© 2018 IJSRSET | Volume 4 | Issue 1 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

A Methodological Survey on Different types of Load Balancing Algorithms in Cloud Computing

Jaimeel Shah
Department of Computer Science Engineering, Parul University, Gujarat, India

ABSTRACT

Cloud Computing is one of the quickly developing area in research industry today. Cloud Computing spins around web based computing which leads to the provision of the resources automatically to the customer based on pay per use service. As cloud computing is the mixture of different technologies such as distributed computing, parallel computing and virtualization which leads to the development of the new era. Due to evolution of the cloud, there is a feasibility about how different services can be provided to the customer through internet. Today due to the recent advancements there are various cloud service providers which provides the services based on the user demand. Although there are various challenges for the cloud service provider but one of the major goal for the cloud service provider is to utilize the resources efficiently and to balance the load among the resources efficiently. In this paper we will try to explore the concept of the load balancing and will also study about the different types of load balancing algorithms along with it we will also try to pinpoint different issues related to the concept of load balancing.

Keywords : Cloud Computing, Load Balancing

I. INTRODUCTION

Nowadays cloud computing is one of the extensively and widely used technology in the domain of Computer Science and information technology enabled services. The several benefits of cloud computing such as high flexibility, better cost effectiveness, pay per use models, scalability, and reliability of cloud computing technology leads several service providers and research scientists towards shifting to it. The main purpose of the cloud computing is to provide the services to the customer at anytime, anywhere based on the pay per use service model [2].

Cloud Computing consist of certain features such as scalability, elasticity and flexibility of cloud computing technology has led several service providers migrate to it. Also, several other features such as 24*7 availability from anywhere, pay-per-use model, and freedom from the capital expenditure has led an upsurge in the adoption of the cloud computing. Although, cloud computing provides several benefits to the clients, but cloud computing consist of many challenges such as load balancing, scheduling, security, data-center energy consumption, service availability and performance monitoring which are still an issues [3]. In this paper we deal with one of the major challenges of cloud computing is Load Balancing. The main purpose of load balancing is to assign the load among the resources to achieve the minimum utilization time, maximum throughput, and to improve the performance. Although, several researches have been conducted in the field of load balancing and task scheduling, our studies show the key role of load balancing and beneficiaries of the load balancing algorithms. In this paper we present a survey of the current load balancing algorithms developed specifically to suit the Cloud Computing environments. We provide an overview of these algorithms and discuss their properties [5]. In addition, we compare these algorithms based on the following properties: the number of attributes taken into consideration such as minimum response time, resource utilization and many more.

The rest of this paper is organized as follows. We discuss the challenges of load balancing in cloud computing in Section II. Moreover, In Section III we will discuss about the current literature and the algorithms proposed to solve the load balancing issues in Cloud Computing. In addition to this, will compare the relevant approaches in Section IV. Lastly we conclude the paper and show possible areas of enhancement and our future plan of improving load balancing algorithms in Section V.

II. LOAD BALANCING ISSUES

Before we could survey the present load balancing approaches for Cloud Computing, we have to recognize the principle issues and difficulties included and that could influence how the algorithm would perform. In this section we will discuss the challenges that takes place while performing the load balancing algorithms. The challenges are as follows[1-8].

1. Virtual Machine Migration: With the help of virtualization technique in cloud computing n number of virtual machines can be created on the physical server but due to increase of task based on the incoming request on virtual machines lead to overload of the server [10]. So to reduce the load on the physical server a virtual machine migration technique is mainly used. However, before performing migrations of the virtual machines some of the questions need to be answered. How to distribute the workload dynamically at the time of migration of VM at the time of bottleneck?
2. Energy Efficiency: Nowadays, due to increase in the growth of the Internet services leads to consumption of the more resources. To handle the energy in the efficient manner a suitable load balancing algorithm for cloud computing is required to allocate the resources keeping in mind of not violating service level agreement and reduce energy consumption is one of the major challenge in cloud computing.

3. Spatial Distribution of nodes in cloud: To increase the performance, load balancing algorithm should consider the geographical distribution of the computing nodes. The challenge is that load balancing should be designed in such a way that request can be handled with minimum response time by considering certain parameters such as network bandwidth, communication speeds, and distance between clients and computing nodes.
4. Fault tolerance: Load balancing algorithm should be designed in such a way that if one of the nodes fails it won’t affect the overall system. So at the time of designing load balancing algorithm one should take care of whether the algorithm is centralized or decentralized.

III. LITERATURE REVIEW

In this section we will discuss about the concept of load balancing algorithm along with that we will also discuss about the different algorithms related to load balancing and the contribution of author’s.

Inderveer et.al.,(2015) have presented a systematic review of different types of load balancing algorithms and job migration techniques [2][13]. The authors compare some of the predefined load balancing algorithms which consist of the properties of the algorithm, parameters, advantages and disadvantages. The author also pays his attention on the challenges of the algorithms. However, their work lacks discussion related to cloud computing as they mainly emphasize on Grid Computing.

Mahmoud, Horbaty, Farrag et.al., (2015) have reviewed an algorithm such as Ant colony optimization, Artificial Bee colony optimization, Genetic algorithm, particle swarm optimization for balancing the load in cloud computing. The author(s) have also proposed a novel algorithm called Ant Lion Optimizer (ALO) which provides better result in the terms of quality of service and response time as compared to the traditional ones[8].

Chen, Wang et.al., (2014) have studied a different type of algorithm such as Min-Min algorithm, Throttled algorithm for balancing the load in cloud computing. The author(s) have proposed a new algorithm based on the user priorities which consist of the extended version of the Min-Min algorithm. However, this works only for the static environment.

Generally, Load balancing algorithm are mainly divided into two types First one is Static Load Balancing algorithm and Secondly is Dynamic Load Balancing Algorithm.

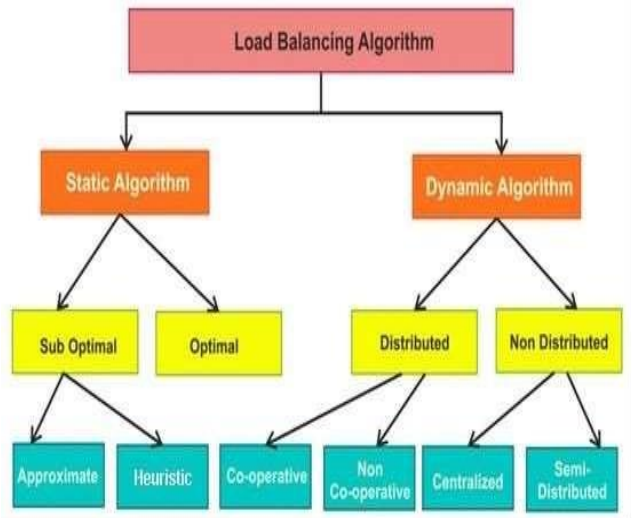

Figure 1: Classification of Load Balancing Algorithm[4]

Static algorithms are mainly divided into two parts optimal algorithm and sub-optimal algorithm (Neeraj et.al., 2014). In optimal algorithms, firstly the data center controller will determine the information regarding the tasks and resources and load balancer makes an optimal allocation within the reasonable time. If the load balancer cannot allocate it within the optimal time then sub-optimal allocation is to be calculated. Sub- optimal algorithms are classified into two classes such as approximate mechanism and heuristic mechanism. In approximate mechanism, load balancing algorithm will terminate after finding the good solution and it will not search the whole solution space whereas in Heuristic mechanism the decisions are taken by assumptions of tasks and resources in load balancing algorithm. Dynamic Load Balancing algorithms are classified into two classes such as distributed and non-distributed. In distributed load balancing algorithm, all the nodes will execute load balancing algorithm in the system and all the respective tasks are shared equally by all the nodes. Moreover, Distributed load balancing algorithms are classified into two parts which consist of co-operative and non- co-operative [6]]. In co-operative algorithm all the nodes work together to achieve the common goal for example to increase the throughput of the nodes. In non-co-operative algorithm all the nodes work independently to achieve their own goal for example to execute the predefined task on the local nodes. Non distributed algorithms are classified into two classes such as centralized and semi-distributed. In centralized load balancing algorithm a single node will act as a central node and central node will be responsible for executing the load balancing algorithm among all other nodes. However if the central node get failed, then to execute the centralized load balancing task is a difficult one. In semi-distributed approach, nodes are divided into the clusters and each cluster acts as a centralized form. Classification of the load balancing algorithm is shown in Figure 2. Different types of Load Balancing algorithms are as follows[10-16].

1) Round Robin Algorithm: In this algorithm, process are partitioned between the processors in such a way that workload between processes are distributed equally based on Turnaround time. The Pseudocode for round robin algorithm are as follows

```
Create P1, P2;
P1 = store ready process
P2= store blocked process
```

```
        New process submitted to end of P1
        If task time interval finished then
                Move to end of P1.
If I/O request or swapped out request is made by process then
 Move process from P1 to P2.
If I/O operation is completed or ready to move  from blocked processes then Move process from P2 to P1.
```

2) Active Clustering Load balancing Algorithm: Active clustering is one of the improved methods of random sampling. The main principle of the algorithm is grouping the similar node together [5]. When an execution starts, first node will select the neighbor nodes and based on the Euclidean distance calculated, the nodes will be placed within the clusters. To compute the distance between the two nodes, Euclidean distance can be calculated by

$$M_w = \frac{1}{T}\sum_{t=1}^{T}\sqrt{(x(t) - x(t-1))^2 + (y(t) - y(t-1))^2}$$

Where $((t) - (t - 1))$ and $((t) - y(t - 1))$ are the coordinate node at time t and (t-1) M_w = Distance between the two nodes.

3) Min–Min Load Balancing Algorithm: Min-Min algorithm is the static load balancing algorithm. Firstly, the minimum completion time of all the tasks need to be calculated. The job which consist of minimum execution time are assigned first to the processors but the task having the maximum execution time have to wait for the specific period of the time[5]. All the tasks in the processor must be updated and the tasks in the queue have to be removed once the task gets completed. The pseudo code of the Min-Min load balancing algorithm is

```
Procedure Min-min(Task Ti)
{
Find execution_Time for each task
Store the execution_Completion_Time of task Ti in orderQueue
{
for each task Ti in orderQueue
{
obtain minimum completionTime from orderQueue; assign task to vm;
update the execution_Completion_Time;
}
Until orderqueue empty;
}
```

4) Genetic Algorithm

The basic parts of the genetic algorithm consist of mutation and crossover methods, with the knowledge that these method will depend upon the type of chromosomes which consider either tree encoding, binary coding, or numeric coding. First of all the crossover method will exchange the parts of the two parents for generating the better children using one or two crossover points. Secondly the mutation method takes place after the crossover method takes place which leads to the decrease in the local optima[17]. The pseudo code of the genetic algorithm is

```
Initialize the population of n chromosomes randomly.
(Each chromosome, consist of randomly assigned each task in the list to a VM)
While (end criterion != satisfied)
Fitness of each chromosome to be calculated.
Good population to be selected for new children generation.
DO cross over operations to be performed on selected chromosomes. With the help of specific mutation probability, Mutate the new children. Update the population with the new children to form the new population
End while
Return the best chromosome of the final population
```

5) Join Idle Queue Load Balancing Algorithm

This algorithm is mainly use for the large scale distributed system. Firstly it utilizes disseminated dispatchers by first load balancing the sit without moving processors crosswise over dispatchers and after that appointing jobs to processors to reduce average queue length at every processor [8]. The disadvantage of this algorithm is that it is not scalable.

6) CARTON Load Balancing Algorithm

CARTON load balancing algorithm are mainly divided into 2 parts. First is Load Balancing and second is Distributed Rate Limiting (DRL) [6]. LB (Load Balancing) is utilized to equally distribute the job among various servers so that the related cost can be minimized and DRL (Distributed Rate Limiting) is utilized to ensure a fair resource allocation. DRL additionally adjusts to server capacities for the dynamic workloads with the goal that execution levels at all servers are equivalent. With low calculation and correspondence overhead, algorithm is simple to execute [18].

7) Power Aware Load Balancing Algorithm

In this algorithm firstly the CPU utilization factor need to be calculate for all the active nodes. Secondly Threshold value for the CPU utilization need to be set (For example the threshold value is 75%). If all the active nodes have more utilization factor as compare to the threshold value the new virtual machine need to be booted on the compute node otherwise the task will be handled by the active nodes[15].

IV. COMPARISION OF DIFFERENT LOAD BALANCING ALGORITHMS

In this section we will compare the different load balancing algorithms based on the performance metrics of the load balancing.

TABLE I Comparison of Load Balancing Algorithm

Algorithm	VM Allocation Dynamicity	VM Type Uniformity	Parameters	Challenges
Round Robin Algorithm	Static	Homogeneous	Waiting Time	Less Resource Utilization
Min Min Algorithm	Static	Homogeneous	Response time	Starvation
Genetic Algorithm	Dynamic	Heterogeneous	Process Utilization	All the jobs have the equal Priority.
Active Clustering Load Balancing Algorithm	Static	Homogeneous	Response Time	Every time iteration takes place for allocation for VMs
Carton Load balancing Algorithm	Dynamic	Heterogeneous	Distributed Rate Limiting	Less Resource Utilization
Throttled Load Balancing Algorithm	Dynamic	Homogeneous and Heterogeneous	Throughput, Resource Utilization	Requires maximum Vm and completion time is high.
Join Idle Queue Load Balancing Algorithm	Dynamic	Heterogeneous	Response Time	Less Resource utilization
Power Aware Load Balancing Algorithm	Static	Homogeneous	Resource Utilization	Maximum Response time.

V. CONCLUSION

Load Balancing in the cloud computing is one of the major challenges in the cloud computing environment. In this paper we have reviewed the multiple algorithms and have also studied about the different load balancing algorithms which exist in the cloud computing. The ultimate goal of load balancing in cloud computing is to provide the profit for cloud service providers by maximum utilization of the resources and to minimize the cost for the cloud consumers so that there task can be done easily.

VI. REFERENCES

[1]. Klaithem Al Nuaimi, Nader Mohamed, Mariam Al Nuaimi and Jameela Al-Jaroodi "A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms" IEEE 2012.

[2]. Surbhi Kapoor, Dr. Chetna Dabas "Cluster Based Load Balancing in Cloud Computing" IEEE 2015.

[3]. Abhinav Hans, Sheetal Kalra "Comparative Study of Different Cloud Computing Load Balancing Techniques" IEEE 2014.

[4]. Reena Panwar, Prof. Dr. Bhawna Mallick "Load Balancing in Cloud Computing Using Dynamic Load Management Algorithm" IEEE 2015.

[5]. Kumar Nishant, Pratik Sharma, Vishal Krishna, Chhavi Gupta and Kuwar Pratap Singh "Load Balancing of Nodes in Cloud Using Ant Colony Optimization" International Conference on Modelling and Simulation IEEE 2012.

[6]. Ariharan V,Sheeja Manakatlu "Neighbor Aware Random Sampling (NARS) algorithm for load balancing in Cloud computing" IEEE-2015.

[7]. A.kumar "Load Balancing in Cloud Data Center Using Modified Active Monitoring Load Balancer" IEEE-2016.

[8]. Garima Rastogi, Dr Rama Sushil, "Analytical Literature Survey on Existing Load Balancing Schemes in Cloud Computing", IEEE 2015.

[9]. Nishant, K. P. Sharma, V. Krishna, C. Gupta, KP. Singh, N. Nitin and R. Rastogi, "Load Balancing of Nodes in Cloud Using Ant Colony Optimization." In proc. 14th International Conference on Computer Modelling and Simulation (UKSim), IEEE, pp: 3-8, March 2012.

[10]. Zhang, Z. and X. Zhang, "A load balancing mechanism based on ant colony and complex network theory in open cloud computing federation." In proc. 2nd International Conference on. Industrial Mechatronics and Automation (ICIMA), IEEE, Vol. 2, pp:240-243,May 2010.

[11]. Kolb, L., A. Thor, and E. Rahm, E, "Load Balancing for MapReducebasedEntity Resolution," in proc. 28th International Conference on Data Engineering (ICDE), IEEE, pp: 618-629, 2012.

[12]. Gunarathne, T., T-L. Wu, J. Qiu and G. Fox, "MapReduce in the Clouds for Science," in proc. 2nd International Conference on Cloud Computing Technology and Science (CloudCom), IEEE, pp:565-572, November/December 2010.

[13]. Ni, J., Y. Huang, Z. Luan, J. Zhang and D. Qian, "Virtual machine mapping policy based on load balancing in private cloud environment," in proc. International Conference on Cloud and Service Computing (CSC), IEEE, pp: 292-295, December 2011.

[14]. Nishant, K., Sharma, P., Krishna, V., Gupta, C., Singh, KP., Nitin, N., and Rastogi,R. (2012). Load Balancing of Nodes in Cloud Using Ant Colony Optimization. 14th International Conference on Modelling and Simulation,3-8.

[15]. Nuaimi, K.A., et al(2012). A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms. in Network Cloud Computing and Applications (NCCA), Second Symposium on. 2012.IEEE.

[16]. Chien,NK.,Son,NH.,andHD.(2016).LoadbalancingalgorithmBasedonEstimatingFinishTimeofServicesinCloud Computing, International Conference on Advanced CommutationTechnology(ICACT), 228-233.

[17]. Gupta, P., M.K. Goyal, and P. Kumar et.al (2013). Trust and reliability based load balancing algorithm for cloud IaaS. in Advance Computing Conference (IACC),IEEE 3rd International. 2013.IEEE.

[18]. Wang, S.-C., et al(2010). Towards a load balancing in a three-level cloud computing network. in Computer Science and Information Technology (ICCSIT), 3rd IEEE InternationalConference.

[19]. Keshvadi, S. and Faghih, B. (2016). A Multi-Agent based Load balancing System in IaaS Cloud Environment. International Robotics and Automation Journal, 1(1).