Cloud Computing Load Balancing Algorithms Comparison Based Survey

Sarika Vasantrao Bodake

Department of Computer Engineering, Asst. Prof., Padmabhooshan Vasantdada Patil Institute of Technology, Bavdhan, Pune, India

***Abstract*— *Cloud computing is an online primarily based computing. This computing paradigm has increased the employment of network wherever the potential of 1 node may be used by alternative node. Cloud provides services on demand to distributive resources like info, servers, software, infrastructure etc. in pay as you go basis. Load reconciliation is one amongst the vexing problems in distributed atmosphere. Resources of service supplier have to be compelled to balance the load of shopper request. Totally different load reconciliation algorithms are planned so as to manage the resources of service supplier with efficiency and effectively. This paper presents a comparison of assorted policies used for load reconciliation.*

***Keywords*— Cloud computing, virtual machines, cloud service provider, load balancing, cloud analyst.**

I. INTRODUCTION

Cloud computing is one amongst the net primarily based service supplier that permits users to access services on demand [1]. It provides pool of shared resources of data, software, databases and different devices in keeping with the consumer request on “pay as you go” basis [2]. Cloud computing architectures are inherently parallel, distributed and serve the requirements of multiple shoppers in several eventualities. This distributed design deploys resources distributive to deliver services with efficiency to users in several geographical channels [3]. Shoppers in a very distributed atmosphere generate request indiscriminately in any processor. The major disadvantage of this randomness is related to task assignment. The unequal task assignment to the processor creates imbalance i.e., a number of the processors are full and a few of them are under loaded [4]. The target of load equalisation is to transfer the load from full method to under loaded method transparently. The key problems related to load equalisation are flexibility and reliableness of services [5]. Cloud computing implements virtualization technique within which one system will be virtualized into variety of virtual systems [6]. Load equalisation decides that consumer can use the virtual machine and that requesting machines are going to be placed on hold. Load equalisation of the complete system will be handled dynamically by mistreatment virtualization technology wherever it becomes potential to remap Virtual Machines (VMs) and physical resources in keeping with the amendment in load. Attributable to these blessings, virtualization technology is being comprehensively enforced in Cloud computing. A Virtual Machine (VM) could be a code implementation of a computing atmosphere within which associate software (OS) or program will be put in and run. The Virtual Machine usually changes a physical computing atmosphere and requests for central processor, memory, hard disk, network and different hardware resources are managed by a virtualization layer that interprets these requests to the underlying physical hardware. VMs are created among a virtualization layer, like a hypervisor or a virtualization platform that runs on prime of a consumer or server software. This software is understood because the host OS. The virtualization layer will be wont to produce several individual, isolated VM environments, wherever multiple requests or tasks will execute in multiple machines [7].

1.1 Load equalisation algorithms

Load equalisation [8] algorithmic rule directly influences the impact of equalisation the server workloads. Its main task is to determine the way to choose consequent server node and transfer a replacement affiliation request to that. Current main load equalisation algorithmic rule is split into static algorithmic rule and dynamic algorithmic rule [9]. The static algorithmic rule is well carried into execution and takes less time, which does not consult with the states of the load nodes; however it will be solely employed in bound specific conditions. The common static algorithms are Round-Robin programming algorithmic rule, Weighted Round-Robin programming algorithmic rule, and Least-Connection programming algorithmic rule etc. Of all, spherical - Robin programming algorithmic rule is that the simplest one that can be most simply be allotted. However, it's solely applicable to the circumstances within which all the nodes in cluster have identical process ability. The dynamic algorithmic rule like 1st come back 1st serve is self-adaptive algorithmic rule, that is best than static algorithmic rule, and appropriate for a good deal of requests that multiply completely different workloads, which might be unable to be forecasted [10]. Self-adaptive load equalisation system primarily includes 2 processes: observation the load states of servers and assignment the request to the servers. The state supervising, that depends on the load info of every node within the cluster monitored and picked up sporadically by the front-end computer hardware, raises the impact of load balance by observation load selection. At identical time, assignment the load carries on operation in keeping with the load info from all nodes, that is, redistributing the load that has to be done.

According to the analysis on top of, the perfect load equalisation algorithmic rule ought to attain the subsequent targets:

- Leave the collections, computing of load node info for every node; stop the front-end computer hardware from being system bottleneck.
- Scale back the complications of load equalisation algorithmic rule as so much as potential.

II. RELATED WORK

Cloud computing is recent rising technology in IT trade leading towards the researches advances in several domains. Jiyni et.al, (2010) have projected a resource allocation mechanism with preemptable task execution that will increase the employment of clouds. They need projected associate degree adaptive resource allocation algorithmic program for cloud system with preemptable tasks however their approach doesn't pertain to price improvement and time improvement [11]. M. Randles etal have projected comparison of static and dynamic load equalisation algorithmic program for cloud computing. [12]. Ram Prasad Padhy, P Goutam Prasad Rao mentioned on basic ideas of Cloud Computing and cargo equalisation and studied some existing load equalisation algorithms, which may be applied to clouds [5]. Additionally thereto, the closed-form solutions for minimum menstruations and coverage time for single level tree networks with completely different load equalisation methods were additionally studied [13]. The performance of those methods with relation to the temporal arrangement and also the impact of link and menstruations speed were studied. The papers delineate the options of a machine to match the performance of 3 dynamic load equalisation algorithms. Cloud Analyst: A Cloud Sim-based Visual creator for Analysing Cloud Computing Environments and Applications [14] Bhathiya Wickrema Singh all gifts however Cloud Analyst will be wont to model and appraise a true world drawback through a case study of a social networking application deployed on the cloud. We've illustrated however the machine will be wont to effectively determine overall usage patterns and the way such usage patterns have an effect on knowledge centres hosting the applying [15].

III. THE ALGORITHMS

3.1 Throttled

In this rule the shopper 1st requests the load balancer to search out an acceptable Virtual Machine to perform the specified operation. It's shown in figure 1.The method 1st starts by maintain an inventory of all the VMs every row is severally indexed to hurry up the operation process. If a match is found on the idea of size and accessibility of the machine, then the load balancer accepts the request of the shopper and allocates that VM to the shopper. If, but there's no VM obtainable that matches the factors then the load balancer returns -1 and therefore the request is queued. The subsequent figure shows however it works [16].

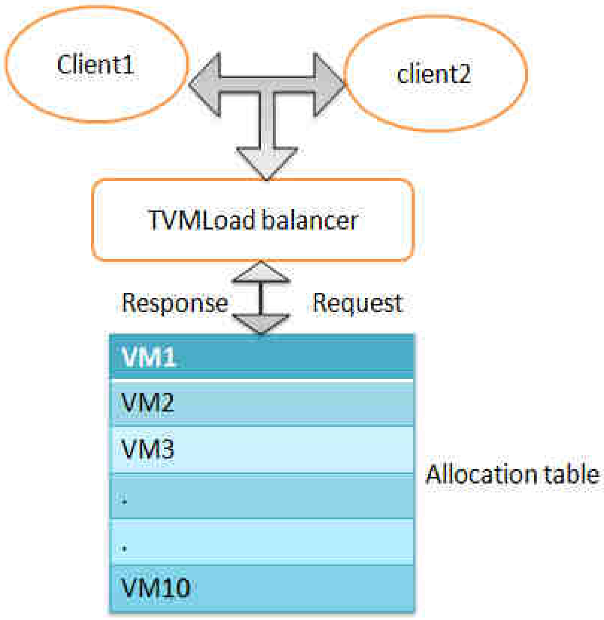

Fig. 1: Throttled Scheduling Process

3.2 Round Robin

It is one amongst the best programming techniques that utilize the principle of your time slices. Here the time is split into multiple slices and every node is given a specific time slice or quantity i.e. it utilizes the principle of your time programming [16]. Every node is given a quantum and during this quantum the node can perform its operations. The resources of the service supplier area unit provided to the requesting shopper on the idea of this point slice. The subsequent figure shows however spherical robin works. The subsequent figure 2 shows each user request is served by every processor among given time quantum. When the time slice is over, following queued user request can come back for execution. If the user request completes among time quantum then user mustn't wait otherwise user need to watch for its next slot.

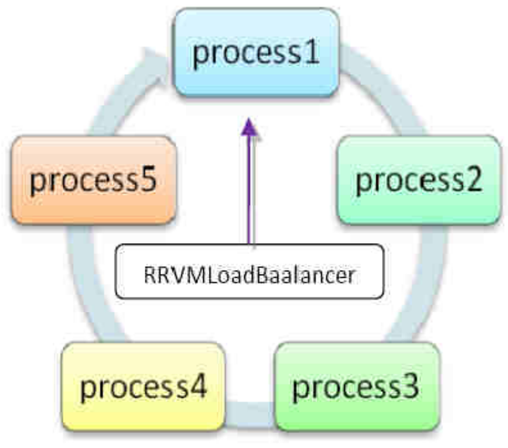

Fig. 2: Round Robin Process

3.3 FCFS

FCFS (First come back 1st Served), employed in parallel task process, is that the simplest task ordering strategy. It chooses and processes them in keeping with the proper order of jobs moving into system [18]. With this theme the user request that comes 1st to the data center controller is allotted the virtual machine for execution 1st. The implementation of FCFS policy is definitely managed with first in first out queue. The data center controller searches for virtual machine that is in idle state or under loaded. Then the first request from the queue is removed and passed to 1 of the VM through the VMLoadBalancer. The subsequent figure shows however FCFS works.

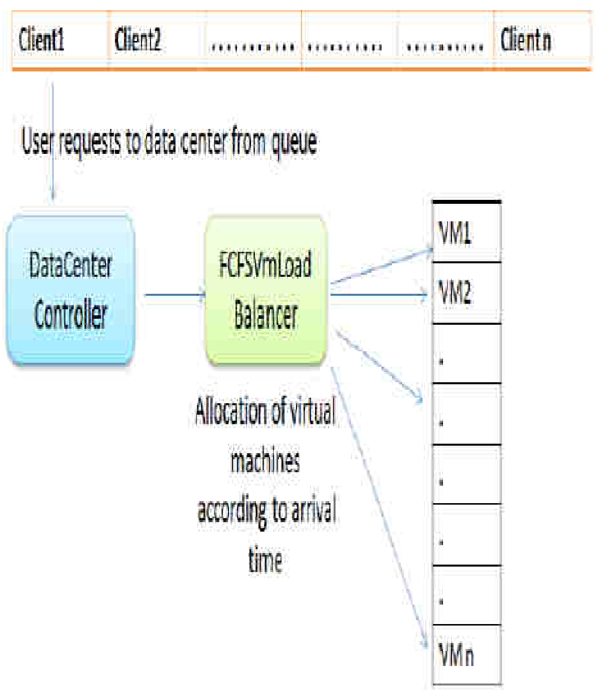

Fig. 3: FCFS Scheduling

3.4 ESCEL Scheduling

It is unfolding spectrum technique during which the load balancer unfolds the load of the duty in hand into multiple virtual machines. The load balancer maintains a queue of the roles that require to use and area unit presently victimisation the services of the virtual machine. The balancer then endlessly scans this queue and therefore the list of virtual machines. If there's a VM obtainable that may handle request of the node/client, the VM is allotted thereto request [17]. If but there's a VM that's free and there's another VM that must be freed of the load, then the balancer distributes a number of the tasks of that VM to the free one thus on cut back the overhead of the previous VM. The roles area unit submitted to the VM manager, the load conjointly maintains an inventory of the roles, their size and therefore the resources requested. The balancer selects the duty that matches the factors for execution at the current time. Though there rule offers higher results as shown in more section, it but needs lots of process overhead. The subsequent figure shows however ESCEL works.

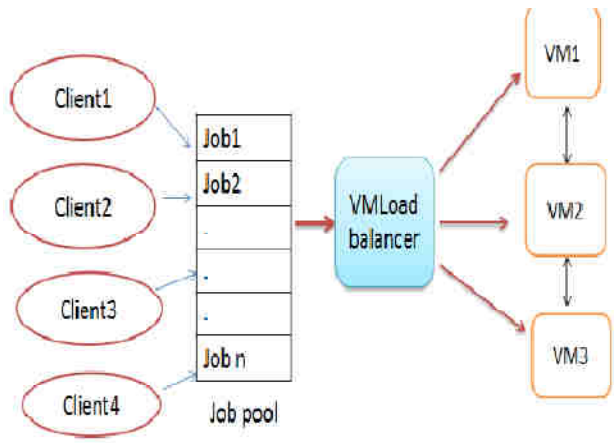

Fig. 4: ESCEL Scheduling

IV. CLOUD ANALYST

Cloud Analyst [14] [16] [17] could be a graphical user interface primarily based tool that's developed on CloudSim design. CloudSim could be a toolkit that enables doing modelling, simulation and alternative experimentation. the most downside with CloudSim is that everyone the work ought to be done programmatically. It permits the user to try and do perennial simulations with slight modification in parameters terribly simply and quickly. The cloud analyst permits setting location of users that area unit generating the appliance and additionally the situation of the info centers. during this varied configuration parameters are often set like range of users, range of request generated per user per hour, range of virtual machines, range of processors, quantity of storage, network information measure and alternative necessary parameters. supported the parameters the tool computes the simulation result and shows them in graphical kind. The result includes interval, interval, cost etc .By acting varied simulations operation the cloud supplier will confirm the simplest thanks to portion resources, supported request that knowledge center to be hand-picked and might optimize value for providing services.

4.1 Simulation Parameters

4.1.1 Region

In the CloudAnalyst the planet is split in half-dozen|to six} 'Regions' that coincide with the 6 main continents within the World. the opposite main entities like User Bases and knowledge Centers belong to 1 of those regions. This geographical grouping is employed to keep up level of realistic simplicity for the big scaled simulation being tried within the CloudAnalyst.

4.1.2 Users

A User Base models a bunch of users that's thought-about as one unit within the simulation and its main responsibility is to come up with traffic for the simulation. One User Base might represent thousands of users however is organized as one unit and also the traffic generated in coincident bursts representative of the dimensions of the user base. The creator might opt to use a User Base to represent one user, however ideally a User Base ought to be wont to represent a bigger range of users for the potency of simulation.

4.1.3 DataCenterController

The Data Center Controller is maybe the foremost vital entity within the CloudAnalyst. One knowledge Center Controller is mapped to one cloudsim. DataCenter object and manages the info center management activities like VM creation and destruction and will the routing of user requests received from User Bases via the web to the VMs. It can even be viewed because the façade utilized by CloudAnalyst to access the center of CloudSim toolkit practicality.

4.1.4 InternetCharacteristics

In this part varied web characteristics area unit modelled simulation, which has the number of latency and information measure ought to be assigned between regions, the number of traffic, and current performance level data for the info centers.

4.1.5 VmLoadBalancer

The responsibility of this part is to portion the load on varied knowledge centers per the request generated by users. one in all the f our given policies are often hand-picked. The given policies area unit spherical robin formula, equally unfold current execution load, throttled and initial return initial serve.

4.1.6 CloudAppServiceBroker

The responsibility of this part is to model the service brokers that handle traffic routing between user bases and knowledge centers. The service broker will use one in all the routing policies from the given 3 policies that area unit nearest knowledge center, optimize interval and reconfigure dynamically with load. The nearest knowledge center routes the traffic to the nearest knowledge center in terms of network latency from the supply user base. The reconfigure dynamically with load routing policy works within the sense that whenever the performance of explicit knowledge center degrades below a given threshold price then the load of that knowledge center is equally distributed among alternative knowledge centers.

In order to investigate varied load equalization policies configuration of the assorted part of the cloud analyst tool ought to be done. We've set the parameters for the user base configuration, application readying configuration, and knowledge center configuration as shown in figure six. As shown in figure the situation of user bases has been outlined in six totally different regions of the planet. We've taken four knowledge centers to handle the request of those users. On DC_1 there are a unit twenty five VMs allotted, fifty VMs area unit allotted to DC_2, seventy five VMs area unit allotted to DC_3 and a hundred VMs area unit allotted to DC_4 severally. Here we've taken six user bases.

Table 1: Setting User Bases and Data Center

Data Center	VMs	Image Size	Memory	BW
DC_1	25	10000	512	1000
DC_2	50	10000	512	1000
DC_3	75	10000	512	1000
DC_4	100	10000	512	1000

USERS	REGION	REQUET PER USER PER HR	DATA SIZE PER REQUEST (BYTES)	PEAK HOURS START (GMT)	PEAK HOURS END (GMT)	AVG. PEAK USERS	AVG. OFF-PEAK USERS
UB1	0	60	100	3	9	1000	100
UB2	1	60	100	3	9	2000	100
UB3	2	60	100	3	9	3000	100
UB4	3	60	100	3	9	4000	100
UB5	4	60	100	3	9	5000	100
UB6	5	60	100	3	9	6000	100

V. RESULT AND RESPONSE TIME

After performing arts the simulation the result computed by cloud analyst is as shown within the following figures. The higher than outlined configuration has been used for every load reconciliation policy one by one and looking on that the result calculated for the metrics like reaction time, request time interval and value in fulfilling the request has been shown. Parameters like average reaction time, knowledge center service time and total price of various knowledge centers have taken for analysis.

Table 2: Comparison of Algorithms Based on Average Response Time

UBS	RR	TR	FCFS	ESCEL
UB1(1000)	50.64	50.65	50.63	50.64
UB2(2000)	51.16	51.15	51.15	51.18
UB3(3000)	51.85	51.82	51.85	51.83
UB4(4000)	52.48	52.49	52.48	52.47
UB5(5000)	301.49	301.56	301.54	301.57
UB6(6000)	200.91	200.90	200.89	200.90

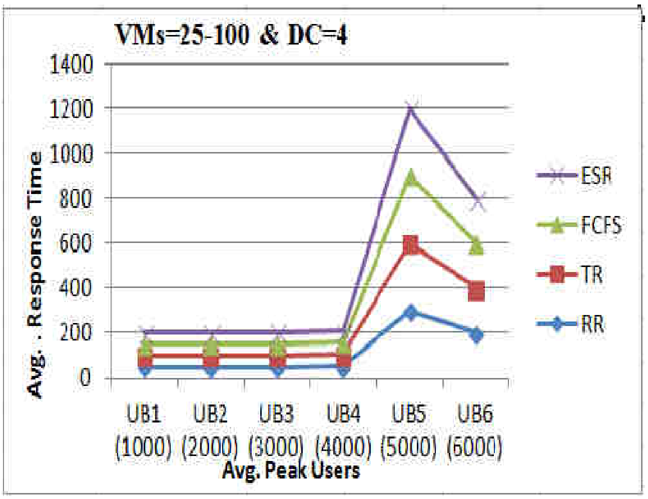

Fig. 5: Result (Avg. Peak Users vs. Avg. Response Time)

Table 3: Avg. Data Center Request Servicing Time

Data Center	RR	TH	FCFS	ESCEL
DC_1	0.786	0.785	0.783	0.784
DC_2	1.567	1.574	1.578	1.577
DC_3	2.1	2.1	2.093	2.097
DC_4	2.773	2.769	2.75	2.77

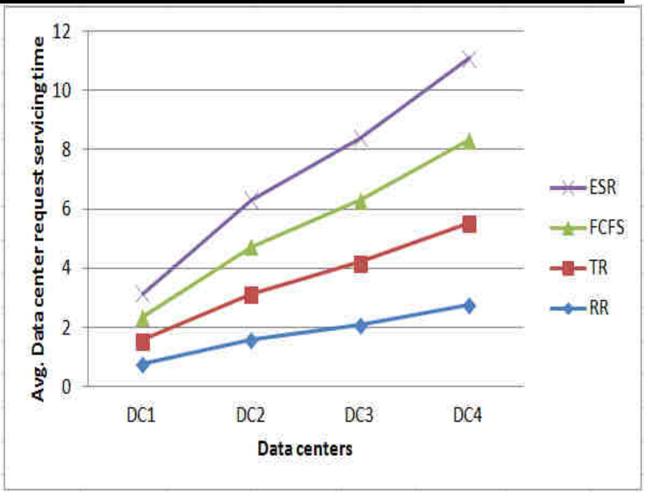

Fig. 6: Avg. Data Center Servicing Time

Table 4: Total Cost Analysis

Data Center	RR	TH	FCFS	ESCEL
DC_1	2.64	2.64	2.64	2.64
DC_2	5.08	5.08	5.08	5.08
DC_3	7.65	7.65	7.65	7.65
DC_4	10.10	10.10	10.10	10.10

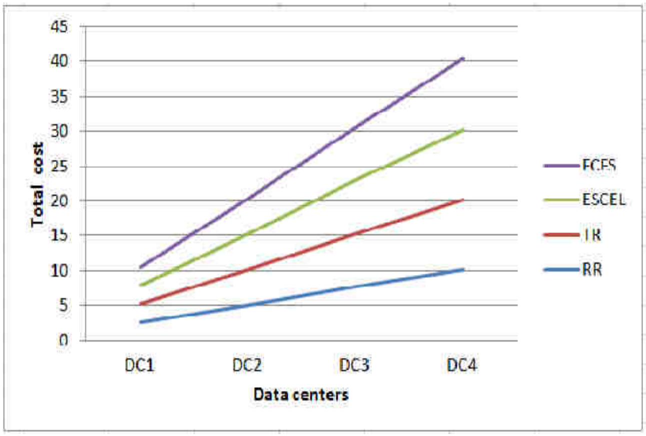

Fig. 7: Costing of Data Center

VI. CONCLUSION

We have simulated four completely different programming algorithms have for corporal punishment the user request in cloud atmosphere. Every rule is ascertained and their programming criteria like average interval, information center service time and total value of various information centers square measure found. In step with the experiment and analysis spherical robin rule has the most effective integrate performance. Future work is supported this rule changed and enforced for real time system. Higher interval is expected if we have a tendency to apply some organic process algorithms like PSO, ACO, and ABC's rather than classical algorithms.

REFERENCES

[1] G. Pallis, “Cloud Computing: The New Frontier of Internet Computing”, IEEE Journal of Internet Computing, Vol. 14, No. 5, September/October 2010, pages 70-73.

[2] Qi Zhang, Lu Cheng, Raouf Boutaba, “cloud computing: state of-the-art and research challenges”, 20th April 2010, Spinger, pp. 7-18.

[3] M. D. Dikaiakos, G. Pallis, D. Katsa, P. Mehra, and A. Vakali, “Cloud Computing: Distributed Internet Computing for IT and Scientific Research”, IEEE Journal of Internet Computing, Vol. 13, No. 5, September/October 2009, pages 10-13.

[4] A. Khiyaita, H. EI Bakkli, M. Zbakh ,Dafir EI Kettani,” Load Balancing Cloud Computing: State Of Art”, 2010,IEEE.

[5] Ram Prassd Pandhy (107CS046), P Goutam Prasad rao (107CS039). “Load balancing in cloud computing system” Department of computer science and engineering National Institute of Technology Rourkela, Rourkela-769008, Orissa, India May-2011.

[6] J. Sahoo, S. Mohapatra and R. lath “Virtualization: A survey on concepts, taxonomy and associated security issues” computer and network technology (ICCNT), IEEE, pp. 222-226. April 2010.

[7] Bhaskar. R, Deepu.S. R and Dr.B. S. Shylaja “Dynamic Allocation Method For Efficient Load Balancing In Virtual Machines For Cloud Computing Environment” September 2012.

[8] R.Shimonski. Windows 2000 & Windows server 2003 clustering and load balancing. Emeryville. McGraw-Hill Professional publishing,CA,USA(2003), p 2,2003.

[9] R.X.T. and X. F.Z..A load balancing strategy based on the combination of static and dynamic, in database technology and applications (DBTA), 2010 2nd international workshops, (2010), pp. 1-4.

[10] Wenzheng Li, Hongyan Shi “Dynamic Load Balancing Algorithm Based on FCFS” IEEE, 2009. pp.1528-1531.

[11] Jiyni Li, Meikang Qui, Jain-Wei Niu, Yuchen, Zhong Ming “Adaptive resource allocation for preemptable jobs in cloud system”. IEEEInternational Conference on intelligent system design and applications, pp. 31-36, 2010.

[12] M Randles, D. Lamb, and A. Taleb-Bendiab, “A comparative study into distributed load balancing algorithms for cloud computing,” 2010 IEEE 24th international conference on advanced information networking and application workshops,2010, pp. 551-556.

[13] Jaspreet Kaur “Comparision Load Balancing Algorithms In A Cloud” International Journal Of Engineering Research And Applications. pp. 1169-1173, 2012.

[14] Bhathiya, Wickremasinghe.”Cloud Analyst: A Cloud Sim-based Visual Modeller for Analysing Cloud Computing Environments and Applications”, 2010, IEEE.

[15] Dr. Hemant, S. Mahalle, Prof. Parag R. Kaveri and Dr. Vinay chavan, “Load Balancing On Cloud Data Centers”. International Journal of Advanced Research in Computer Science and Software Engineering, Jan-2013.

[16] A. Singh, P. Goyal, S. Batra : Anoptimized round robin scheduling algorithm for CPU scheduling, International journal of computer and electrical engineering (IJCEE), vol. 2, No. 7, pp 2383- 2385, December, 2010.

[17] Tanvee Ahmed, Yogendra Singh “Analytic Study Of Load Balncing Techniques Using Tool Cloud Analyst” . International Journal Of Engineering Research And Applications. pp. 1027-1030, 2012.

[18] A. Khiyati, M. Zbakh, H. El Bakkali, D. El Kettani “Load Balancing Cloud Computing: State Of Art”IEEE, 2012.

[19] R. Buyya, R. Ranjan, and R. N. Calheiros, “Modeling and Simulation of Scalable Cloud Computing Environments and the CloudSim Toolkit: Challenges and Opportunities,” Proc. of the 7th High Performance Computing and Simulation Conference (HPCS 09), IEEE Computer Society, June 2009.