IJCSC 0973-7391 Volume 7 • Number 1 Sept 2015 - March 2016 pp. 254-258 available online at www.csjournalss.com

Review of various Load Balancing Techniques in Cloud Computing

Jyoti Rathore
Research Scholar Computer Science & Engineering,
Suresh Gyan Vihar University, Jaipur Email: Jyoti.rathore131@gmail.com
Dr. Bright Keswani
Professor, Department of Computer Applications, Suresh Gyan Vihar University, Jaipur
Email: kbright@rediffmail.com
Dr. Vijay Singh Rathore
Professor, Department of Computer Science & Engineering, Jaipur Engineering College & Research Centre, Jaipur
Email: Vijaydiamond@gmail.com

Abstract: Cloud computing is a new concept in IT environment. Infrastructure and resources are in great demand in cloud computing. Cloud computing deliver the cloud services effectively and efficiently to the consumers on pay-per usage basis. Load Balancing is an important aspect of cloud computing environment. Efficient load balancing scheme handles the user's request in efficient way. Load Balancing may even support prioritizing users by applying appropriate scheduling criteria. This paper presents various load balancing schemes in different cloud environment.
Keywords: Cloud Computing, Load Balancing, etc.

I. INTRODUCTION

A Cloud computing new technology of large scale distributed computing [10]. Cloud computing providers, such as Google, Amazon, Yahoo and Microsoft deliver services to customers all over the world. There is no need to install software's in PC's as cloud computing is platform independent and it provide online applications to users on-demand [1]. There are many problems prevalent in cloud computing. The main problem in Cloud computing is load balancing [2]. Load balancing is about distributing the load among various nodes to improve both resource utilization and job response time. It avoid a situation where some of the nodes are idle or doing little work while other are heavily loaded [10].

II. CLOUD COMPUTING

The " Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, data storage, software applications and other computing services) that can be rapidly provisioned and released with minimal management effort or service provider interaction "[4]. Because of resource sharing the overall cost reduces. It delivers all services through the internet dynamically, such as operating system, network, storage, software, hardware and resources [11]. Cloud is a pool of heterogeneous resources. It is a huge infrastructure. It has no relevance with its name "Cloud". Infrastructure refers to both the applications delivered to end users as services over the Internet and the hardware and system software in datacenters that is responsible for providing those services. To efficiently use these resources and make sure that resources are available to the end users "Computing" is done [5].

III. LOAD BALANCING

A load balancing algorithm avoid overloading or under loading of any specific node [4]. The selection of load balancing algorithm is not easy because it involves additional constraints like security, reliability, throughput etc. So, the main goal of a load balancing algorithm is to improve the response time of job by distributing the total load of system [6]. The algorithm makes sure that no specific node is overloaded. While developing such algorithm important things to consider are the estimation of load, the comparison of load, the stability of different system, the performance of system, the interaction between the nodes, the nature of work to be transferred, the selection of nodes and many other ones [4]. This load considered can be in terms of CPU load, amount of memory used, delay or Network load.

There are mainly two types of load balancing algorithms

1) Static Algorithm requires a prior knowledge of system resources, so that shifting of the load does not depend on the current state of system and traffic is divided evenly among the servers. Static algorithm is proper for the system which has low variation in load.

2) In Dynamic Algorithm lightest server in the whole network is searched and preferred for balancing a load. For this real time communication with network is needed. To manage the load current state of the system is used.

IV. TECHNIQUES SUPPORTING LOAD BALANCING: A REVIEW

The Round Robin utilizes the principle of time slices. Here node is given a particular time slice or time interval. Each node is given a time quantum and its operation. The Resources are provided to the requesting client on the basis of time slice [17]. In max-min algorithm first all the available tasks are submitted to the system and minimum completion time is calculated, then task which having maximum is chosen and allocated to the machine which can execute that task in less time. This algorithm performs well when short tasks are in high numbers then long ones. For e.g. if in a task set only a single long task is presented then ,Max Min algorithm runs short tasks concurrently along with long task. The make span focus on how much small tasks will get executed concurrently with the large ones. The algorithm demerit is it suffers from starvation [12][8].

In Ant Colony Optimization Technique pheromone table is maintained which was updated by ants as per the resource utilization and node selection. Ants move in forward direction in search of the overloaded or under loaded node. As the overloaded node is found, then ants traverse back to fill the recently encountered under loaded node [16].

The Min-Min Algorithm first calculates the minimum completion time for the entire task. The task having the completion time minimum is chosen and assigned to the node which can execute that task in less time. This is done till all the tasks have been assigned to the equivalent machines. The algorithm performs better when small tasks are greater in number than the large tasks. The algorithm has a disadvantage that it leads to starvation [8].

In Equally Spread Current Execution technique load balancer maintains an index table of Virtual machines as well as number of requests currently assigned to the Virtual Machine (VM). If the request comes from the data centre to allocate the new VM, it scans the index table from top for least loaded VM. In case there are more than one VM is found than first VM is selected. The load balancer also returns the VM id to the data centre controller. The data centre communicates the request to the VM identified by that id. By increasing the allocation count of identified VM the index table is updated. When VM completes the assigned task, a request is communicated to data centre which is further notified by the load balancer. The load balancer again updates the index table by decreasing the allocation count for identified VM by one [3].

The Honey Bee algorithm is a nature inspired load balancing technique. It achieves load balancing across heterogeneous virtual machine of cloud computing environment and maximizes the throughput. First the current workload of the VM is calculated to decide the VM states whether it is over loaded, under loaded or balanced. The priority of the task is taken into consideration after removed from the overloaded VM which are waiting for the VM. Then the task is scheduled to the lightly loaded VM. It reduces the response time of VM and also reduces the waiting time of task [13].

In throttled algorithm the load balancer maintains index table of virtual machines with its state (busy or ideal). When task came the data center queries the load balancer for allocation of the VM. The load balancer checks the index table from top to find first available VM, if the VM is found, the load data center communicates the request to the VM identified by the id. Further, if appropriate VM is not found, the load balancer returns -1 to the data center. When the VM completed the allocated task, a request is acknowledged to data center, which is apprised to load balancer to de- allocate the same VM whose id is already communicated. [10].

The Active Clustering is a self aggregation technique. Here same types of nodes are grouped together and work on these groups is done. Here network is rewired to balance the load of the system. Systems optimize using similar job assignments by connecting similar services. The performance of the system is enhanced with high resources thereby increasing the throughput by using these resources effectively [2].

The Biased Random Sampling is based on virtual graph construction. Each node of the graph corresponds to the computer of the cloud system. Edges b/w nodes are two types as Incoming edge and outgoing edge that is used to consider the load of particular system and also allotment the resources of the node [14]. It is scalable, reliable and effective load balancing approach that is mainly developed to balance the load of distributed system. Load balancing is achieved without the need to monitor the nodes for their resources availability [2].

The Decentralized Content Aware is named as workload and client aware policy (WCAP). It uses a unique and special property (USP) to specify the unique and special property of request as well as computing nodes. USP helps the scheduler to decide the best suitable node for processing the requests. This strategy is implemented in a decentralized manner with low overhead. It also helps in reducing the idle time of the computing nodes hence improving their utilization [16].

The Map Reduced based Entity Resolution load balancing technique is based on large datasets. There are two main tasks: Map task and Reduce task. The PART method is executed for mapping task, where the request entity is partitioned into parts. And then COMP method is used to compare the parts and finally GROUP method is use to group similar entities and then similar task are group using Reduce function and this also reduces the results of the tasks. Map task reads the entities in parallel and process them, so that overloading of the task is reduced [2].

The Opportunistic Load Balancing (OLB) is a static load balancing algorithm whose goal is to keep each node in the cloud busy so does not consider the current load on each node. It dispatches the selected job to a randomly selected available VM. OLB does not consider the execution time of the task in that node. This may cause the task to be processed in a slower manner and will cause some bottlenecks since requests might be pending waiting for nodes to be free [15].

The Cloud Partitioning Concept is a dynamic load balancing technique for cloud based on partitioning concept with a switch mechanism to choose different strategies for different situations. Cloud Partitioning is for Public cloud which has numerous nodes with distributed computing resources at different geographic locations. Here cloud divides into several cloud partitions. The cloud has main controller that chooses the suitable partitions for arriving jobs and there is balancer for cloud partition which chooses the

correct load balancing strategy. If the load status of a cloud partition is idle and normal, this partitioning accomplished job locally and when load status is not normal, another cloud partition is searched [9].

The Carton is a technique that is a combination of Load Balancing (LB) and Distributed Rate Limiting (DRL). LB (Load Balancing) is used to equally distribute the jobs to different servers so that the associated costs can be minimized. DRL ensure the equal distribution of resources. Work load is dynamically assigned to improve the performance and spread the load equally to all the servers. With very low computation and communication overhead, this algorithm is simple and easy to implement [7].

The Compare and Balance algorithm uses the concept of compare and balance to reach an equilibrium condition and manage unbalanced system's load. On the basis of probability (no. of virtual machine running on the current host and whole cloud system), current host randomly select a host and compare their load. If load of current host is more than the selected host, it transfers extra load to that node. Each host of the system performs the same procedure [2]. This algorithm assures that the migration of VMs is always from high cost physical hosts to low-cost host but assumes that each physical host has enough memory which is a weak assumption [7].

V. CONCLUSION

This paper discussed that there are infinite computing capabilities with attractive pay-per-use scheme in cloud computing. Request on the cloud are in random manner because of this randomness load balancing occurs. Load balancing is major concern in cloud computing. Particular node get overloaded which slow down the node and results in poor system efficiency. Because of which efficient load balancing algorithms is required for improving the utilization of computing resource. In this paper, we have analyzed various load balancing algorithms for cloud computing. The main task of load balancing is to distribute load dynamically among the nodes and to make maximum resource utilization by reassigning the total load to individual node. So the performance of the system is increased.

REFERENCES

[1]. Bala K. et. al., 'A Review Of The Load Balancing Techniques At Cloud Server', International Journal of Advances in Computer Science and Communication Engineering (IJACSCE), Vol. 2, Issue I, 2014, pp 6-11.

[2]. Desai T. and Prajapati J., 'A Survey of Various Load Balancing Techniques and Challenges in Cloud Computing', International Journal Of Scientific & Technology Research, Vol. 2 , Issue 11, 2013, pp 158-161.

[3]. Katyal M. and Mishra A., 'A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment', International Journal of Distributed and Cloud Computing, Vol. 1, Issue 2, 2013, pp 5-14. 4.

[4]. Sahu Y. and Pateriya R., 'Cloud Computing Overview with Load Balancing Techniques', International Journal of Computer Applications , Vol. 65– No.24 , 2013, pp 40-44.

[5]. Gabi D. et. al., 'Systematic Review on Existing Load Balancing Techniques in Cloud Computing', International Journal of Computer Applications, Vol. 125 – No.9, 2015, pp 16-24.

[6]. Palta R. and Jeet R., 'Load Balancing in the Cloud Computing Using virtual Machine Migration: A Review', International Journal of Application or Innovation in Engineering & Management (IJAIEM), Vol. 3, Issue 5, 2014, pp 437-441.

[7]. Ramesh B., 'Load balancing in Cloud Computing – An Analysis', International Conference on Security and Authentication - SAPIENCE141, 2014, pp 125-131.

[8]. Gopinath G. and Vasudevan S., 'An in-depth analysis and study of Load balancing techniques in the cloud computing environment', 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15), 2014, pp 427- 432.

[9]. More S. and Mohanpurkar A., 'Load Balancing Strategy Based On Cloud Partitioning Concept', Suresh Gyan Vihar University Journal of Engineering & Technology (An International Bi-Annual Journal) Vol. 3, Issue 2, 2017, pp 48-52 ISSN: 2395-0196 52 Multidisciplinary Journal of Research in Engineering and Technology, Vol. 2, Issue 2, 2015, pp 424-431.

[10]. Kaur P. and Kaur P., 'Efficient and Enhanced Load Balancing Algorithms in Cloud Computing', International Journal of Grid Distribution Computing, Vol. 8, No.2 , 2015 pp 9-14.

[11]. Panwar R. and Mallick B., 'A Comparative Study of Load Balancing Algorithms in Cloud Computing', International Journal of Computer Applications, Vol. 117 – No. 24, 2015, pp 33- 37.

[12]. Sadhu R. and Vania J., 'Survey on Various Load Balancing Techniques in Cloud Computing Environment', International Journal of Science and Research (IJSR), Vol. 4, Issue 10, 2015, pp 1881-1884.

[13]. Vouk M., 'Cloud Computing – Issues, Research and Implementations', Journal of Computing and Information Technology – CIT, 2008, pp 235–246.

[14]. Megharaj G. and Mohan K.G., 'Two Level Hierarchical Model of Load Balancing in Cloud', International Journal of Emerging Technology and Advanced Engineering, Vol. 3, Issue 10, 2013, pp 307-311.

[15]. Singh A. et. al., 'Comparative Analysis Of Load Balancing Algorithms In Cloud Computing', International Journal of Advanced Technology & Engineering Research (IJATER), Vol. 4, Issue 2, 2014, pp 18-21.

[16]. Gupta R., 'Review on Existing Load Balancing Techniques of Cloud Computing', International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 2, 2014, pp 168-171.

[17]. Pasha N. et. al., 'Round Robin Approach for VM Load Balancing Algorithm in Cloud Computing Environment', International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 5, 2014, pp 34-39.