International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Index Copernicus Value (2015): 62.86 | Impact Factor (2015): 3.791

Load Balancing Techniques – A Review

Kuldeep[#1], Ravinder Kumar[#2]

[1]kuldeepboora8[at]gmail.com
[2]rssdhanda[at]gmail.com

Abstract: ***In the cloud storage, Load balancing is one of the central issues in cloud computing. It is a mechanism that disseminates the dynamic nearby workload equitably over every node in the entire cloud to avoid a circumstance where few nodes are intensely loaded while others are idle or doing little work. It accomplishes a high client fulfilment and resource usage proportion, thus enhancing the overall performance and resource utilization of the system. It likewise guarantees that each processing resource is distributed effectively and decently. In cloud system the resources are distributed to all computer nodes fairly and inefficiently manner. If a node has extra workload as the comparison of other nodes then we use load balancing techniques or process for transferring the extra load of nodes to others networks node. This paper provides a literature review of various load balancing algorithms in cloud computing.***

Keywords: Cloud Computing, Load Balancing, Round Robin, FCFS, Throttle

1. Introduction

Load balancing provides a procedure by which we distribute the dynamic workload across all cloud network nodes equally to escape a condition where various network nodes are greatly loaded while some other nodes are futile or doing slight work. In the cloud storage, Load balancing is one of the central issues in cloud computing. It is a mechanism that disseminates the dynamic nearby workload equitably over every node in the entire cloud to avoid a circumstance where few nodes are intensely loaded while others are idle or doing little work. It accomplishes a high client fulfilment and resource usage proportion, thus enhancing the overall performance and resource utilization of the system. It likewise guarantees that each processing resource is distributed effectively and decently.

2. Load Balancing

In cloud computing a CSP (Cloud Service Provider) is used in load balancing for distribution of resources across all the server, data enters and hard drives used in the network system of cloud the CSP work as ISP. It allots all application requirements across any number of application distributions nodes which are positioned in different data centers. Load balancing can be generally characterized by numerous techniques. These are: Integrated or decentralized, active or static, and periodic or non-periodic.

In cloud computing environment, a few resources are intensely loaded while others are idle or doing little work its fundamental reason is random arrival of task and random usage of CPU service time thus, in cloud computing load balance and resource control is principle key challenging issue. To achieve minimum response time, less energy efficiency, optimal resource utilization and remove overload of a node it provides a procedure to allocate workload across multiple computers, or nodes. The fundamental reason of load balancing is to enhance execution performance by balancing load among different resources viz. network links, disk drives, and central processing units. Different load balancing algorithms are used to distribute the load on to different various systems.

3. Basic Algorithm of Load Balancing

1) Round Robin Load Balancing: Round Robin (RR) algorithm focuses on the fairness. It is a static load balancing technique in which tasks are scheduled on the basis of time quantum. RR uses the ring as its queue to store jobs. The fundamental preferred standpoint is that it is a starvation free. Each procedure will be executed by CPU for settled time cut. So along these lines no procedure left for its round to be executed by the CPU. All tasks will performed with no prioritization. It chooses the load randomly and leads to the circumstance where a few nodes are intensely loaded and few are softly loaded. In spite of the fact that the algorithm is exceptionally straightforward and simple to actualize however there is an extra load on the scheduler to choose the time quantum [1].

2) Min–Min algorithm: The Min-Min algorithm first finding the minimum expected time of all tasks in meta-task. The task having the minimum expected completion time is selected and assigned to the corresponding resource. This step is iterated until meta-task is not empty. Here, a big task has to wait for the completion of smaller ones [1].

3) Max-Min algorithm: The Max-Min algorithm expected completion time of each task as per the available resource is calculated. A task which has overall maximum completion time is scheduled over a resource with overall minimum execution time. This step is repeated until meta-task is not empty. Here, the waiting time of a larger task is reduced [1].

4. Literature Review

Load balancing is the process of finding overloaded nodes and then transferring the extra load to other nodes. There are many load balancing algorithms for balancing the load among the servers of cloud data centers. Literature reviews of various load balancing work done in the past are explained below.

A. Fang et al. (2010)

In their paper, "to obtain high resource utilization and meet dynamic requirements of task by providing a two level task

scheduling mechanism based on load balancing in cloud computing. They paper improve the response time, resources utilization by mapping task to VMs and then VMs to host resources. They use the first level of scheduling (from user's application to the VM) to create a description of VM including the task of computing resources, network resources, storage resources, etc. and used the second level scheduling (from the VM to host resources) to find appropriate resource for VM. This approach may have succeeded in improving the resource utilization, but we think that using two levels of task scheduling would increase the response time compared with other load balancing algorithms"[2].

B. Sethi et al. (2012)

Introduced "a load balancing algorithm using fuzzy logic with Round Robin (RR) algorithm. The algorithm is based on various parameters such as processor speed, and assigned load in VM and etc. The algorithm maintains the information of each VM and numbers of requests currently allocated to VM. When a new request is received, the load balancer searches for the least loaded VM and allocate it, but if there are more than one VM, the selection will be based on processor speed and load in VM using fuzzy logic. This algorithm enhanced the performance of load balancer and decreased the response time. In addition, the results referred that its performance is better than RR algorithm. The drawback of this approach that authors had focused only on how to decrease the response time of job scheduling and they ignored talk about processing cost. In addition, the researchers compared their results with only RR algorithm which had been enhanced and improved by many researchers before" [3].

C. Ratan Mishra et al.(2012)

Described "Load balancing ensures that all the processor in the system or every node in the network does approximately the equal amount of work at any instant of time. Many methods to resolve that problem had been came into existence like Particle Swarm Optimization, hash method; genetic algorithms and several scheduling based algorithms are there. In that paper they are proposed a method based on Ant Colony optimization to resolve the problem of load balancing in cloud environment"[4].

D. Sharma et al. (2013)

Proposed in their paper "a new algorithm to enhance response time of each VM. The proposed algorithm collects information about all VMs in a list and uses it to allocate appropriate VM where status is available. When a new request is received, the load balancer looks at the table and identifies VM whose current allocation count is less than max allocation, and then check its status. The result is returned to the datacenter and then the data center allocates this resource to the request. When the VM is finished, it notifies the datacenter to de-allocate it. The drawback of this algorithm is in some case such as the high workload it may increase the waiting queue because the allocation depends on the available status only"[5].

E. Amandeep Kaur Sidhu et al. (2013)

Discussed "a couple of existing scheduling algorithms can keep up load balancing and give better methodologies through efficient job scheduling and resource allocation methodologies to. Keeping in mind the end goal to increase most extreme benefits with load balancing algorithms, it is important to use resources productively. Their paper examined a portion of the current load balancing algorithms in cloud computing and furthermore their difficulties"[6].

F. Singh et al. (2014)

Developed "a new heterogeneous load balancing algorithm to distribute the load across a number of servers. They create VMs of different datacenters according to host specification including core processor, processing speed, memory, storage etc. Then allocate weight count according to the RAM allocated to the VMs in the datacenter. They use a data structure to maintain weight count and the current allocation count of the VM. They allocate the VM which have available status and have a higher RAM. When allocating a new VM, the algorithm returns the VM id to the Data Center Controller, and then updates the allocation count for that VM and adding the new allocation to the busy list. When the VM finishes processing the request the algorithm de-allocates the VM and removes the VM from the busy list. The main drawback of the algorithm is the authors allocates the VM which have higher RAM specification, but they ignores others specification such as processor power. On other hands they do not present any results and comparison with other algorithms" [7].

G. Singhal and Jain (2014)

Proposed "a load balancing algorithm using Fuzzy Logic, the algorithm focuses on a public cloud. The main idea of the algorithm is partitioning the cloud to several partitions and each partition having its own load balancer, and there is a main controller which manages all these partitions. With the idle partition status they use a fuzzy logic and in the normal partition status they use a global swarm optimization based load balancing strategy. The result shows enhancements in resource utilization and availability in cloud computing environment. The drawback of this approach is the difficulty of testing the technique in a real environment to make sure that it has achieved good results"[8].

H. Abhay Kumar Agarwal et al. (2015)

Proposed an algorithm that we named as a New Static load balancing algorithm in cloud computing. "The proposed algorithm is utilizing the idea of both Active Monitoring Load Balancing Algorithm and Throttled Load Balancing Algorithm. The point by point outline, pseudo code and execution of algorithm are likewise introduced in this paper. The outcomes (Overall Response Time and Datacenter Processing Time) got are contrasted and the consequences of Throttled Load Balancing Algorithm. This correlation is done in the wake of executing and breaking down each of the current algorithms talked about in this paper, and found that Throttled Load Balancing Algorithm is best among all the current algorithms"[9].

I. Bhavisha Patel et al. (2015)

Described that "Cloud Computing is the utilization of processing resources that are conveyed as a service over a network. In cloud computing, there are many undertakings requires to be executed by the accessible resources to accomplish best execution, use the resources proficiently under the state of substantial or light load in the network, limit the response time and delay, for keep up system stability, to enhance the execution, increment the throughput of the system, to diminish the communication overhead and to limit the computation cost. This Paper Describe different load balancing algorithms that can be connected in cloud computing"[10].

J. Mithun Dsouza et al. (2016)

Described that "Cloud computing has become popular due to its attractive features. The load on the cloud is increasing tremendously with the development of new applications. Load balancing is an important part of cloud computing environment which ensures that all devices or processors perform same amount of work in equal amount of time. In this paper we are mentioned about different techniques in load, we aim to provide a structured and comprehensive overview of the research on load balancing algorithms in cloud computing. This paper surveys the state of the art load balancing tools and techniques over the period of 2004-2016"[11].

K. Mamta Khanchi et al. (2016)

Described that "Cloud computing is a computing provided over the internet. The guideline part of cloud computing is virtualization that arrangements with the development and administration of virtual machines productively. As the number of clients and requests for the services are expanding step by step in cloud computing, in this way load balancing is a vital research territory for taking care of the clients' requests effectively. For proficient and viable administration and use of cloud service provider's resources, many load balancing algorithms have been now proposed. This paper proposed and executed a hybrid approach for virtual machine level load balancing"[12].

5. Challenges in Load Balancing

The scientific communities define some scientific challenges which are persisting unsolved mainly load balancing challenges are:

1) Automated Service Provisioning: Cloud computing provide a key feature of elasticity; in which service/resources are allocated or released automatically to the user when it required. How keep the record of resources which are used by same traditional systems which have the same performance and use optimal resource how decide which resource is use or release?

2) Virtual Machines Migration: Due to virtualization each and every VM saw as a file or set of files if a when any PM is greatly overloaded and want to unload for this unloading move VM to various physical machines. A problem is arising how we can distribute the load dynamically when moving the virtual machine to physical machines. How avoid blocks in Cloud computing systems?

3) Energy Management: The adoption economy of scale is the main profits in the cloud system. For allowing global economy energy saving is a basic point which allows how reduced providers will support set of global resources that each process as sits own resources. So each has keeping acceptable performance. How can we use a part of datacenter?

4) Stored Data Administration: Management of stored data in a n/w or thorough a n/w which increase exponential establishments that provide the source to the individuals which manage own data at various years points. The main challenge in cloud computing is how to manage the management of data storage. So when a process wants to maintain access very fast. How optimum storage data can is distributed in the cloud system.

5) Manifestation of Small Data Centers for Cloud Computing: Small data centers are useful than large data center because it has various issue such as more energy consumption and expensive. Small data centers are best it provides the best diversity computing with less energy consumes. The main problem occur at large scale in cloud computing with optimal distribution with sufficient response time.

6. Conclusions

In the cloud storage, Load balancing is one of the central issues in cloud computing. It is a mechanism that disseminates the dynamic nearby workload equitably over every node in the entire cloud to avoid a circumstance where few nodes are intensely loaded while others are idle or doing little work. It accomplishes a high client fulfilment and resource usage proportion, thus enhancing the overall performance and resource utilization of the system. It likewise guarantees that each processing resource is distributed effectively and decently. This paper provided a literature review of various load balancing algorithms in cloud computing.

References

[1] Dharmesh Kashyap, Jaydeep Viradiya, "A Survey Of Various Load Balancing Algorithms In Cloud Computing", International Journal of Scientific & Technology Research Volume 3, Issue 11, November 2014.
[2] Fang, Y., Wang, F., and Ge, J., "A Task Scheduling Algorithm Based on Load Balancing in Cloud Computing", Lecture Notes in Computer Science, Jg. 2010(6318): pp. 271-277, (2010).
[3] Sethi, S., Anupama, S., and Jena, K., S, "Efficient load Balancing in Cloud Computing using Fuzzy Logic", IOSR Journal of Engineering (IOSRJEN), 2(7): pp. PP 65-71, (2012).
[4] Ratan Mishra and Anant Jaiswa "Ant colony Optimization: A Solution of Load balancing in Cloud",

International Journal of Web &Semantic Technology (IJWesT) Vol.3, No.2, April 2012.

[5] Sharma, T. and Banga, V. K., "Proposed Efficient and Enhanced Algorithm in Cloud Computing", International Journal of Engineering Research & Technology (IJERT), 2(2), (2013).

[6] Amandeep Kaur Sidhu, Supriya Kinger, "Analysis of Load Balancing Techniques in Cloud Computing", International Journal of Computers &Technology Volume 4 No. 2, March-April, 2013, ISSN 2277-3061.

[7] Singh, A., Bedi, R., and Gupta, S., "Design and implementation of an Efficient Scheduling algorithm for load balancing in Cloud Computing", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 3(1), (2014).

[8] Singhal, U. and Jain, S., "A New Fuzzy Logic and GSO based Load balancing Mechanism for Public Cloud", International Journal of Grid Distribution Computing, 7(5): pp. 97-110, (2014).

[9] Abhay Kumar Agarwal, Atul Raj, "A New Static Load Balancing Algorithm in Cloud Computing", International Journal of Computer Applications (0975 – 8887) Volume 132 – No. 2, December 2015.

[10] Bhavisha Patel, Mr. Shreyas Patel, "Various Load Balancing Algorithms in cloud Computing", IJARIIE-ISSN(O)-2395-4396, Vol-1 Issue-2 2015.

[11] Mithun Dsouza, Mohammed Rizwan, Ramnath Gaonkar and Dr. S. Sathyanarayana, "Scheduling and Load Balancing Techniques in Cloud Computing: A Survey", International Journal of Latest Trends in Engineering and Technology Special Issue SACAIM 2016, pp. 309-316

[12] Mamta Khanchi, Sanjay Tyagi, "An Efficient Algorithm for Load Balancing in Cloud Computing International Journal of Engineering Sciences & Research Technology", June, 2016