International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 01
January2017

A Review on Various Approaches of Load Balancing In Cloud Computing

Muskan Garg & Rajneesh Narula
Research Scholar IT department at HOD of IT Adesh Institute of Engg.& Technology, FaridkotAdesh Institute of Engg.& Technology, Faridkot
CSE department at Adesh Institute of Engg.& Technology, FaridkotAdesh Institute of Engg.& Technology, Faridkot
muskangrg96@gmail.com ; rajnarula51@gmail.com

Abstract-*In cloud computing various users sends request for the transmission of data for different demands. The access to different no. of user increases load on the cloud servers. Due to these cloud server does not provides best efficiency. In this paper the main problem in the scheduling process is to allocation of virtual machines to different user-bases for response to their requests. The allocation of VM must be in such manner to each UB so that minimum response time has been provided.*

Keywords: *load balancing; cloud computing, Round-Robin, min-min, VM, response time, Data Center Response time and Load***.**

INTRODUCTION

Cloud computing is an evolving area that allows users to organize applications with enhanced scalability, availability and fault tolerance. Cloud computing provides internet based platform that is used for computer technology. It describes a diversity of computing concepts [8].Cloud computing accumulates all the computing resources and manages them automatically. Nowaday's world depends on cloud computing to store the public as well as personal information. Cloud computing provides relevant hardware, software and service according to the requirement that users put forward. A cloud computing structure is categorized by its on-need self-service, access over internet, pooling of resources, elasticity of service availability and measurement of services utilized by individual users. Cloud computing provides a collective group of resources, including data storage space, networks, computer processing power and specialized corporate and user application. There are four deployment models in cloud computing. They are

- Public
- Private
- Community
- Hybrid.

1.1 Cloud Services

Cloud computing provides a number of clouds for providing services. Services can be larger or smaller, and use of a service is measured and customers are billed accordingly [9]. Service Models are the orientation models on which cloud computing is based. These can be categorized into three basic service models as listed below:

- Infrastructure as a Service (IaaS)
- Platform as a Service (PaaS)
- Software as a Service (SaaS).

1.2 Benefits of Load Balancing

- **Redundancy**

It describes the process of running two or more, the same servers thus providing a guaranteed event that one server becomes occupied.

- **Scalability**

Even though modest resources requirements are offered, scalability must always be considered for finding the correct host solution.

- **Resource Optimization**

Through load balancing, one can optimize how traffic is circulated to the server cluster so, that it guarantees the best performance.

- **Security**

In security, only one IP is exposed to the web with load balancing, which significantly reduces the amount of break points in case of attack.

2. REVIEW OF LITERATURE

Magade, Krishnanjali A. et al The IEEE 802.11 standard does not provide any mechanism to resolve load imbalance in the network. To reduce this deficiency, various loadbalancing techniques have been designed. Loadbalancing provides a cost-effective, efficient and transparent method to expand the bandwidth of network devices and servers, increase the throughput, and enhance the data process capability, thus increasing the flexibility and availability of networks. There are different techniques based on persistent algorithm for loadbalancing in Wireless LAN. The goal of the proposed work is to use Persistence weighted round robin algorithm for loadbalancing in wireless LAN. This algorithm is able to distribute mobile stations among all APs and the signal strengths between stations and access points are also being maximized at the same time. This technique will be useful to reduce the congestion in the network, maintain the load in balance condition on network, as well as improve the bandwidth utilization.

Yean-Fu Wen et al Author discuss about the loadbalancing problem is one of an open issues for the cloud computing. A good load balancing mechanism enhances the performance of network processing, optimizes the use of resources, and ensures that no overloading a single node or link case. The existing load balancing cloud computing research mainly unilateral the fairness of the transmission network or stand only for the system. Hence, this work considers the handling of both network and system load balancing to obtain high performance. To assign the tasks to the same type of nodes along the links with minimum processing and transmission delays subject to the capacities of nodes and links. Three task assignment schemes FCFS, Min-Min, and Min-Max are adopted along with dynamic clustering, which is a method to group the same type of cloud servers. This study changes in the variables manipulated with the number of nodes and the number of tasks and records the maximal end-to-end delay, average end-to-end delay and fairness index, to analyze the loadbalancing results. The results show that the Min-Max combination with dynamic clustering has a good effect.

de Mello, M.O.M.et al The aim of the author to introduce a new loadbalancing routing heuristic, called BPR - Bottleneck, Path length and Routing overhead, which offers an efficient online solution by taking into consideration all these aspects of the loadbalancing routing problem. We have carried out a simulation study to compare the performance of BPR with a recent related work. BPR obtains bottleneck values within desirable bounds, while reducing the average path length. As a result, BPR also notably reduces the number of route updates in the network, i.e. the routing overhead. Finally, we show that BPR is simple and has low demand for processing requirements.

Biased Random Sampling Algorithm: This algorithm[3] is based on the construction of the virtual graph having connectivity between the all nodes of the system where each node of the graph is corresponding to the node computer of the cloud system. Edges b/w nodes are of 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. It is scalable technique to balance the load of the cloud system. It is also reliable and effective load balancing approach that is mainly developed to **balance the load of distributed system. Compare and Balance Algorithm**: This algorithm[6] 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.

3. APPROACHES USED

3.1 Min-Min Load Balancing Algorithm

Min-Min is a static load balancing algorithm, where the parameters associated to the job are recognized in advance. In Min-Min algorithm, the execution and completion time of the unassigned waiting in queue are identified by the cloud manager. The jobs with minimum execution in time are being assigned first to the processors, so that the task is completed in time. But the tasks with maximum execution need to wait for a specific period of time. As such, all the all the tasks in the processor must be updated and the tasks in the queue must be removed. The task with minimum time execution performs better than the maximum time execution. The main disadvantage of this algorithm is that it leads to starvation. The terminology related to static load balancing for Min-Min is [5]

Excepted Time of Compute (ETC) - The running time excepted for tasks in all the nodes are stored in ETC,

Minimum Execution Time Algorithm (MET) – It finds the best job-processor-pair, were current load is not considered, and

Minimum Completion Time Algorithm (MCT) –It allocates the tasks based on the minimum completion time.

3.2 Round Robin Algorithm

Round Robin is one of the static load balancing algorithms, where preceding states are not taken into account. It is simple and uses the Round Robin Method for job allocation. It selects the first node at random and then allocates the job to all the other nodes evenly in Round Robin Method. The main advantage of Round Robin is that it does not need any interposes communication. There is no prior information about the processors' running time, so that some tasks may get heavily loaded. To overcome this, weighted Round Robin algorithm is being proposed. Here each node assigned has a specific weight. Based on the nodes weight, they would get the requests. If all nodes are equal, then the node is indicated to traffic.

3.3 Opportunistic Load Balancing Algorithm

It is also one of the static load balancing algorithms, which do not consider the present workload of the VM. It usually keeps each and every node busy. This deals with the unexecuted tasks quickly and in random order to the current node, where each one of task is assigned to the node randomly. This algorithm provides a load balancing schedule but does not produce a good result. The tasks are processed in a slow manner, where the current execution time of the node is not calculated

3.4 Ant Colony Optimization Based Load Balancing Algorithm

This algorithm is designed to seek out the optimal path among the food and colony of ant, based on its actions. The main aim of this approach is to distribute the work load among the nodes in an efficient manner. The regional load balancing node is preferred as head node in Cloud Computing Service Provider [4]. As the request is being sent, the ant starts is first movement from the head node [5]. The ants collect the information from the cloud node and assign the tasks to the particular node. Once the task is assigned to the head node, the ant moves in a forward direction with the overloaded node to the next node checking whether the node is overloaded or not. During the movement, if it finds any loaded node again it moves in a forward direction, else it finds the overloaded node, it moves in backward direction and replaces were the node found before [6].Once the job gets successful it is updated, then the result is reported based on the individual result of the ant. After receiving the individual result they are combined together to build the complete report. The solution set is updated automatically, when the ant updates the result for every movement. To prevent backward movement, the ant commits suicide when it reaches the target node

3.CONCLUSION

In cloud computing various users sends request for the transmission of data for different demands. The access to different no. of user increases load on the cloud servers. Due to these cloud server does not provides best efficiency. To provide best efficiency load has to be balanced main problem in the paper is that different jobs can be divides in tasks. The job dependency checking is done on the basis of directed a cyclic graph. The dependency checking the make span has to create on the basis of shortest job first and pound robin approach. The minimization can be done on the basis of using hybrid RR andmin-min algorithm.

REFRENCES

[1] Amir Nahir"Distributed Oblivious Load Balancing Using Prioritized Job Replication", ISSN 978-3-901882-48-7, IEEE, 2012.

[2] Hong Tao "A dynamic data allocation method with improved load-balancing for cloud storage system", ISSN 978-1-84919-707-6, PP 220 – 225, IEEE, 2013

[3] Yuqi Zhang "Dynamic load-balanced multicast based on the Eucalyptus open-source cloud-computing system",ISSN 978-1-61284-158-8, pp. 456 – 460, IEEE, 2011.

[4] Magade, Krishnanjali A. "Techniques for load balancing in Wireless LAN's**",** ISSN 978-1-4799-3357-0, PP 1831 – 1836, IEEE, 2014.

[5] Yean-Fu Wen "Load balancing job assignment for cluster-based cloud computing", ISSN 14517061, PP 199 – 204, IEEE, 2014.

[6] De Mello, M.O.M.C "Load balancing routing for path length and overhead controlling in Wireless Mesh Networks", ISSN 14630778, PP 1-6, IEEE, 2014.

[7] R. Angel Preethima, Margret Johnson, "Survey on Optimization Techniques for Task Scheduling in

Cloud Environment", IJARCSSE,Volume 3, Issue 12, December 2013.

[8] AkhilGoyal, Bharti, "A Study of Load Balancing in Cloud Computing using Soft Computing Techniques ",International Journal of Computer Applications (0975 – 8887) Volume 92 – No.9, April 2014

[9] NavjotKaur,Taranjit Singh Aulakh,Rajbir Singh Cheema, "Comparison of Workflow Scheduling Algorithms in Cloud Computing",(IJACSA) International Journal of Advanced Computer Science and Applications, vol. 2, No. 10, 2011.

[10] Mayank Singh Rana, Sendhil Kumar , Jaisankar N, "Comparison of Probabilistic Optimization Algorithms for Resource Scheduling in Cloud Computing Environment" International Journal of Engineering and Technology (IJET)

[11] C.Kalpana,U.Karthick Kumar, R.Gogulan, "Max-Min Particle Swarm Optimization Algorithm with Load Balancing for Distributed Task Scheduling on the Grid Environment", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 1, May 2012.

[12] ZHANG Yan-huaa, Feng Leia, Yang Zhia, "Optimization of Cloud Database Route Scheduling Based on Combination of Genetic Algorithm and Ant Colony Algorithm", Science direct,Procedia Engineering 15 (2011), pp. 3341 – 3345.

[13] Foster, I., Y. Zhao, I. Raicu and S. Lu, "Cloud Computing and Grid Computing 360-degree compared," in proc. Grid Computing Environment Workshop.

[14] Buyya R., R. Ranjan and RN. Calheiros, "InterCloud: Utility oriented federation of cloud computing environments for scaling of application services," in proc. 10th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP), Busan, South Korea, 2010.

[15] Martin Randles, EnasOdat, David Lamb, Osama Abu- Rahmeh and A. Taleb-Bendiab, "A Comparative Experiment in Distributed Load Balancing", 2009 Second International Conference on Developments in Systems Engineering