Available online at: http://www.ijmtst.com/vol5issue03.html

International Journal for Modern Trends in Science and Technology

ISSN: 2455-3778 :: Volume: 05, Issue No: 03, March 2019

A review on load balancing algorithms in cloud computing

Amanpreet Kaur | Amreen Kaur | Anupam Garg

Assistant Professor ,Computer Science and Engineering Department, BGIET Sangrur, Punjab, India.

To Cite this Article

Amanpreet Kaur, Amreen Kaur and Anupam Garg, "A review on load balancing algorithms in cloud computing", *International Journal for Modern Trends in Science and Technology*, Vol. 05, Issue 03, March 2019, pp.-25-30.

DOI: https://doi.org/10.46501/IJMTST050328

Article Info

Received on 02-March-2019, Revised on 21-March-2019, Accepted on 28-March-2019.

ABSTRACT

Cloud computing is the method for figuring, by means of the web that shares PC assets as opposed to utilizing programming or capacity on a nearby PC. It stores the information and assets in the open condition. So now daily's measure of information stockpiling increment rapidly. load Balancing is the primary issues in Cloud which is required to disseminate the dynamic remaining task at hand over numerous hubs to guarantee that no single hub is overpowered. load Balancing gives appropriate use of assets and upgrading the execution of the framework. The few existing calculations that can give load adjusting and furthermore give better systems through productive employment booking and asset planning methods also. So as to pick up boost the benefit and adjusting calculations, it is important to use assets effectively. This paper talks about a portion of the current load adjusting calculations in distributed computing.

Keywords: *cloud, client, load, PC, internet etc.*

I. INTRODUCTION

Cloud computing is an on interest administration in which shared assets, data, programming and different gadgets are given by the customers necessity at specific time. It's a term which is commonly utilized in the event of Internet. The entire Internet can be see as a cloud. Resources and operational costs can be cut utilizing distributed computing.

There is no standard meaning of Cloud registering. For the most part it comprises of a gathering of circulated servers known as experts, giving interest administrations and assets to various customers known as customers in a system with adaptability and dependability of server farm. The dispersed PCs give on-request benefits. Administrations might be of programming assets (for example Programming as a Service) or physical assets or equipment/foundation (for example Equipment as a Service or Infrastructure as a Service). Amazon EC2 (Amazon Elastic Compute Cloud) is a case of distributed computing administrations. In the examination work diverse calculations are utilized to keep up the heap that is talked about in this work.

Cloud Components

A Cloud system consists of 3 major components such as clients, data centre and distributed servers. Each element has

a definite purpose and plays a specific role.material.

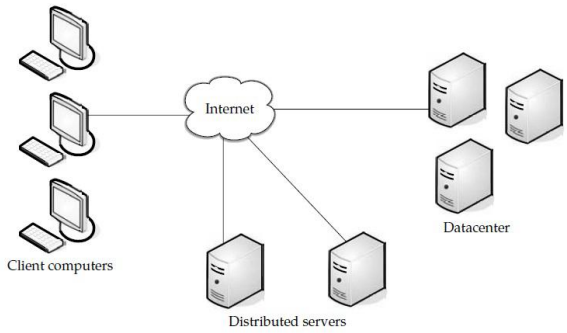

Figure 1: Three components make up a cloud computing solution

(a)Clients

End clients interrelate with the customers to oversee data identified with the cloud. Customers for the most part drop into three classifications as given in [1]:

•Mobile: Windows Smartphone, advanced cells like a Blackberry or I telephone.

•Thin: They don't do any calculation work. They just present the data. Servers do all the instrument for them. Slender customers don't have any inward memory.

•Thick: These apply diverse programs like IE or Mozilla Firefox or Google Chrome to interface with the Internet cloud.

Presently a-days flimsy customers are increasingly prominent as contrast with different customers in view of their low cost, security, low utilization of intensity, less commotion, effectively replaceable and repairable and so on.

(b) Data Center

Server farm is only a gathering of servers facilitating diverse applications. End clients associate with the server farm to buy in various applications. A server farm may exist at an extensive separation from the customers.

(c)Distributed Servers

Distributed servers are the parts of a cloud which are present throughout the Internet hosting different applications. But using the application from the cloud, the user will be aware of that he is using this application from its own machine.

Type of Clouds

Based on the sphere of influence or environment in which clouds are used, clouds can be divided into 3 types:

_ Public Clouds

_ Private Clouds

_ Hybrid Clouds (combination of both private and public clouds)

-Community Clouds

Services provided by Cloud computing

Service means different types of applications provided by different servers across the cloud. It is generally given as “as a service”. Services in a cloud are of 4 types as given in [1] :

_ Anything as a Service (XaaS)

_ Software as a Service (SaaS)

_ Platform as a Service (PaaS)

_ Hardware as a Service (HaaS) or Infrastructure as a Service (IaaS)

•Anything as a Service(XaaS)

It is an aggregate term said to represent various things including "X as an administration", and furthermore called everything as an administration. The short structure alludes to an expanding number of administrations that are conveyed over the Internet as opposed to gave locally or on location. The most widely recognized models are Software as a Service, Infrastructure as an administration and Platform as an administration. The join utilization of these three is some of the time alluded to as the SPI demonstrate.

• Software as a Service (SaaS)

In software as a service, the user use different software applications from different servers through the Internet. The user uses the software as it is without any change and do not need to make lots of changes or don’t require integration to other systems.

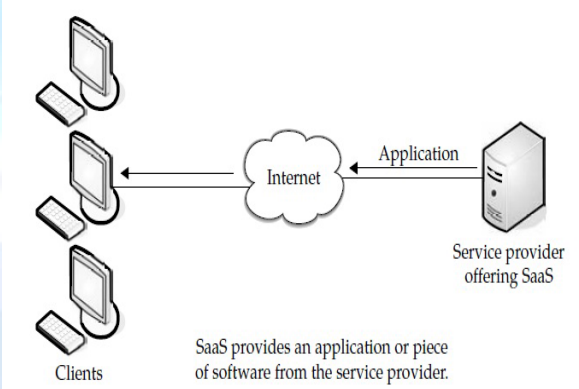

Figure 2: Software as a service

The client will have to pay for the time he use the software. The software that does a clear-cut task without any need to interact with other systems makes it an ideal candidate for Software as a Service.

Some of these applications include:

_ Customer resource management (CRM)

_ Video conferencing

_ IT service management

_ Accounting

_ Web analytics

_ Web content management

•Platform as a Service (PaaS)

Stage as an administration gives every one of the assets that are required for building applications and administrations totally from the Internet, without download or introduce programming. It's administrations are programming plan, improvement, testing, utilization, and facilitating. Different administrations can be group participation, database incorporation, web administration reconciliation, information security, storage room and forming and so forth.

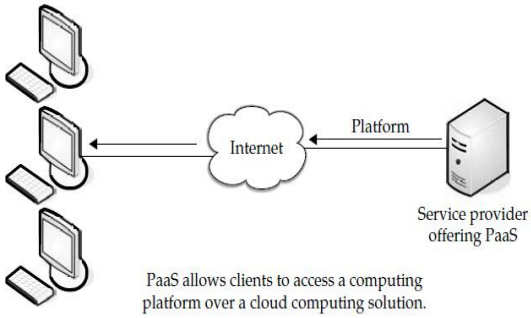

Figure 3: Platform as a service

•Hardware as a Service (HaaS)

It is also known as Infrastructure as a Service. It offers the hardware as a service to an organisation so that it can put anything into the hardware according to its will [1].

It allows the user to "rent" resources (taken from [1]) as

_ Server space
_ Network equipment
_ Memory
_ CPU cycles
_ Storage space

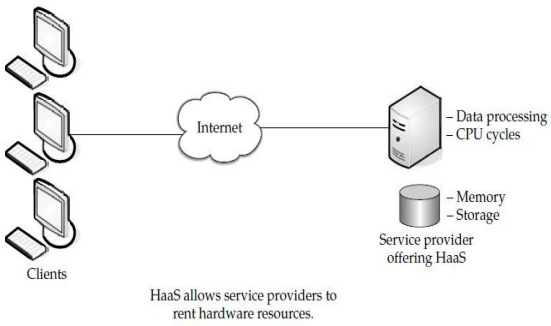

Figure 4: Hardware as a service

Cloud computing provides a Service Oriented Architecture (SOA) and Internet of Services (IoS) type applications, excluding fault tolerance, high scalability, accessibility flexibility, reduced information technology overhead for the user concentrated cost of ownership, on demand services etc. Middle to these issues lies the establishment of an efficient load balancing algorithm.

II. LOAD BALANCING IN CLOUD

Load Balancing is a PC organizing strategy to disperse remaining task at hand over different PCs or a PC pack, net connections, focal handling units, plate drives and different assets, to get ideal asset utilization, misuse throughput, diminish reaction time, and maintain a strategic distance from over-load. Utilizing different system with load adjusting, rather than a solitary module, may expand consistency through laying-off. The heap adjusting administration is commonly given by committed programming or equipment, for example, a multilayer switch or a Domain Name System server. Load adjusting is one of the focal issues in distributed computing [2]. It is an instrument that circulates the dynamic nearby outstanding task at hand uniformly over every one of the hubs in the entire cloud to stay away from a circumstance where a few hubs are vigorously stacked while others are inactive or doing little work. It finishes a high client satisfaction and asset use proportion, subsequently recouping the general execution and asset utility of the framework. It additionally guarantee that each registering asset is circulated productively and decently [3]. It further forestalls bottlenecks of the framework which may happen because of load disparity. When at least one segments of any administration come up short, load adjusting helps in support of the administration by execute reasonable over, for example in provisioning and de-provisioning of examples of uses no matter what. The reason for load adjusting is enhancing the execution by adjusting the heap among these different assets (organize joins, focal handling units, circle drives) to accomplish ideal asset utilization, greatest throughput, most extreme reaction time, and stay away from over-load. To appropriate load on various frameworks, distinctive kinds of load adjusting calculations are utilized.

When all is said in done, load adjusting calculations pursue two noteworthy groupings:

•Depending on how the charge is dispersed and how forms are assigned to hubs (the framework load);

•Depending on the data position of the hubs (System Topology).

In the principal case it planned as a unified methodology, conveyed methodology or half breed approach and in the second case as static methodology, dynamic or versatile methodology.

a) Classification According to the System Load

•Centralized approach: In this only one node is responsible for managing the distribution within the whole system.
•Distributed approach: In this each node independently builds its own load vector by collecting the load information of other nodes. Decisions are made using local load vectors. This approach is more suitable for generally distributed systems such as cloud computing.
•Mixed approach: A combination of above two approaches to take advantage of each approach.

b) Classification According to the System Topology

•Static approach: This approach is generally defined in the design or implementation of the system.
•Dynamic approach: This approach takes into account the current state of the system during load balancing decisions. This approach is more suitable for widely distributed systems such as cloud computing.
•Adaptive approach: This approach adapts the load distribution to system status changes, by changing their parameters dynamically and even their algorithms.

There are different types of load balancing algorithms that discussed in this paper.

III. LITERATURE SURVEY

Saeed javanmardi et al. [2018] In this paper with the aid of genetic algorithm and fuzzy theory, present a hybrid job scheduling approach, which consider the load balancing of the system and reduces total execution time and execution cost. The main goal of this research is to assign the jobs to the resources with considering the VM MIPS and time-span of jobs. The new algorithm assigns the jobs to the resources with considering the job length and resources capacities. Evaluate the performance of the approach with some famous cloud scheduling models. The result of the experiments shows the efficiency of the proposed approach in term of execution time, execution cost and average degree of imbalance [4].

Hitesh A. Ravani et al. [2017] this paper discusses that Resource Scheduling is the process of mapping tasks to available resources on the basis of tasks characteristics and requirements. The received tasks are group on the basis of data and resources. Resource selection is done on the basis of its cost and turnaround times both using greedy approach and task selection on the basis of a priority. This way of resource selection and task selection gives better results over sequential scheduling. The available resources should be utilized efficiently without affection the service parameters of cloud. Main aim of this paper is to analyze the various scheduling algorithm and manage the resources which are precisely available at certain fixed times and for fifed intervals of time. Find the optimizes scheduling algorithm for resource so the cloud provider get benefits in term of efficient resource management which provide more resources to allocate without postponing or declining any user requests. Cloud users also get benefits in term of their monetary gains at each front [5].

Florin Pop et al. [2017] in this paper evolutionary computing offers different methods to solve NP-hard problems, finding a near-optimal solution. Task scheduling is a composite problem for large environments like Clouds. Genetic algorithms are a superior method to find a solution for this problem considering multi-criteria constrains. This is also a method used for optimization. In these types of environments service provider want to increase the profit and the customers (end-users) want to minimize the costs. So, it's all about money and minimum two optimization constrains. On the other hand, a good performance to ensure the QoS is to use the reputation of resources offered. This aspect is very important for service providers because represents a ranking method for them. In this paper a reputation guided genetic scheduling algorithm for independent tasks in inter-Clouds environments. The characters is considered in the selection phase of genetic algorithm as evolutionary criteria for the algorithm and evaluate the proposed solution considering load-balancing as a way to measure the optimization impact for providers and maxspan as a metric for user performance [6].

Jianfeng Zhao et al. [2016] it's a basic requirement in cloud computing that scheduling virtual resources to physical resources with balance load. The simple scheduling methods can not meet this requirement. This paper proposed a virtual resources scheduling model and solved it by advanced Non-dominated Sorting Genetic Algorithm II (NSGA II). This model was evaluated by balance load, virtual resources and physical resources were abstracted a lot of nodes with attributes based on analyzing the flow of virtual resources scheduling. NSGA II was engaged to address this model and a new tree sorting algorithms was adopted to improve the efficiency of NSGA II. In experiment, verified the correctness of this model. Comparing with Random algorithm,

Static algorithm and Rank algorithm by a lot of experiments, at least 1.06 and at most 40.25 speed-up of balance degree can be obtained by NSGA II [7].

Lucio Agostinho [2016] In cloud computing the allocation and scheduling of multiple virtual resources, such as virtual machines (VMs), are still a challenge. The optimization of these processes brings the advantage of improving the energy savings and load balancing in large datacenters. Resource allocation and scheduling also impact in federated clouds where resources can be leased from partner domains. This paper proposes a bio-inspired VM allocation method based on Genetic Algorithms to optimize the VM distribution across federated cloud domains. The main contribution of this work is an inter-domain allocation algorithm that takes into account the capacity of the links connecting the domains in order to avoid quality of service degradation for VMs allocated on partner domains. Architecture to replicate federated clouds is also a contribution of this paper [8].

Andrew J.Younge et al. [2015] This paper represent the notion of Cloud computing has not only reshaped the field of distributed systems but also fundamentally changed how businesses utilize computing today. While cloud computing provides many advanced features, it still has some shortcoming such as the relatively high operating cost for both public and private clouds. The area of Green computing is also becoming increasingly important in a world with limited energy resources and an ever-rising demand for more computational power. In this paper a new framework is accessible that provides efficient green enhancements within a scalable cloud computing architecture. Using power sensitive scheduling techniques, variable resource management, live migration, and a minimal virtual machine design, overall system efficiency will be vastly improved in a data center based cloud with minimal performance overhead [9].

IV. CONCLUSION

In this paper the current conventions are not fitting to those WSNs that are sent in huge areas since it utilizes single jump steering where every sensor hub can convey specifically to the CH and the BS. In this way, it causes issues of vitality lopsidedness. In this paper in various scientists investigate work is surveyed and distinctive issue are confronted.In Cloud computing there are many existing issues like Load Balancing, virtual machine migration, Energy management etc, which have not been fully addressed. Central of these issues the main issue is load balancing, that is required to distribute the dynamic local workload to all the nodes in the whole cloud to achieve a higher satisfaction and resource utilization ratio. This paper presents a concept of load balancing and its algorithms.

REFERENCES

[1] Anthony T.Velte, Toby J.Velte, Robert Elsenpeter, " Cloud Computing A Practical Approach", TATA McGRAW-HILL Edition 2015, pp.334

[2] B. P. Rima, E. Choi, and I. Lumb, "A Taxonomy and Survey of Cloud Computing Systems", Proceedings of 5th IEEE International Joint Conference on INC, IMS and IDC, Seoul, Korea, August 2009, pp. 44-51

[3] A. M. Alakeel, "A Guide to dynamic Load balancing in Distributed Computer Systems", International Journal of Computer Science and Network Security (IJCSNS), Vol. 10, No. 6, June 2015, pp. 153-160.

[4] Saeed Javanmardi, MohammadShojafar, Danilo Amendola, Nicola Cordeschi, " Hybrid Job scheduling Algorithm for Cloud computing Environment", adfa, 2016, pp.2

[5] Hitesh A. Ravani, Hitesh A. Bheda, "Genetic Algorithm Based Resource Scheduling Technique in Cloud Computing", International Journal of Advance Research of Computer Science and Management Studies, Volume 1,Issue 7,Fecember 2016,ISSN:2321-7782

[6] Florin Pop, Valentin Cristea, Nik Bbbessis, Stelios Sotiriadis, "Reputation guided Genetic Scheduling Algorithm for Independent Tasks in Inter-Clouds Environments", International Conference on Advanced Information Networking and Applications Workshops, 2016, pp. 772-776

[7] Jianfeng Zhao, Wenhua Zena, Min Liu, Guangming Li, "Multi-Objective Optimization Model of Virtual Resources Scheduling under Cloud Computing and It's Solution", International Conference on Cloud and Service Computing, Grant no 60903129, 2016

[8] Lucio Agostinho, Guilherme Feliciano, Leonardo Olivi, Eleri Cardozo, "A Bio-Inspired Approach to Provisioning of Virtual Resources in Federated Clouds", IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing, 2016, pp. 548-604

[9] Andrew J. Younge, Gregor von Laszewski, Lizhe Wang, Sonia Lopez-Alarcon, warren Carithers, "Efficient Resource Management for Cloud Computing Environments", IEEE, 2015, pp.

[10] Yu-lung Lo, Min-Shan Lai , "The Load Balancing of Data base Allocation in the Cloud", International MultiConference of Engineers and Computer Scientists, Volume I, March 13-15,2016, ISSN: 2078-0958

[11] D.T. Pham, A. Ghanbarzadeh, E. Koc, S. Otri, S. Rahim, M. Zaidi, "The Bees Algorithm-A Novel Tool for Complex Optimisation Problem", Cardiff CF243AA UK, 2015

[12] Er. Shimpy, Mr. Jagandeep Sidhu, "Different Scheduling Algorithms In Different Cloud Environment", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 9, September 2018, ISSN: 2278-1021

[13] Binitha S, S Siva Sathya, "A Survey of Bio inspired Optimization Algorithms", International Journal of Soft Computing and Engineering, Vol. 2, Issue 2, May 2017, ISSN: 2231-2307

[14] Omid Bozorg Haddad, Abbas Afshar, Miguel A. Marino, "Honey Bees Mating Optimization (HBMO) Algorithm", Water Resource Management, DOI: 10.1007/s11269-005-9001-3, 2006, pp. 661-680