEE.
[5] S. S. MOHARANA, R. D. RAMESH and D. POWAR, "ANALYSIS OF LOAD BALANCERS IN CLOUD COMPUTING," International Journal of Computer Science and Engineering (IJCSE) , pp. 102-107, 2013.
[6] N. K. Chien, N. H. Son and H. D. Loc, "LoadBalancing Algorithm Based on Estimating Finish Time of Services in Cloud Computing," ICACT, pp. 228-233, 2016.
[7] Domanal, S. G. and Reddy, G. R. M. (2013). Load balancing in cloud computing usingmodified throttled algorithm In Cloud Computing in Emerging Markets (CCEM), 2013 IEEE International Conference on, pages 1-5. IEEE.
[8] Domanal, S. G. and Reddy, G. R. M. (2014). Optimal load balancing in cloud computingby efficient utilization of virtual machines. In Communication Systems and Networks (COMSNETS), 2014 Sixth International Conference on, pages 1-4. IEEE.
[9] VishwasBagwaiya, “A hybrid approach using throttled and esce load balancing algorithms in cloud computing.”
[10] Tushar Desai. “A survey of various load balancing techniques and challenges in cloud computing” International Journals of scientific and technology research volume 2,Issue11,Nov2013.
[11] UpendraBhoi. “Enhanced max-min Task scheduling Algorithm in cloud computing”. International Journal of Application or Innovation in Engineering &management (IJAIEM), April 2013.
[12] GytisVilutis, “Model of load balancing and scheduling in cloud computing”. Proceedings of the ITI 2012 34th Int.Conf. on Information Technology Interfaces, June 25-28,Cavat,Croatia.
[13] R.-S. Chang, J.-S. Chang and P.-S. Lin, "An ant algorithm for balanced job scheduling in grids," Future Generation .Computer Systems 25 (2009) 20–27, pp. 21-27, 2009.
[14] Z. Chaczko, V. Mahadevan, S. Aslanzadeh and C. Mcdermid, "Availability and Load Balancing in Cloud Computing," International Conference on Computer and Software Modeling IPCSIT vol.14 (2011) © (2011) IACSIT Press, Singapore, pp. 134-140, 2011.
[15] Nguyen, QuyetThang; Quang-Hung, Nguyen; Tuong, Nguyen Huynh; Tran,VanHoai; Thoai, Nam, “Virtual machine allocation in cloud computing for minimizing total execution time on each machine,”Computing, Management andTelecommunications (ComManTel), 2013 International Conference on , vol.,no., pp.241,245, 21-24 Jan. 2013.
[16]. Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov, Cesar A. F. De Rose,andRajkumarBuyya, “CloudSim: A Toolkit for Modeling and Simulationof Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms,”Software: Practice and Experience (SPE), Volume 41, Number 1,Pages: 23-50, ISSN: 0038-0644, Wiley Press, New York, USA, January, 2011.
[17] Kumar Nishant, , P. Sharma, V. Krishna, Nitin and R. Rastogi, "Load Balancing of Nodes in Cloud Using Ant Colony Optimization," IEEE, pp. 3-9, 2012.
[18] Y. Xu, L. Wu, L. Guo,, Z. Chen, L. Yang and Z. Shi, "An Intelligent Load Balancing Algorithm Towards Efficient Cloud Computing," AI for Data Center Management and Cloud Computing: Papers from the 2011 AAAI Workshop (WS-11-08), pp. 27-32, 2011.