International Journal of Computer science engineering Techniques-– Volume 2 Issue 3, Mar - Apr 2017

RESEARCH ARTICLE OPEN ACCESS

# A Study of Load Balancing Algorithm on Cloud Computing

[1]Dr. Vijay S. Gulhane

[1](Information Technology, Sipna Collage of Engineering. & Technology, Amravati)

## Abstract:

Nowadays, Cloud computing has become buzzword in the Information Technology and is a next stage in the evolution of Internet, It provides very large amount of computing and storage services to users through the internet. Cloud computing employs a variety of computing resources to facilitate the execution of large-scale tasks. Load balancing is a method which allocates workload across different nodes to ensure that none of the node is overwhelmed or is lacking resources. Load balancing algorithms are chiefly classified under static and dynamic algorithms there are various algorithms under these two classifications like Honey Bee Foraging Algorithm, Throttled load Balancing Algorithm, Ant Colony Optimization Algorithm etc. this paper will give the study of all the load balancing algorithms which are used in cloud computing

*Keywords* — **Cloud Computing, load Balancing, Workload**

## I. INTRODUCTION

The cloud computing is a distributed internet based paradigm, designed for remote sharing and usage of different resources and services like storage, computational capabilities and applications etc. with high reliability over the large networks. This services are mainly divided into three categories like infrastructure as a service (IaaS), software as a service (SaaS), platform as a service (PaaS). Cloud computing facility can deliver services in form of software (e.g. email, web browser), platform (e.g. development tools) and infrastructure (e.g. storage space) it is a service oriented application many firms are relying on cloud computing paradigm to cater to the needs of users it has its own share of benefits and challenges and has tremendous scope for Future[1]. Load Balancing is a methodology to distribute the workload across multiple computers or other resources over the network links to achieve optimal resource utilization, minimum data processing time and to avoid overload

.

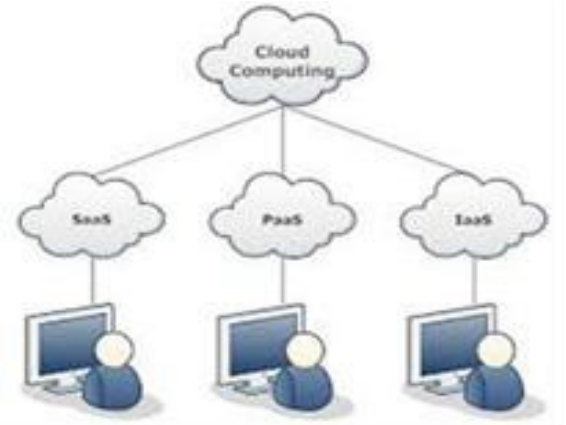


Fig. 1 Cloud Computing Model

Load balancing ensures that all the processors in the system as well as in the network do approximately the equal amount of work at any instant of time. Load balancing serves two important needs, primarily to promote availability of cloud resources and secondarily to promote performance. Load Balancing is a methodology to distribute the workload across multiple computers or other resources over the network links to achieve optimal resource utilization, minimum data processing time and to avoid overload. Load balancing ensures that all the processors in the system as well as in the network do approximately

the equal amount of work at any instant of time. Load balancing serves two important needs, primarily to promote availability of cloud resources and secondarily to promote performance. Following are the important parameters for load balancing algorithms[11].

**Goals of Load Balancing**

Following parameters are available in literature for measuring efficiency of a load balancing algorithm[2]

i.*Reliability*: The algorithm must be reliable, since process failure while transferring job from one location to other may lead to increased waiting time and customer dissatisfaction.

ii.*Adaptability:* Algorithm must be capable of adapting the dynamically changing user requests and provide task allocation in minimal amount of time.

iii.*Fault Tolerance:* The algorithm must ensure fault tolerance, so that in case of a problem in the system complete load balancing mechanism does not stop working.

iv.*Throughput:* the algorithm must ensure increased throughput at minimal expense. If a load balancing algorithm doesn't increase system throughput, it defeats its own purpose.

v.*Waiting Time:* Algorithm should minimize wait time of a task for allocation of resources to it.

Next subsection elaborates major components of a dynamic load balancing algorithm.

## I. Classification of Load Balancing

### *A. Static algorithms:*

The execution of these algorithms do not take into account the current state of the system hence these algorithms don't depend on the current state which the system is in, static algorithms have prior knowledge related to system resources and details of all tasks. These algorithms allocate workload among processors before execution of algorithm Depending on their performance such as arrival time execution time, amount of resources needed, the workload is distributed in the start by master processor the slave processor calculate its allocated work and The execution of these algorithms do not take into account the current state of the system hence these algorithms don't depend on the current state which the system is in, static algorithms have prior knowledge related to system resources and details of all tasks. These algorithms allocate workload among processors before execution of algorithm Depending on their performance such as arrival time execution time, amount of resources needed, the workload is distributed in the start by master processor the slave processor calculate its allocated work and submits the result to the master the goal of SLB method is to reduce the overall execution time of a concurrent program and minimizing the communication delays [3].

### *B. Dynamic Load Balancing (DLB) Algorithms*

In dynamic load balancing algorithms work load is distributed among the processors at runtime. The master assigns new processes to the slaves based on the new information collected [4, 5]. Unlike static algorithms, dynamic algorithms allocate processes dynamically when one of the processors becomes under loaded. Instead, they are buffered in the queue on the main host and allocated dynamically upon requests from remote hosts. Dynamic load balancers continually monitor the load on all the processors, and when the load imbalance reaches some predefined level, the redistribution of work takes place. But as this monitoring steals CPU cycles so care must taken as when it should be invoked. This redistribution does incur extra overhead at execution time.

### *C. Sender initiated algorithm:*

This type of load balancing algorithm is initiated by the sender here the sender transmits the request messages until it gets a receiver that can accept the workload

### *D. Receiver initiated algorithm:*

This type of load balancing algorithm is initiated by the receiver here the receiver transmits the request messages till it finds a sender that can accept the workload

### *E. Symmetric algorithm:*

This algorithm is a combination of sender initiated and receiver initiated algorithm

*F. **Centralized approach**:*

In centralized approach the tasks and workload is passed from a centralized location to different processes there is a master slave relation between the centralized location and processes

*G. **Decentralized approach**:*

In this approach the workload is passed to arbitrary processes

## II. Review of Load Balancing Algorithm

Several load balancing have been proposed. Some of them are discussed here as follows.

### A. *Dynamic energy aware capacity provisioning: -*

The algorithm proposed in paper [6] is based on saving energy in data centers by dynamically adjusting data center capacity by turning off unused machines or to set them to a power saving state. This algorithm uses a model predictive control (MPC) which minimizes the total energy cost while meeting performance objective in terms of task scheduling delay. Parameters used in this algorithm are Resource utilization, task scheduling delay (SLA cost), machine reconfiguration cost, and electricity price. It also predicts the future usage of resources in system i.e. CPU and memory. ARIMA model is used for prediction. Controller is responsible for reducing total operational cost of system. Which is the sum of SLA cost and energy cost. A high task scheduling delay affects the performance of some tasks. MPC algorithm adjusts the no. of servers/VM's to track the optimality condition while considering switching cost of machines. Bottleneck resource (a resource having high utilization) plays important role in this. Capacity provisioning module decides which machine to be added or removed based on certain criterion like usage of machine and its location. This algorithm gives the better results in saving energy with high performance and average resource utilization. However this algorithm considers all machines to be homogeneous with identical resource capacities

.

### B. *Virtual machine placement algorithm: -*

In [7], Virtual machine (VM) placement is the process of selecting the most suitable server in large cloud data centers to deploy newly-created VMs. Several approaches have been proposed to find a solution to this problem. Existing solutions only consider a limited number of resource types, thus resulting in unbalanced load or in the unnecessary conditions. Here, we propose an algorithm, called MaxBRU that improves the utilization of resources and maintains the usage of resources across multiple dimensions. This algorithm uses multiple resource-constraint metrics that help to find the most suitable server for deploying VMs in large cloud data centers. Parameters used in this algorithm are Virtual machines, resource utilization ratio (RU), resource balance ratio (RB). Resource may be CPU capacity, Memory and Network bandwidth. The main advantage of Max- BRU algorithm is that First it increases the resource utilization by minimizing the amount of physical servers used. Second, it effectively uses the multiple type of resources.

### C. *Round robin load balancing: -*

In [8], the algorithm is proposed called Round robin, which uses the time slicing mechanism. As name implies this algorithm works in the round manner in which each node has given a time slice and has to wait for their turn. The time is divided into interval is allotted to each node in which nodes have to perform their task. It uses random selection procedure where first node is selected randomly and jobs are allocated to other nodes in a round robin fashion. The key point of this algorithm is that it yields no starvation and gives a faster response in case of equal workload distribution among processes. But as different processes have different processing times, therefore at any time some nodes may be heavily loaded while others remain idle and underutilized.

### D. *Central load balancing decision model:*

In [9] proposed algorithm is called as CLBDM. It is based on the human administrator point of view. It is a combination of the Round Robin Algorithm

and session switching at the application layer. Round Robin is a used for the load balancing. In round robin algorithm, it assigns the task to the node with the least number of loads. In this algorithm a threshold time is considered. If connection time between the client and the node in the cloud is greater than the threshold time then there is an issue. If an issue is found, then the connection between client and node will be terminated and the task will be moved to another node using the regular Round Robin rules. The key point of this algorithm is that it is superior to Round Robin Algorithm as automated tasks forwarding reduces the need for a human administrator. But there are some drawbacks of this algorithm are that if this algorithm fails to work properly then whole process/system fails.

### *E. Min- Min load balancing: -*

In the proposed algorithm known as Min- Min [10], initially there is a task set which is not assigned to any of node. For all the available nodes the minimum completion time is calculated. On finding the minimum time the task having the completion time minimum is chosen and assigned to the respective node. The execution time of all other tasks available in that machine is updated and the task gets discarded from the available task set. Once all the tasks have been assigned to proper machine this process is repeated. The advantages of this algorithm are that it is a simple and fast algorithm yields improved

## III. CONCLUSIONS

The Cloud computing has revolutionized the way resources and services are availed by users over the Internet but it has its challenges efficient load balancing is one of the key issues concerning any cloud service provider as an even distribution of workload across different nodes is a pivotal requirement for high resource utilization and user satisfaction there are different classifications of load balancing algorithms each algorithm gives optimum results in a particular circumstance and scenario, depending on objectives of the cloud environment and given resources an algorithm is selected. The performance of the load balancing algorithms is evaluated by different parameters like throughput, response time, fault tolerance, scalability etc. load balancing is the requirement of a cloud environment and how well this requirement is met depends on the algorithm chosen.

## REFERENCES


[1] Pragati Priyadarshinee, Pragya Jain ―Load Balancing and Parallelism in Cloud Computing‖ International Journal of Engineering and Advanced Technology (IJEAT) Volume-1, Issue-5, June 2012

[2] Aarti Singha, Dimple Junejab, Manisha Malhotra, Autonomous Agent Based Load Balancing Algorithm in Cloud Computing (ICACTA-2015) 45 ( 2015 ) 832 – 841.

[3] Rajesh George Rajan, V.Jeyakrishnan ―A Survey on Load Balancing in Cloud Computing Environments‖ International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 12, December 2013.

[4] Malik, S., "Dynamic Load Balancing in a Network of Workstation", 95.515 Research Report, 19th November, 2000.

[5] Wang Y. and Morris, R., "Load balancing in distributed systems", IEEE Trans. Computing, C-34, 3, 204-217, March, 1985.

[6] Qi Zhang , Shuo Zhang, Raouf Boutaba and Mohmed Faten Zhani, "Dynamic energy capacity provisioning for cloud computing environments" In Proceedings of the International Conference on Autonomic Computing (ICAC), 2012.

[7] Nguyen Trung Hieu, Mario Di Francesco, and Antti Yl¨a-Jaaski, "A Virtual Machine Placement Algorithm for Balanced Resource Utilization in Cloud Data Centers", International Conference on Cloud Computing,2014.

[8] Nusrat Pasha, Dr. Amit Agarwal and Dr.Ravi Rastogi,"Round Robin Approach for VM Load Balancing Algorithm in Cloud Computing

Environment" International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, May 2014.

[9] Radojevic, B. and M. Zagar, "Analysis of issues with load balancing algorithms in hosted (cloud) environments." In proc.34th International Convention on MIPRO, IEEE, 2011

[10] T. Kokilavani and Dr. D.I. George Amalarethinam, "Load Balanced Min-Min Algorithm for Static Meta-Task Scheduling in Grid Computing"International Journal of Computer Applications Volume 20– No.2, pp.0975- 8887, April 2011..

[11] Amritpal Singh" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 7, July 2015"