Optimized Load Balancing in Cloud Computing using Hybrid Approach of Round Robin and Min-Min Scheduling

Muskan Garg[1], Rajneesh Narula[2]

Research Scholar IT Department, Adesh Institute of Engg.& Technology

HOD of IT& CSE department, Faridkot Adesh Institute of Engg.& Technology, Faridkot

Abstract: *Cloud computing is a new emerging are in academics and industry. In cloud computing various service providers provide different services to customers for data storage, data processing. Cloud analyst is tool that is much useful for development and simulation of cloud environment before actual deployment is real world application. In the processing of cloud analyst various service broker polices and load balancing policies have been used for response of different users. In this paper a new hybrid approach has been purposed for load balancing in cloud computing. This approach use hybrid allocation VM's to different user base requests.*

Keywords: load balancing and cloud computing, Round-Robin and min-min algorithm, response time, Data Center Response time and Load

1. Introduction

Cloud computing is a new technology .It providing online resources and online storage to the user's .It provide all the data at a lower cost. In cloud computing users can access resources all the time through internet. They need to pay only for those resources as much they use .In Cloud computing cloud provider outsourced all the resources to their client. There are many existing issues in cloud computing. The main problem is load balancing in cloud computing. Load balancing helps to distribute all loads between all the nodes. It also ensures that every computing resource is distributed efficiently and fairly. It helps in preventing bottlenecks of the system which may occur due to load imbalance. It provides high satisfaction to the users. Load balancing is a relatively new technique that provides high resource utilization and better response time. [1] [2] [3] [4] Cloud computing provide many advantages to the users.

A. Cloud computing consists of several characteristics: [5] [6].
- On demand service- Cloud computing provide services to users on their demand.Users can access the services as they want.
- Broad Network Access- In cloud computing capabilities are available over the network.All the capabilities are accessed through different mechanisms.
- Resource Pooling- Different models are used to pooled the resources which provide by the providers to their consumers. All the resources dynamically assigned and reassigned according to consumer demand.
- Rapid Elasticity- Quantity of resources is increase at any time according to the customer's requirements.
- Measured Service- In cloud computing resource usage can be monitored, controlled for both the provider and consumer of the all service.

B. Challenges in Cloud Computing there are many challenges in cloud computing:

1) Security
2) Efficient load balancing
3) Performance Monitoring
4) Consistent and Robust Service abstractions
5) Resource Scheduling
6) Scale and QoS management
7) Requires a fast speed Internet connection.

1.1 Goals of Load balancing:

1) To improve the performance substantially.
2) To have a backup plan in case the system fails even partially.
3) To maintain the system stability.
4) To accommodate future modification in the system.

2. Review of Literature

Er. AmandeepKaur et al [4] Author proposed that as the cloud computing is a new style of computing over internet. It has many advantages along with some crucial issues to be resolved in order to improve reliability of cloud environment. These issues are related with the load management, fault tolerance and different security issues in cloud environment. In this paper the main concern is load balancing and security issues in cloud computing. The load can be CPU load, memory capacity, completion time of each job and security issues to prevent the data from unauthorized user. From decades well known algorithms like FCFS, Priority has been seen into action to reduce the server load. But with the increase in the complexity of the server needs, they have failed to cope up with the current scenario. In this paper an approach has been that named Cross Breed Job scheduling technique which would be a combination of FCFS, Priority and would be monitored by RBAC (Role based access control). RBAC is a system which checks that whether the user of the system has the access to particular content or not. If the user doesn't have the access to the content, he will be denied and the server's load would be minimized.

Magade, Krishnanjali A. et al [5] The IEEE 802.11 standard does not provide any mechanism to resolve load imbalance in the network. To reduce this deficiency, various loadbalancing techniques have been designed. Loadbalancing provides a cost-effective, efficient and transparent method to expand the bandwidth of network devices and servers, increase the throughput, and enhance the data process capability, thus increasing the flexibility and availability of networks. There are different techniques based on persistent algorithm for loadbalancing in Wireless LAN. The goal of the proposed work is to use Persistence weighted round robin algorithm for loadbalancing in wireless LAN. This algorithm is able to distribute mobile stations among all APs and the signal strengths between stations and access points are also being maximized at the same time. This technique will be useful to reduce the congestion in the network, maintain the load in balance condition on network, as well as improve the bandwidth utilization.

Yean-Fu Wen et al [6] Author discuss about the loadbalancing problem is one of an open issues for the cloud computing. A good loadbalancing mechanism enhances the performance of network processing, optimizes the use of resources, and ensures that no overloading a single node or link case. The existing loadbalancing cloud computing research mainly unilateral the fairness of the transmission network or stand only for the system. Hence, this work considers the handling of both network and system loadbalancing to obtain high performance. To assign the tasks to the same type of nodes along the links with minimum processing and transmission delays subject to the capacities of nodes and links. Three task assignment schemes FCFS, Min-Min, and Min-Max are adopted along with dynamic clustering, which is a method to group the same type of cloud servers. This study changes in the variables manipulated with the number of nodes and the number of tasks and records the maximal end-to-end delay, average end-to-end delay and fairness index, to analyse the load balancing results. The results show that the Min-Max combination with dynamic clustering has a good effect.

De Mello, M.O.M.et al [7] The aim of the author to introduce a new loadbalancing routing heuristic, called BPR - Bottleneck, Path length and Routing overhead, which offers an efficient online solution by taking into consideration all these aspects of the loadbalancing routing problem. A simulation study has been carried out to compare the performance of BPR with a recent related work. BPR obtains bottleneck values within desirable bounds, while reducing the average path length. As a result, BPR also notably reduces the number of route updates in the network, i.e. the routing overhead. Finally, we show that BPR is simple and has low demand for processing requirements.
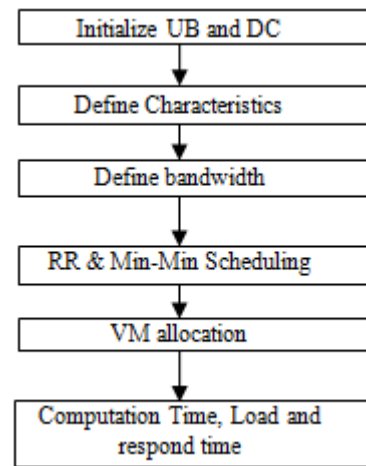
3. Methodology

Figure 3.1: Flow of purposed work

In the purposed work cloud job scheduling and load balancing policy has been used for allocation of virtual machine to different user bases requests. In this process user base has been deployed in different regions of the world that transmit their request to different datacenters allocated by third party cloud server. These datacenters have memory capacity, bandwidth and processors that divide their memory to different VM's for allocation to different user bases for respond their requests. In the purposed work cloud analyst tool has been used for responding simulation of cloud environment that use different cloud service model, load balancing policies and cloud service brokers.

Various components cloud-sim execution has been described below.

- **User Base**

User based is a component that used to represents a group of users, their requests; peak hours; region and size average peak users and average off peak users available that transmit their request on cloud environment.

- **DataCenterController**

Data center controller is the component that has been used to control all the activities that has been occurred on cloud environment. Datacenter memory, number of virtual machines, sharing policy has been decided through data centered controller. Virtual machine cost and data transmission cost has been defined by this module.

- **InternetCharacteristics**

Internet characteristics is component of cloud analyst tool that has been used for defining different bandwidth from one region to another region and for delay matric that can be allowed to data transmission between a datacenter to other datacenter and user base to datacenter.

- **VM LoadBalancer**

Load balancing policy is component that has been used in cloud job scheduling to manage load on different cloud center in such a way so that minimum time taken has been done by cloud service provider to respond user requests. In this paper round robin load balancing policy has been modified using nature inspired approaches that works in collaboration with Genetic algorithm and pollination based optimization approach.

- **CloudApplication ServiceBroker**

Cloud service broker policy is used to route traffic routing between different user bases and datacenters. These service broker policies have been used for data transmission from user base to datacenter on the basis of credentials available in service broker policy. Three main service broker policies have been defined by the cloud analyst simulation tools that are closest data center, optimize response time and reconfigure dynamically with load.

The closest data center routes the traffic to the closest data center in terms of network latency from the source user base. The reconfigure dynamically with load routing policy works in the sense that whenever the performance of particular data center degrades below a given threshold value then the load of that data center is equally distributed among other data centers

In the purposed work load balancing policies have been used for management of load on different services that has been used for cloud request response. These policies have been used and results have been obtained and discussed in next section.

- **User Base Configuration**
In this section user bases have been deployed in different regions of the world and these user bases have been configured for transmitting different request for transmission data to different datacenters.

User base configuration parameters
a) **Name**
This parameter define name of the user base that has been deployed in cloud simulation environment.
b) **Region**
This parameter define region of the user base that has been deployed in cloud simulation environment. The whole world map has been divided into six different regions number from 0 to 5.
c) **Request/user/hour**
In this parameter number of request have been defined by a particular use in an hour has been defined. On the basis of these request loads have been transmitted on cloud environment.
d) **Size/request/**
In this parameter size of the request has been defined. The size has been defined in bytes. Large number of bytes responds to transmit a heavy request on cloud.
e) **Peak hours**
This parameter defines peak hours that have been used user bases. Peak hours belong to time in which maximum users are using cloud services.
f) **Average peak and off-peak users**
In this parameter average numbers of user have been defined in peak hours and in off peak hours.

- **Data center Configuration**
In the processing of jobs of different datacenter various parameter for simulation has been defined at datacenters.
a) **Name**
This parameter define name of the datacenter that has been deployed in cloud simulation environment.
b) **Region**

This parameter define region of the datacenter that has been deployed in cloud simulation environment. The whole world map has been divided into six different regions number from 0 to 5.
c) **Architecture and OS**
In this parameter architecture and operating system used by datacenters have been defined. In this 32-bit architecture Linux based operating system has been used for simulation.
d) **VMM**
Virtual machine management has been used for defining virtual machine architecture in cloud computation. XEN open source virtual machine architecture has been used for data management. XEN divided CPU memory into different multi processorssystem that acts as virtual machines.
e) **Cost**
In the cost parameter of datacenters configuration various types of costs has been defined. VM cost, Memory cost, storage cost and data transmission cost has been defined for differentdatacenters that have been deployed.
f) **Physical H/W Units**
Number of physical units of H/w has been defined and these hardware units have been configured by providing id, memory, and storage, number of processors, processor speed and VM policy.

- **Internet characteristics Configuration**
a) **Transmission Delay between regions**
Transmission delay of data from one region to other region has been defined in the internet characteristics. On the basis of internet characteristics different metric has been defines to compute transmission delay region to another.

Table 4.1: Transmission Delay matrix in ms

R/R	0	1	2	3	4	5
0	25	100	150	200	250	100
1	100	25	250	500	350	200
2	150	250	25	150	150	200
3	250	500	150	25	500	500
4	250	350	150	500	25	500
5	100	200	200	500	500	25

b) **Bandwidth between Regions**
Bandwidth between different regions defines provided bandwidth for data transmission from one region to another using internet connection. Bandwidth has been defined in Mbps.

Table 4.2: Bandwidth matrix in Mbps

R/R	0	1	2	3	4	5
0	2000	1000	1000	1000	1000	1000
1	1000	800	1000	1000	1000	1000
2	1000	1000	2500	1000	1000	1000
3	1000	1000	1000	1500	1000	1000
4	1000	1000	1000	1000	500	1000
5	1000	1000	1000	1000	1000	500

4. Results

In the purposed work different simulations have been done for load balancing in cloud computing using different load balancing policies. In purposed work simulation model have

been developed by deploying various user bases and datacenters in different regions. These user bases have been allocated in different regions for transmitting requests to different users for processing their request. Datacenters have been used for responding user request by allocating virtual machines for different user requests.

Simulation configuration

In the purposed work six user bases and two datacenters have been defined in cloud computing environment. These user bases and datacenters have been allocated in different regions for processing. Region boundaries have been defined by different color regions.
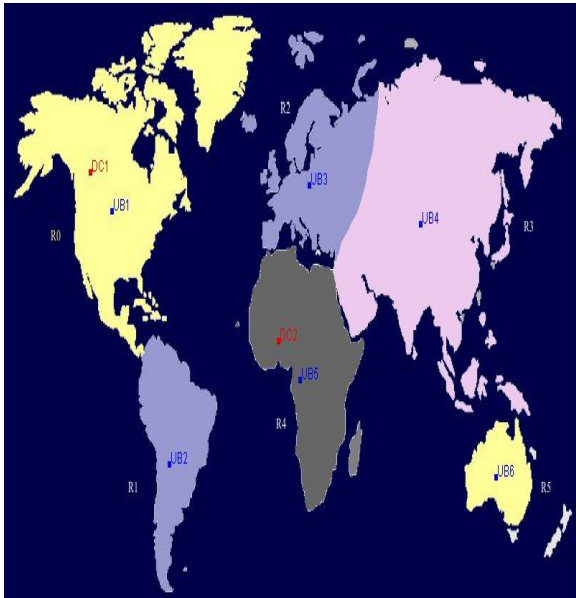

Figure 5.1: User base and DC allocation in Cloud Analyst

This figure represents location of different user bases and datacenters in different regions.

Table 5.1: User Base Configuration

Name	R	Request /Hr	Size	Peak Hr	Peak users	Normal users
UB1	0	600	100	1-3	1500	100
UB2	1	1600	1000	3-5	500	1000
UB3	2	2000	1000	5-7	500	1000
UB4	3	1000	1000	7-9	1600	600
UB5	4	1000	1000	9-11	2000	700
UB6	5	7000	300	11-12	2000	800

This table represents user base configuration that has been used for simulation of cloud computing environment.

Table 5.2: Datacenter Configuration

Name	R	Cost Vm/hr	Memory cost/s	Data transfer/ Gb	Speed	H/w Units
DC1	0	0.1	0.5	0.1	10000	2
DC2	4	0.1	0.8	0.1	10000	3

This table represents datacenter configuration that has been used for simulation of cloud computing environment.

Table 5.3: Other Simulation parameters

Parameter	Value
User Grouping Factor in User Base	1000
Request Grouping Factor	10
Executable instruction length/request	100
Load Balancing Policy	RR, Min-Min, RR + Min-Min
Simulation Duration	60.0 min
VM Image Size	10000
VM Memory	512
VM Bandwidth	1000
Data Center Architecture	X86
Data Center OS	Linux
Data Center VMM	Xen

On the basis of these simulation parameters various different load balancing policies have been used for simulation of cloud computing. On the basis of these simulations various parameters have been evaluated for performance evaluation of purposed approach.

Performance evaluation parameters

Various performance evaluation parameters have been computed for performance evaluation of purposed work. These parameters are used for performance evaluation of purposed approach.

1) **Over all Response Time**
   Over all response time has been measured for user base request completion and datacenter processing time. These parameters provide information about time taken by cloud service provider to respond requests.
2) **Data center Request servicing Time**
   Datacenter request servicing time provide information about a singledatacentersfor serving requests of different user bases.
3) **Cost analysis**
   Cost analysis parameters provide information about different load balancing approaches using data transmission cost and total VM cost occurred.

Table 5.4: Over all response time using Different policies

Parameter	RR	Min-Min	RR + Min-Min
ORT	278.46	272.72	271.91
DRPT	99.13	97.47	96.73

This table represents over all response time summary using different load balancing approaches.

Table 5.5: Datacenter request servicing time using Different policies

Data Centre	RR	Min-Min	RR + Min-Min
DC1	114.88	112.83	111.86
DC2	35.55	33.66	33.66

This table represents Datacenter request servicing time using different load balancing approaches

Table 5.6: Total Data transfer cost using Different policies

Cost	RR	Min-Min	RR + Min-Min
DC1	2958.10	2951.98	2945.23
DC2	1470.58	1450.29	1457.03

This table represents Datacenter request servicing time using different load balancing approaches

2. Conclusion& Future Scope

Cloud computing is used for responding different user requests. There are many problems in load management in service providing. In this paper load balancing in cloud computing has been donning using new purposed approach that uses optimization for allocation VM on the basis of fitness and number of requests allocating to virtual machine. In this process fitness of all VM's has been measured and on the basis of global fitness VM has been allocated to user base request for response. In the proposed approach round robin hybrid withMin-Min scheduling approach has been used for load balancing and by analyzing various performance evaluationparameters we can conclude that proposed approach provide better results in terms of overall response time and datacenter processing time.

References

[1] Amir Nahir"Distributed Oblivious Load Balancing Using Prioritized Job Replication", ISSN 978-3-901882-48-7, IEEE, 2012.

[2] Hong Tao "A dynamic data allocation method with improved load-balancing for cloud storage system", ISSN 978-1-84919-707-6, PP 220 – 225, IEEE, 2013

[3] Yuqi Zhang "Dynamic load-balanced multicast based on the Eucalyptus open-source cloud-computing system",ISSN 978-1-61284-158-8, pp. 456 – 460, IEEE, 2011.

[4] Magade, Krishnanjali A. "Techniques for load balancing in Wireless LAN's**",** ISSN 978-1-4799-3357-0, pp. 1831–1836, IEEE, 2014.

[5] Yean-Fu Wen "Load balancing job assignment for cluster-based cloud computing", ISSN 14517061, PP 199 – 204, IEEE, 2014.

[6] De Mello, M.O.M.C "Load balancing routing for path length and overhead controlling in Wireless Mesh Networks", ISSN 14630778, PP 1-6, IEEE, 2014.

[7] R. Angel Preethima, Margret Johnson*, "Survey on Optimization Techniques for Task Scheduling in Cloud Environment"*, IJARCSSE,Volume 3, Issue 12, December 2013.

[8] AkhilGoyal, Bharti, *"A Study of Load Balancing in Cloud Computing using Soft Computing Techniques "*,International Journal of Computer Applications (0975 – 8887) Volume 92 – No.9, April 2014

[9] NavjotKaur,Taranjit Singh Aulakh,Rajbir Singh Cheema, *"Comparison of Workflow Scheduling Algorithms in Cloud Computing"*,(IJACSA) International Journal of Advanced Computer Science and Applications, vol. 2, No. 10, 2011.

[10] Mayank Singh Rana, Sendhil Kumar , Jaisankar N, *"Comparison of Probabilistic Optimization Algorithms for Resource Scheduling in Cloud Computing Environment"* International Journal of Engineering and Technology (IJET)

[11] C.Kalpana,U.Karthick Kumar, R.Gogulan, *"Max-Min Particle Swarm Optimization Algorithm with Load Balancing for Distributed Task Scheduling on the Grid Environment"*, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 1, May 2012.

[12] ZHANG Yan-huaa, Feng Leia, Yang Zhia, *"Optimization of Cloud Database Route Scheduling Based on Combination of Genetic Algorithm and Ant Colony Algorithm"*, Science direct,Procedia Engineering 15 (2011), pp. 3341 – 3345.