ISSN No:-2456-2165

An Efficient Load Balancing Technique in Cloud Computing

Kunal Sharma
Department of Computer Engineering
Bharati Vidyapeeth College of Engineering (Deemed to Be University)
Pune, India

Dinesh Thakur
Department of Computer Engineering
Bharati Vidyapeeth College of Engineering (Deemed to Be University)
Pune, India

Anant Sharma
Department of Computer Engineering
Bharati Vidyapeeth College of Engineering (Deemed to Be University)
Pune, India

Rohini Khalkar
Department of Computer Engineering
Bharati Vidyapeeth College of Engineering (Deemed to Be University)
Pune, India

Abstract:- **Load balancing and Resource allocation are very important in cloud computing. Any process related to the computation is done with help of the legit resources which is mandatory to follow up. To use the resources in efficient manner an optimized balancing and scheduling algorithm is used to get on with cloud scheduling problems. By deploying or adding VM's in proper locations to get better speed of locating better allocation method with in turn increase optimized utilization of resources available. Scheduling the resources to get over assignment problem that are encountered at the time of processing. Due to the sudden increase in the work load over the internet there is a need of some Load Balancing Algorithm to counter that load which can cause severe breakdowns and downtime in the working of any organization. In this paper, we are presenting an efficient method to counter load balancing which is Round Robin Algorithm.**

Keywords:- *Cloud computing, Load balancing, Round Robin.*

I. INTRODUCTION

With the help of cloud computing one can provide the huge amount of potential to provide the power of computation to everyone at reduced cost of expenditure. To get in touch with the other competition in the market one must reduce inefficiencies and increase productivity. In manufacturing, productivity is directly linked to how well you can optimize the resources. The hardest call is to select the best way to optimize the performance in a computational process. Actually, cloud is used to give on-demand capacity as a utility, but it can differ among the different cloud providers but the best method is to provision virtualized resources as a service(VRaaS). Cloud computing is basically outsourcing computer program which in turn helps the user to access the resources from anywhere they want to with help of the internet that are being uploaded on the cloud it helps users to not to worry about the memory and processing power and can directly get the end results.

II. THEORY

A. *Load Balancing*

Due to the increase in web traffic and different apps over the web the data is also increasing day by day and so millions of data is created every second. Load balancing has become a very essential part in research parts due to need of balancing the load of this heavy traffic. Cloud computing is basic concept of that uses virtual machines instead of physical machines to host. In the method of Load balancing we divide the whole load with various nodes in distributive system. If any failure occurs in any node it leads to the isolation of web and hinders the organization in this case load balancing helps in providing availability and scalability. Main definition of cloud computing is that it is a set of distributed computers which further consist of parallel computing nodes which gives on-demand resources or services over a network. Load balancing is the process of redistributing the incoming traffic load to the individual servers of server pool which helps in increasing the system efficiency. For creating the load balancing mechanism many strategies and techniques are used. In this load should be distributed over all the present resources so that each resource does almost the same amount of work at any point of time. Cloud vendors use load balancing services, which allow its users to adjust amount of processors for their surplus demands. That proves load balancing provides two purposes, primarily to increase availability of cloud assets and secondly increase the performance.

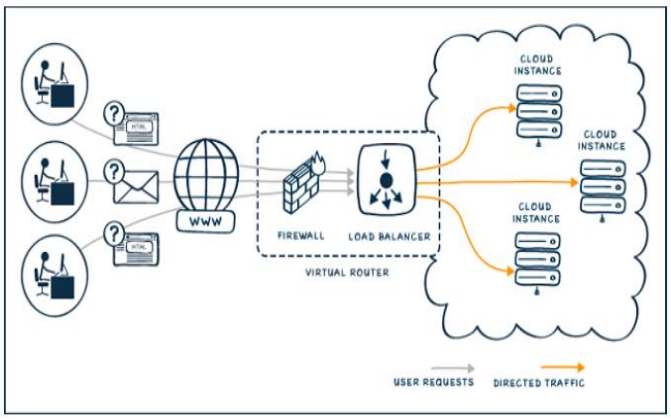

Fig 1:- Load Balancing

B. Goals of load balancing algorithms

- *Cost effectiveness:* It is method to improve the performance of our system by reducing the resource cost.

- *Scalability and flexibility:* The cloud computing system in which the algorithm is deployed can vary in topology or size. Thus, the algorithm should be scalable and adjustable such that changes can be managed easily.

- *Priority:* It means we can set different priority of the resources or jobs we need beforehand through the algorithm itself for better management of the crucial jobs in spite of equal time provision for rest of the jobs irrespective of their origin.

C. Load balancing classification

Load balancing algorithm can be further categorized into 2 different sub categories

- *Static approach:* - In this approach traffic load is divided equivalently between all the active severs.
- *Dynamic approach:* - In this approach consider only the present state of the system during resource allocation. Dynamic approach is optimal for large or hugely distributed networks such as parallel & distributive computing.

Dynamic load balancing further has 2 different categories:

- *Centralized approach:* - In this method, only a specific single node is responsible for managing and distribution of entire system.
- *Distributed approach:* - In this approach, every node has to build load vector on its own without any help. Vector collects the information regarding load of other nodes. All future decisions are done using these local vectors.

D. Metrics for Load Balancing:

- *Throughput:* It means how many data amount of data or the entities are passing or are in a process at the moment. Throughput should be always high for good performance.
- *Fault Tolerance:* As the name suggests it should be able to tolerate the fault.
- *Migration time:* It is the time taken to transfer from one process to another in a system. If the migration time is less the system performance is more.
- *Response Time:* It can be described as the time consumed by the resource allocation algorithm to response back to the task. If response time is less than the algorithm is more efficient.
- *Scalability:* It is the capability of an algorithm to execute Resource Allocation for any number (should be finite) of nodes of the system.

E. Policies of Load Balancing algorithm

Policies active in load balancing which are as follows:

- *Information policy:* It defines that which information is needed and what are the steps used to get that information.
- *Triggering policy:* It defines the exact time of execution when the resource allocation task is starting to work on load.
- *Resource type policy:* It defines all the resources that are available to perform load balancing during the process.
- *Location policy:* It make use of all the end results of the resource policy and after looking at the results from the policy it finds out the exact pair or the best counter-part for a particular server or the receiver.
- *Selection policy:* It uses the results and finds out the task which is the task which has to be transferred from overloaded node to the vacant node.

III. PROPOSED SYSTEM

In this design we present the implementation of the automatic load balancing and resource management system that gets a acceptable balance between avoidance of overload and green computing. In this model we have created VM's (Virtual Machines) inside the PM's so as to support green computing to cope up with the overload of traffic on one particular PM's which can degrade the performance of its VM's. We have developed and implemented a load balancing algorithm that can avoid overload in systems effectively.

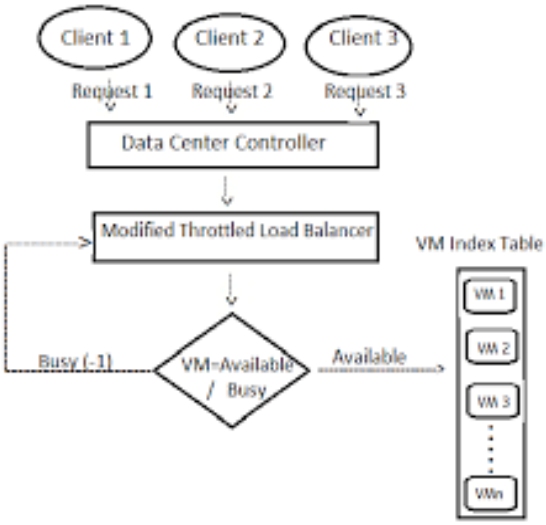

Fig 2:- Flow Chart

- *Round Robin:* This is a load balancing algorithm in this algorithm all the processes are divided in between all the processes. In this algorithm each process is assigned to the processor in a round robin order. In this traffic load is distributed in between all the available processors are equal. In this algorithm different processes may not have same job processing at the same time. At most of the time some nodes may be fully loaded and others remain free. In Round Robin Scheduling the time quantum plays a vital role. When the value of the time quantum is very high then Round Robin Scheduling Algorithm is exactly the same as the first come first serve Scheduling and time quantum is very less then RR Scheduling is known as Processor Sharing Algorithm.

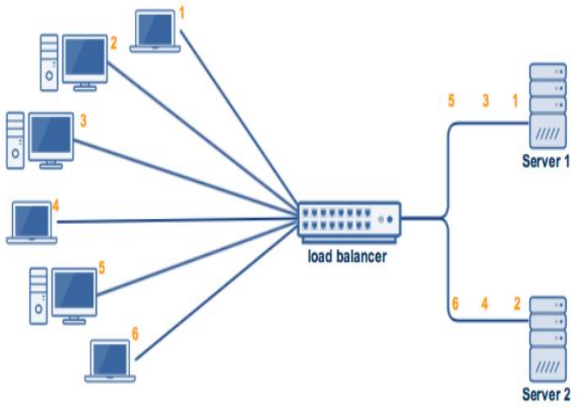

Fig 3:- Round Robin

IV. CONCLUSION

With the use of this load balancing algorithm to address scheduling of available resources problem in cloud services with the help of this one can achieve the proper optimization and sometimes further more sub-optimization in cloud scheduling problems. By the help of this model for resource scheduling it will increase the desired performance with the increase in efficiency in resource allocation. Representation of the scheduling problem as unbalanced assignment problem with the help of server node concept and the addition of the legit scheduler to the idle nodes to assign the incoming traffic of a specific server to perform the task and with the help of this rate of utilization of resources also increases with the better precision.

V. FUTURE WORK

This can be implemented and deployed with the dynamic load balancing concept on cloud environment. The concept of Big Data in cloud is also a bid name to consider, it consists of numerous nodes to be monitored with lot of precision to deal with the incoming request dynamically to the idle nodes or servers. Requests that are requested by the users will be satisfied by the reassigning of the incoming request to other server nodes that are less loaded from the overloaded nodes of the server as the server consist of 3 different states: idle, busy and normal. With the use of proper well-defined algorithm, one can get efficiency and scalability in providing unbreakable services.

REFERENCES

[1]. Wen, Wei-Tao, Chang-Dong Wang, De-Shen Wu, and Ying-Yan Xie. "An ACO-Based Scheduling Strategy on Load Balancing in Cloud Computing Environment." In Frontier of Computer Science and Technology (FCST), 2015 Ninth International Conference on, pp. 364-369. IEEE, 2015.

[2]. Issawi, Sally F., Alaa Al Halees, and Mohammed Radi. "An Efficient Adaptive Load Balancing Algorithm for Cloud Computing Under Bursty Workloads."Engineering, Technology & Applied Science Research 5, no. 3 (2015): pp-795.

[3]. LD, Dhinesh Babu, and P. Venkata Krishna. "Honey bee behavior inspired load balancing of tasks in cloud computing environments." Applied Soft Computing 13, no. 5 (2013): 2292-2303.