An Analytical Review on Various Load Balancing Algorithms in Cloud Environment

Megha Dubey, Soumya Mukharji
O.P. Jindal University, Raigarh (C.G.)

Abstract- Cloud computing is rising as a new technology that delivers secure data storage center, unlimited computational resources and flexible data processing capabilities. Load balancing is one of the major issues in cloud computing which is balance the workload evenly across all the nodes in cloud data centers. The main aim of cloud data centers is how to distribute and service the requests comes from user bases efficiently and accurately. It helps to improve the performance of cloud system and make proper utilization of resources by fair allocation. This paper concluded several existing techniques for load balancing with their comparisons and performance analysis on the basis of different parameters.

Keywords- Cloud Computing, Load balancing, Existing Algorithms, Comparison, Performance Analysis.

I. INTRODUCTION

Cloud computing is a web-based approach where computing has been pre installed and exist as a service. All the applications and files are hosted on a cloud which consists of thousands of computers interlinked together in a complex way. The main idea behind cloud computing is to provide the ability to store, share and access the data, resources, and services over the internet without any hardware need. Cloud computing is the bundle of virtual resources and services through internet. As the number of users growing regularly in a cloud platform, load balancing has become major issues for cloud service provider. The goal of load balancing is to utilize the cloud resource in such way that throughput, efficiency and response time type of parameter should be increased. Load balancing balances the load evenly across all the nodes and execute the task in the shortest period of time. [2][3][4]

A. CLOUD SERVICES MODELS

- Software as a Service (SAAS): The application software that are provided by the providers. User's access, retrieve all the applications and services by web interfaces But they haven't the authority to control the internal function.
- Platform as a Service (PAAS): It provides platform and all the resources to the user for developing and deploying applications in cloud infrastructure. These applications and services are building by using programming languages, various library files and simulation tools provided by cloud service providers.
- Infrastructure as a Service (IAAS): It provides the services for accessing the internal resources such as processing, networking physical machines, virtual machines, virtual storage as per demand.

B. Characteristics of Cloud Computing[1]

- *Network access*: Cloud provides infinite possibility for the users available over the internet.
- *On demand service*: Users can access the services as per requirements.
- *Resource pooling*: Cloud providers develop various deployment models are distributed the resources among the users according to their demands of users.
- *Pay per use*: The cost of computing is based on the resources, data, and services used by customers.
- *Rapid elasticity*: Quality of service parameters (QoS) will not be affected as increase or decrease the quantity of user requirement.

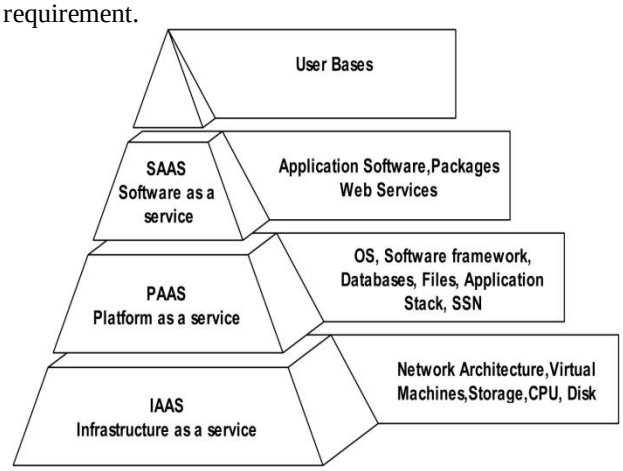

Fig.1: Cloud computing Models

a. LOAD BALANCING

Load balancing is a technique that balances the excess dynamic workload evenly across all the cloud data centers. Load balancing is used for attaining better service provisioning, resource utilization and improving the overall performance of the system in terms of throughput, efficiency, fault tolerance and response time. For the proper load distribution a load balancer is used to determine which virtual machine should assigns to which request group. [5]

C. Classifications of Load Balancing Algorithms

- *Static Load Balancing:* This algorithm does not depend on recent status of the system and current state of data. It uses preconception and previous knowledge about the task, requesting groups and the availability of resources.
- *Dynamic Load Balancing***:** Unlike Static approach, this algorithm does not require previous knowledge about the system. In this algorithm the workloads are allocating in run time according to the current status and availability of the resources. These algorithms are considered as

complex, but have better fault tolerance and overall performance over static algorithms. [5][7]

a. LOAD BALANCING ALGORITHMS

In this paper following pre existing load balancing techniques are analyzed and compared on the basis of load balancing parameters.

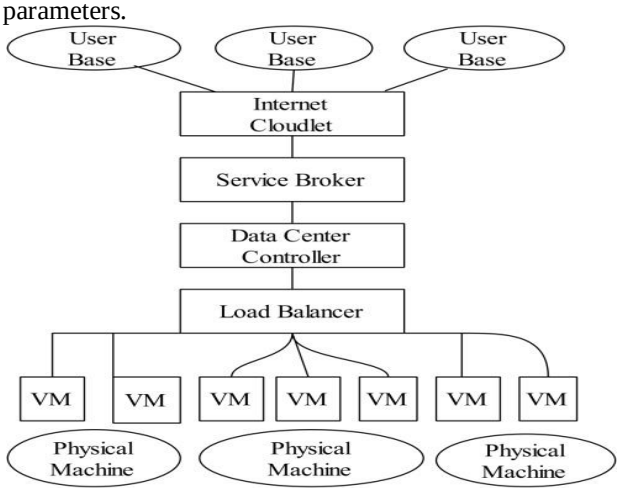

Fig.2: Load Balancing in cloud computing

A. Round-Robin Algorithm [8]

In this algorithm, the assigned jobs to each resource according to time slices (also known as time quantum) in circular order without any priority consideration. This algorithm is a type of static algorithm which uses the round robin fashion for allocating job. It selects the first task randomly then equally distributes all those tasks to multiple nodes according to their time slices. Because of the different processing time of each job when it will assign to resources some nodes get overloaded and some nodes get under loaded at many points in time.

B. Min-Min & Max-Min Algorithm [10]

In both these algorithms minimum execution time and minimum completion time is calculated for all the requesting tasks.

- Minimum Execution Time: It allocates each task to those resources which having least execution time without considering that those resources are available or busy.
- Minimum Completion Time: It allocates the tasks to those resources which finished them in least time without considering their minimum execution time

In Min-Min algorithm initially all the unassigned tasks in waiting queue. Expected minimum completion time is calculated for the entire task and the completion time for all the tasks is calculated on all the resources. The task with the minimum predicted completion time from waiting queue is taken and it will assign to the similar resource. When task is executed this process is repeatedly done until all tasks are completed.

In Max-Min algorithm, the maximum expected completion time for entire task is measured for all the resources. The task with the maximum predicted completion time is chosen and it will allocate to the similar resource. When the task which is completed that is removed from the queue and the process is constantly run until all tasks are executed.

C. Active Monitoring Algorithm [8]

This algorithm balances the tasks among available virtual machines to even out the number of active tasks on each virtual machine at any given time. In this algorithm active load balancers maintain all the details about the VM whenever any request comes to data center controller it will parses the table for available VM in active load balancer. Active load balancer returns the VM Id then data center controller distributes equal amount of workloads on all the available virtual machines.

D. Throttled Load Balancing Algorithm [8]

In this algorithm data center controller explores a available virtual machine for allocating the specific request. In this algorithm throttled load balancer maintains a list of Busy/Available VM's. Whenever a request arrives it parses the list of VM when the first available VM found the load balancer returns the VM Id and ensures that only a pre-defined number of request groups are assigned to a single VM. If more request groups are present than the number of available VM's, then the requests groups will have to be queued until the data center controller finds next available VM.

E. Ant Colony Optimization Algorithm [5]

This algorithm is used to search for an optimal solution for efficient distribution of work load among all the nodes. In this algorithm the pheromone-based behavior of ants foraging is used. The foraging behavior of ants is based on finding the shortest path between source of food and their colony. When the new request initializes ant starts moving from their head node to the direction of food node during the foraging process, ant's keeps track their records by leaving their pheromones on the path so that the other ants of the colony find their path for food. If the selected path is the shortest then it will be confirmed otherwise the pheromones evaporate and ants searches for new optimal path.

F. Honeybee Foraging Algorithm [3]

This algorithm adequately distributes the work load among all virtual machines and increases the throughput. The main aim of this algorithm is to assign the priority based tasks to the virtual machines in such a way that they have minimum amount of waiting time in queue. The current work load of VM is calculated and decides the state of VM whether it is over loaded, under loaded or balanced. According to tasks priority the work load from Over loaded VM is migrated to under loaded VM. This load balancing algorithm is based on the honey bee foraging strategy whenever the tasks are removed from VM's are treated as honey bees.

G. Active Clustering Algorithm [6]

In this algorithm collecting similar types of nodes together based on the concept of match maker node. It forms a cluster of similar tasks and similar resources and services for balances

the load of system. Matchmaker forms a connection between its adjacent nodes it's like as the initial node. Then the matchmaker node disconnects the connection with initial node. This process is repeating until the work load balances. This algorithm increases the throughput and performance of the system because of the high availability of resources and efficient resource utilization.

H. Biased Random Sampling Algorithm [1]

In this algorithm the distribution of loads in each nodes and the connectivity between the nodes and request groups are represented by virtual graph. Each node is taken as a vertex of the graph and each in degree represents the available resources hold by particular nodes. When a request comes load balancer allocates the resources of nodes on the basis of their in degree edges. When these resources are assigned the in degree of particular node is decrement by one and it will automatically increase when request executed. Random sampling used the threshold value, which indicates the maximum traversal from one node to another node. The length of traversal is known as walk length. The adjacent node of the current node is selected for the traversal. When a request comes, load balancer arbitrarily selects a node and compares the current walk length with the threshold value. If the current walk length is equals to or greater than the threshold value, the job is executed at that node. Otherwise, the walk length of the request is incremented and another node is picked randomly.

I. Genetic Algorithm (GA) [9]

This Algorithm is based on three operations: selection, genetic operation, and replacement. GA works on selection of individuals from population and apply cross over on best fittest pair of individuals where fitness will calculate by fitness function. Now mutate the new offspring with a mutation probability and apply replacement operation to incorporate new individual into population. This algorithm used to balance the load by trying to minimize the make span of a given tasks set. Advantage of this algorithm is that it can handle a vast search space, applicable to complex objective function and can avoid being trapping into local optimal solution.

II. CONCLUSION

Load Balancing is major issue in cloud computing environment to achieve maximum utilization of resources. In this paper, we have discussed the different types of load balancing algorithms and their performance analysis based on different types of parameters like throughput, response time, migration time, fault tolerance, resource utilization etc. In Table1 we also described benefits and drawbacks for these algorithms in different conditions. By comparing and analyzing the existing algorithms we have find that there is many scope for improving the performance of algorithms in terms of priority, cost and more dynamic in nature. Future work is related to designing a new dynamic load balancing algorithm along with more Quality of Service parameters in cloud computing environment.

III. REFERENCES

[1]. Dharmesh Kashyap, Jaydeep Viradiya, "A Survey Of Various Load Balancing Algorithms In Cloud Computing, International Journal Of Scientific & Technology Research", Volume 3, Issue 11, 2014.

[2]. Sukrati Jain, Dr. Ashendra K. Saxena, "A Survey of Load Balancing Challenges in Cloud Environment", 5th International Conference on System Modeling & Advancement in Research Trends, IEEE 2016.

[3]. Dhinesh Babu L.D., P. Venkata Krishna, "Honey bee behavior inspired load balancing of tasks in cloud computing environments", Applied Soft Computing Science Direct 13 (2013) 2292–2303.

[4]. Mohit Kumar, S.C.Sharma, "Dynamic load balancing algorithm for balancing the workload among virtual machine in cloud computing", 7th International Conference on Advances in Computing & Communications Science Direct 2017.

[5]. Sushil Kumar, Deepak Singh Rana, (2015) Various Dynamic Load Balancing Algorithms in Cloud Environment: A Survey, International Journal of Computer Applications (0975 – 8887) Volume 129 – No.6.

[6]. Rajwinder Kaur and Pawan Luthra, "Load Balancing in Cloud Computing", Proc. of Int. Conf. on Recent Trends in Information, Telecommunication and Computing 2014.

[7]. Abhijit Patil, Harshal Gala Jai Kapoor "Dynamic Load Balancing in Cloud Computing using Swarm Intelligence Algorithms" International Journal of Computer Applications (0975 – 8887) Volume 130 – No.15,2015.

[8]. B. Wickremasinghe, "Cloud Analyst: A CloudSim-based Tool for Modeling and Analysis of Large Scale Cloud Computing Environments", MEDC Distributed Computing Project, CSSE Dept., University Of Melbourne 2009.

[9]. Kousik Dasguptaa, Brototi Mandalb, Paramartha Duttac, Jyotsna Kumar Mondald, Santanu Dame, "A Genetic Algorithm (GA) based Load Balancing Strategy for Cloud Computing", International Conference on Computational Intelligence: Modeling Techniques and Applications, Science Direct 2013.

[10]. Rajwinder Kaur, Pawan Luthra, "Load Balancing in Cloud System using Max-Min and Min-Min Algorithm", International Journal of Computer Applications (0975 – 8887) National Conference on Emerging Trends in Computer Technology, 2014.

[11]. Sidra Aslam, Munam Ali Shah , "Load Balancing Algorithms in Cloud Computing: A Survey of Modern Techniques", National Software Engineering Conference (NSEC 2015) IEEE.

[12]. Ms. Shalini Joshi, Dr. Uma Kumari, "Load Balancing in Cloud Computing: Challenges & Issues", IEEE 2016.

b. COMPARISON [11] [12]

Algorithm	Benefits	Drawbacks
Round-Robin	Used fixed time quantum. Easy to understand Performs better for short CPU burst.	Larger tasks take long time. Can occur more context switches due to short quantum time
Min-Min and Max-Min	Used minimum execution and completion time value. Requirements are prior known.	Starvation occurs and tasks variation can't be predicted. It takes long time to complete the task.
Active Monitoring	load balancing attempts to maintain equal work loads on all the available VMs	Poor performance in case of fault tolerance.
Throttled	Good performance; Virtual machine state list is used to manage the tasks.	Tasks need to be waited when resources are busy.
Ant Colony Optimization	Minimizes make span.; Computationally intensive It is adaptive to dynamic environments.	Network is over headed so search takes long time. No clarity about the number of ants.
Honeybee	Increases throughput; Minimize response time	High priority tasks can't work without VM machine.
Active Clustering	Performs well with high utilized resources. Increases the throughput.	The performance is poor when there is an increase in variety of nodes
Biased Random Sampling	It is performing excellent when similar resources provided.	Corrupts when load increases.
Genetic Algorithm	Load balancing is based on three operations: selection, genetic operation, and replacement	It is not used distributed balancing approach.

Fig2 -Performance of various load balancing algorithms

Parameters / Algorithms	Throughput	Response Time	Overhead	Fault tolerance	Migration time	Resource Utilization	Scalability	Performance
Round Robin	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Active Monitoring	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Min-Min	Yes	Yes	Yes	No	No	Yes	No	Yes
Max-min	Yes	Yes	Yes	No	No	Yes	No	Yes
Throttled	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Active clustering	No	No	Yes	No	Yes	Yes	No	No
Biased Random Sampling	No	No	Yes	No	No	Yes	No	Yes
HoneyBee	No	No	No	No	No	Yes	No	No
Genetic Algorithm	No	No	No	No	No	Yes	No	Yes
Ant Colony Optimization	Yes	No	No	No	No	Yes	Yes	Yes