ISSN No:-2456-2165

A Review of Cloud Computing Scheduling Algorithms

Vaibhav Sharma, Harish Chandra Sharma
School of CA & IT, SGRR University,
Dehradun, Uttarakhand, India

Sanjay kumar
Tula's Institute Dehradun,
Uttarakhand, India

Abstract:- Cloud Computing is basically an on demand computing done using the internet. The resources are shared with users using the internet to provide efficient computing. In Cloud Computing, the Load Balancing algorithms can be applied in various environments i.e. static, dynamic and centralized environment. The paper compares some of the load balancing strategies of cloud computing environment and summarizes the findings. The paper also reviews the different Load Balancing Algorithms in Cloud Computing with their advantages and disadvantages.

Keywords:- *Cloud Computing, Load Balancing, Static Load Balancing, Dynamic Load Balancing.*

I. INTRODUCTION

Cloud Computing [1] is an emerging and trending technology. It provides an ICT platform of various resources to the end-user, so that the user can allocate and reallocate the resources as per the requirements. It is the latest technology for resource sharing. Cloud Computing can deliver the computing services i.e. resources, hardware infrastructure, application etc. to end-users any time as Pay as you use system. It is the collection of resources such that hardware resources, software resources, storage and network services and devices that are collected together to provide powerful computing as a service.

There are different reasons for using cloud computing.
- Dynamic and Static Virtual Machine allocation
- No high end hardware and software is required
- No specific software required.
- Operating System not required
- Scalability
- Pay as you use
- Massive abstract infrastructure

Cloud Computing is an emerging technology trend that uses requirement bases hardware and software as its base. The hardware can be configured as desired and can be upgraded any time without affecting the behavior of cloud.

It uses a product –based software container system. E.g. a service that can be moved from one cloud provider to any other cloud provider without any effect on the service. It uses the concept of virtualization and abstraction layer for software and hardware. The some of the services provided by the Cloud Computing are:
- Software-as-a-Service (SaaS)
- Platform-as-a-Service (PaaS)
- Infrastructure-as-a-Service (IaaS)

These services determine the type of computing resources provided to the consumers.

II. LOAD BALANCING IN CLOUD COMPUTING ENVIRONMENT

Load Balancing in Cloud Computing is basically a process of distributing workloads and resources in such a way that tasks are allocated to all processors that helps in efficient utilization of all computing resources. [2]To achieve the simplified load balancing globally in cloud computing, we need some proper scheduling algorithms. Load Balancing distributes the load instruction to multiple nodes available across the network. The main responsibility of load balancing algorithm is to determine the next task in such a way that helps in reducing execution time, utilization of resources at the datacenters. There are many load balancing algorithms that are used in Cloud Computing for load balancing across the network.

Load Balancing Algorithm aims at:
- Improved network throughput.
- Provide reliable services.
- Minimize the system overheads.
- Maintains system stability and scalability
- Efficient resource utilization
- Efficient fault toleration system
- Highly Improved performance

Load Balancing in Cloud computing can be classified as:
- Static Load Balancing
- Dynamic Load Balancing

A. Static Load Balancing

Static Load Balancing [3] is highly dependent on the prior knowledge of the system. It is no way concern with the current state of the system. It needs the information about the data, processing power, attributes, computing resources, memory capacity, etc. in advance, so that it can be provide to algorithm to execute efficiently. The static load balancing algorithms are of non-preemptive types.

B. Dynamic Load Balancing

Unlike the static load balancing, The Dynamic load balancing is purely dependent on the current state of the system. It does not require any prior knowledge of the system.[4]

In this, jobs dynamically move to the under loaded nodes from the overloaded nodes, this movement of the jobs are decided at execution time, so the movements are not fixed previously. The classification of the dynamic load balancing algorithms can be done as centrally oriented &

ISSN No:-2456-2165

distributed spreading across the network. Dynamic load balancing can be classified as distributed and non-distributed algorithms.[5][6]

Cloud Sim simulator can be used for simulating the load balancing algorithm and measuring its effectiveness. [7]

III. ALGORITHMS FOR LOAD BALANCING

Load Balancing is one of the most important topic in the Cloud Computing. At any particular moment, a certain node is overloaded with maximum load then the load must be shared with other under loaded nodes to achieve the efficient resource utilization for the minimum completion time of specified tasks. Load Balancing helps to avoid the overloading and under loading of any nodes in the Cloud Computing environment.

It helps in efficient use of the computing resources, resource utilization, increased throughput, less waiting time, reduced power consumption, etc. There are many algorithms to achieve the load balancing in Cloud Computing, some of them are described below:

A. Round Robin Algorithm

Round Robin [8] is a simple algorithm that work as a circular queue. It allocates the task to the next VM in circular queue fashion irrespective of the load on the particular VM. At every particular time, when a task assigned to a VM, it moves to the end of the queue. In this algorithm, hardware resource allocates a time slice to execute a task on each node. It provides higher performance as compared to FCFS (First Come First Serve), but in some scenarios when time slice is huge then it behaves like FCFS.

B. Weighted Round Robin Algorithm

Weighted Round Robin algorithm is a modified form of Round Robin. Unlike the Round Robin, it considers the resource capacity of the VM and the VM with higher capacity get the maximum number of task. Servers can be allocated a weight, which represents the processing power of VM. [10].

C. Dynamic Round Robin Algorithm

The Dynamic Round Robin algorithm is designed to reduce the energy consumption in the system. It can be described by the following steps:

- If the physical machine has completed its execution then it will go into retiring state means it will not accept any more VM and the system will shut down after completing its execution.
- If the system stuck in to the retiring state for quite a long period of time then the VM can be shifted to a different physical machine.

D. Throttled Load Balancer

The Throttled Load Balancer is totally based on the request and allocation of the Virtual Machine. In this, client first request to the load balancer to check which VM can hold the particular load and perform task easily. It also shows that which VM is available and which VM is busy. This status is recorded in a Table called VM state table. If the request matches with the any VM state and capacity, it is accepted otherwise the request is send to queue. [10]

E. Central Load Balancer

It is the updated or modified form of Throttled Load Balancer. The central load balancer also maintains a table which contains the id and states of virtual machine. It balances the load among VM using various hardware resources. The preference of VM is also recorded along with the state of the VM. The preference usually calculated by the speed of the processor and the capacity of the memory [13]. For any request, the highest order VM is checked in the table, if is available then it is allocated for the request else the next order VM is checked. The process will continue untila VM is found for the particular request.

This algorithm can efficiently balance the load in any extreme heterogeneous conditions, however it has a drawback as well. All the requests are returned to central load balancer, which create overheads on the load balancer.

F. Active Monitoring Load Balancing (AMLB) Algorithm

AMLB is designed to minimize the response time in Cloud Computing environment. It maintains a table that consist of the data structure of virtual machine. For any request, it checks for the VM with minimal load or the VM in idle state and return the VM id. It is then the responsibility of Data Center Controller to assign the VM id to the particular request. Data Center then inform the load balancer about the allocated VM to the request.

G. VM Assigned Load Balancing Algorithm

It is similar to the Active Monitoring Load Balancing Algorithm (AMLB). It also maintains a table of VM. When any request arrives, the load balancer search throughout the VM table to find the least loaded VM to allocate for the request. If found, then it checks if this VM is selected in the last iteration. If no, then this VM is selected. Otherwise the next VM with least load is selected. The load balancer then returns the VM is to the Data Center.

H. Load Balancing Min-Min Algorithm

Min-Min is a simple algorithm used to provide the enhanced performance. It picks the task with minimal execution time. It helps in scheduling the task with least execution time and improve the overall produce period. In this, the task with minimum execution time are executed first and the task with larger execution time are executed at last. Thus, if the task with larger execution time are more than the task with low execution time, it gives the poor resource utilization and the larger task have to wait for the smaller task to finish first.[9]

I. Load Balancing Improved Min-Min(LBIMM)

The Min-Min [10] algorithm is improved to minimize the completion time of jobs, it called as Load balancing improved Min-Min algorithm (LBIMM). LBIMM executes same as Min-Min algorithm, but it picks the task with minimum execution time from the highly loaded resource and calculate the least completion time of the task on the other available resources. In this, if the total execution time generated by LBIMM is less than Min-Min than the task is

ISSN No:-2456-2165

submitted to the resource that produce it and time of the resources gets updated.

J. Max-Min Algorithm

Max-Min algorithm overcome the drawback of Min-Min algorithm where the task with larger execution time are greater than the task with minimum execution time. In Max-Min algorithm, the task with maximum execution task are selected first to execute and after the complete execution of larger task, the task with minimum execution time are executed.

K. User Priority Awarded Load Balance Improved Min-Min

User preference used by H. Chen et. al. [11] is included with LBIMM rule to develop PA-LBIMM. The algorithm works on two teams. All tasks are divided into two group G1 and G2, G1 is for high priority task group G2 is for ordinary task. The high priority tasks are executed first Min-Minalgorithm. The ordinary priority tasks are scheduled next. Finally load balancing algorithm produces makes pan of the executed task. [10]

L. Opportunistic Load Balancing (OLB)

Opportunistic Load Balancing (OLB) is a static/fixed load balancing algorithm, so it do not consider the current load as well as the prior information of the resources. OLB continuously tries to keep all the nodes busy in the system. OLB assign the tasks to the nodes randomly and doesn't consider the execution time of the task. This results in slow processing of the task.

M. Honey Bee Foraging

Honey Bee is a nature motivated decentralized load balancing algorithm which helps to achieve the load across the resources through local server action and enhanced the throughput. In this, the current load of a node is calculated then the states of the VM are set accordingly. The VM are then grouped according to their current load then the task from the higher load VM are assigned to the VM with minimal load. It reduces the response time of VM and the waiting time of particular task.

Honeybee Foraging servers are divided into virtual server having its own queue to maintain the server request and compute its processing time.

- Calculate profit for certain request. Profit can be set as needed. Generally calculating request takes the time of the waiting and CPU time.
- It uses response time as a parameter.
- The server stays only when the profit is high, or else proceeds forage by indicating that whether the state is loaded, overloaded or under loaded**.**

N. Exponential Smoothing Based Weighted Least Connection(ESBWLC)

ESBWLC algorithm uses the information from the previous data and historical data and makes efficient choice of all the information as classical knowledge and categorize the knowledge through smoothing issue to create a deeper impact through the new knowledge. It collects the information by using the advantages of electronic equipment, power, memory, variety of connections, etc.

O. Weighted Active Monitoring Load Balancing

Various VM have different efficiency and processing power, So the weights can be allocated to VMs by accessible processing power of VM. The next VM allocation can be done by choosing the powerful VM according the assigned weights of particular VM. The approach is projected by Harish Sharma et. al [3] which is unification of Weighted Round Robin and Active Monitoring Load Balancing algorithms. Thus the allocation of VM is done using not only the load but the processing capability of the VM. This approach overcome the drawback of Active Monitoring Load Balancing algorithm.

ISSN No:-2456-2165

S.no.	Algorithm	Description	Advantage	Disadvantage
1	Round Robin	Request allocation is done only for limited time period	Load is distributed equally	No prior information of the task. For a larger task context switching increases.
2	Weighted Round Robin	VM weight is assigned according to the processing capacity only	Optimal resource utilization	Processing time not taken into consideration, can lead to slow execution of tasks.
3	Dynamic Round Robin	Maintains retiring state of the VM	Reduced power consumption	Not idle for large data centers.
4	Throttled LB Algorithm	Maintain states of VM i.e. busy or idle	Load is distributed evenly across all nodes	Current state of the VM is not considered before allocation
5	Central Load Balancer	Maintains the states of all available VM	Centrally load balancing approach	Priority is fixed.
6	Active Monitoring Load Balancing	The request to allocated to the least loaded VM	Current load is considered before allocation	Does not consider the processing power of the particular VM
7	VM Assign LB Algorithm	VM is allocated as and when available	Proper utilization of VMs	NA
8	Weighted Active Monitoring LB Algorithm	All available VM's are assigned a weight according to their processing power.	Processing power and weights of the VM is considered before allocation	It increases the complexity
9	Min-Min Algorithm	Allocate tasks with least execution time first.	Simple to execute	Does not consider existing load and tasks with large execution time
10	Load Balancing Improved Min- Min LBIMM	From the all available task, the task with smallest completion time from the most densely loaded resource is calculated. If it is less than make span of Min-Min then the task is assigned again to the resource that produces it. The same process is repeated.	Overall completion time is reduced	Does not consider priority
11	Max Min Algorithm	Allocates task with higher execution time first.	Reduced make span with respect to Min-Min	Shorter jobs have to wait until all the tasks with higher execution time finishes.
12	User Priority Awarded Improved Min- Min	According to user priority, divides the task in two groups	Considers the priority and make span	No specific deadline
13	Opportunistic Load Balancing (OLB)	Uses static load balancing algorithm attempt to allocate selected job available in VM	Try to Keep all available VM busy	Previous load on VM is not considered
14	Honey Bee Foraging	Distributed load balancing for self-organization	Highly suited for heterogeneous environment	Resource increased but not efficiency
15	Weighted Least Connection	Assign work to the node having the minimum number of connection	Efficient Load Balancing	Does not consider the processing speed
16	Exponential Smoothing Forecast based WLC (ESBWLC)	Task are assigned according to the processing power and memory of the node.	Each VM is examined	Calculations are Complex

Table 1: Advantage and Disadvantages of Load Balancing Algorithm

ISSN No:-2456-2165

Algorithm	Environment	Challenges
Min-Min	Static	Starvation problem for the tasks with higher execution time
Max-Min	Static	Starvation problem for the tasks with least execution time
Round Robin	Static	Lack of proper resource utilization
OLB	Static	Slow Executions
Honey Bee	Dynamic	Takes more time to choose the appropriate VM
Active Clustering	Dynamic	Works on similar nodes, performance reduces with system diversity [18], power consumption
ESWLC	Dynamic	High complexity with more processing time

Table 2 : Challenges in existing load balancing algorithm

Algorithm	Static Load Balancing	Dynamic Load Balancing	Centralized Load Balancing	Distributed Load Balancing
Round Robin	**1**	**0**	**1**	**0**
Min-Min	**1**	**0**	**1**	**0**
Max-Min	**1**	**0**	**1**	**0**
CLB	**1**	**0**	**1**	**0**
LBMM	**0**	**1**	**0**	**0**
Active Clustering	**0**	**1**	**0**	**1**
OLB	**1**	**0**	**1**	**0**
PA-LBIMM	**1**	**0**	**1**	**0**
WLC	**0**	**1**	**1**	**0**
ESWLC	**0**	**1**	**1**	**0**
Honey Bee Foraging	**0**	**1**	**0**	**1**

Table 3: Load Balancing Environment for Different Algorithms

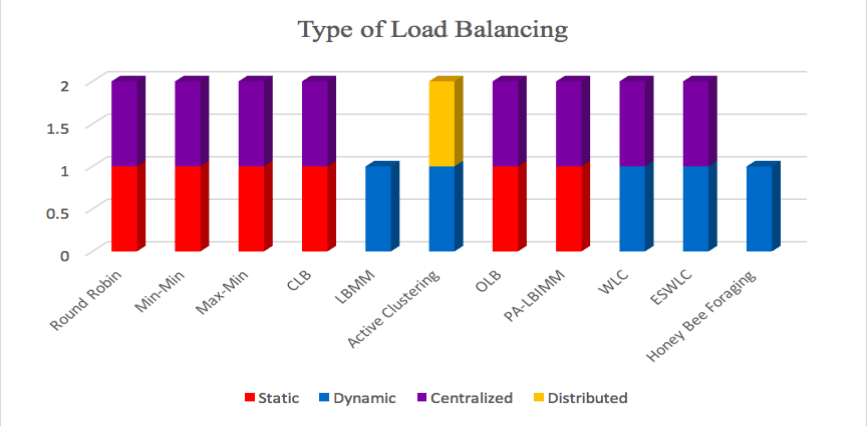

Fig. 1: Type of Load Balancing usage in different algorithms

IV. LOAD BALANCING APPROACH

Virtual Machine (VM) migration is a process of moving VM from one physical host to the other. It is also a method which is used for load balancing at the server. Dynamic movement of VM is the server can give more options for load balancing:

- Non Live VM Migration
- Partial Migration
- Local Area Live VM Movement
- Live Storage Migration
- Network Connection Migration

V. CONCLUSION

The load balancing in most important topic in cloud computing. It helps to improve and enhance the performance and throughput of the system. The paper collectively summarizes different load balancing algorithms of cloud computing along with their advantages and disadvantages. The above discussed load balancing algorithms focuses on increasing throughput, enhancing performance, reducing the response time and reducing the make span. The paper also summarizes the VM migration approach for load balancing in Cloud Computing environment.

It will remain a challenge for the future researchers to reduce the power consumption and effectively uses the bandwidth among the different data center in the cloud computing load balancing algorithms.

REFERENCES

[1.] Peter Mell, Timothy Grance. "The NIST Definition of Cloud Computing", Technical report. NIST, September 2011.
[2.] Harish Sharma, Pradeep Semwal, "A Review of Load Balancing Algorithms in Cloud Computing", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.9, Issue 3, pp.2786-2791, March 2021.
[3.] Harish C. Sharma, Meenakshi Bisht, "Best Fit Resource Allocation in Cloud Computing", International Journal of Computer Sciences and Engineering, E- ISSN : 2347-2693, Vol.7, Issue.3, pp.928-932, March, 2019
[4.] Gulshan Soni, Mala Kalra, "A Novel Approach for Load Balancing in Cloud Data Center", IEEE International Advance Computing Conference (IACC), Gurgaon, India, 2014, pp. 807-812.
[5.] Bhushan Ghutke, Urmila Shrawankar, "Pros and Cons of Load Balancing Algorithms for Cloud Computing", 2014 IEEE, International Conference on Information Systems and Computer Networks (ISCON), Mathura, 2014, pp. 123-127.
[6.] A. Kumar and M. Kalra, "Load balancing in cloud data center using modified active monitoring load balancer", 2016 International Conference on Advances in Computing, Communication, & Automation, ICACCA, Dehradun, 2016, pp.1-5
[7.] Boutaba, R., Zhang, Q., & Zhani, M. F. (n.d.). "Virtual Machine Migration in Cloud Computing Environments. Benefits, Challenges,and Approaches", IGI Global. 2014, pp. 383–408.
[8.] Harish C. Sharma, Himanshu Bahuguna, "A Survey of Load Balancing Algorithm in Cloud Computing", International Journal of Computer Engineering and Applications, issn: 2321-3469, Vol. XI, Issue XII, December 2017, pp. 150-160
[9.] Harish C. Sharma, et al, "Load Balancing in Cloud Computing Using a Novel Minimum Make span Algorithm", IJARCET, Vol. 5 issue 4, April 2016, ISSN :2278 – 1323 pp. 1160-1164
[10.] Harish C. Sharma, Pradeep Semwal, "Performance Comparison of Min-Min and Max-Min Scheduling Algorithm in Cloud Computing" NCRAST-2015, ISBN:978-93-85437-03-8, pp 379-385, Aug 22-23, 2015