REVIEW ON LOAD BALANCING ALGORITHMS IN CLOUD COMPUTING

[1]Mona Daharwal and [2]Mayank Namdev

[1]M.Tech Scholar, Sagar Institute of Research and Technology Bhopal

[2]Assistant Professor, Sagar Institute of Research and Technology Bhopal

Abstract: Cloud computing is a rising innovation in the present and the days to desire giving diverse sort of administrations to the end clients. It concentrates predominantly the dynamic administrations utilizing enormous versatile and virtualized assets over the Internet. The designation of the gathering of assets may begin an issue of accessibility of these assets and conveying the remaining load of all the VMs among themselves called as burden adjusting. Burden adjusting is a difficult wonder in distributed computing. It guarantees that no single hub is over-burden by circulating the outstanding task at hand among the hubs appropriately. This is done to improve the aftereffects of the framework. Because of oddity of distributed computing field, there is no numerous standard loads balancing algorithms utilized in cloud condition. Subsequently productive usage of assets must be significant and for that heap adjusting assumes a significant job to get most extreme profit by the assets. Having comprehended the crucial job of load balancing, this paper covers different load balancing algorithms identified with cloud computing and shows a similar report on load balancing algorithms.

Keywords: Cloud Computing, Load balancing, response time, cloud computing, processing time, cloud analyst,

I. INTRODUCTION

Cloud computing is an on-request pay innovation where the payment is made according to the use of the administration. Cloud computing has been one of the present apparatuses grasped by both industry and scholarly. It gives a vigor and adequacy to store and recover the information. The object is to accomplish high asset use and decrease accordingly time while guaranteeing the most extreme use of assets. It transmits all prerequisites through the web progressively when client demands, for example, working framework, organize, capacity, programming, equipment and assets. The administrations are basically partitioned into to three kinds to be specific SaaS, PaaS and IaaS. Further, In distributed computing Everything as a Service (Xaas) takes this innovation to people to come. The Cloud computing is isolated into three gatherings, for example, Public, Private and Hybrid cloud. Open Cloud: A cloud is known as an "open cloud" when the administrations are open for open use. Private Cloud: This cloud arrangement purposes exclusively for a solitary endeavor, regardless of whether took care of inside or by an outsider, and facilitated either inside or remotely. Crossover Cloud: It is a blend of both open just as private cloud arrangement. [1]

This paper has been coordinated as pursues. Segment I gives presentation on distributed computing. Area II expresses the heap adjusting system in distributed computing. Segment III represents the writing study on significant burden adjusting calculations. Segment IV portrays execution examination of burden adjusting calculations. Segment V gives the end.

II. LOAD BALANCING MECHANISM IN CLOUD COMPUTING

Cloud balancing Mechanism is a segment in cloud OS that conveys high asset time and powerful asset utilization by Designating the absolute burden among the different cloud hubs, side by side it tackles the issue of misuse of virtual machines. Cloud balancing Mechanism gives answer for the issue of overburdening and spotlights on higher throughput, boosting asset usage and decreasing the reaction time. It is the essential for boosting the cloud execution and using the assets effectively. The usage of mists has been upgraded by an asset portion strategy which has pre-emptible assignment execution. The heap adjusting is a proficient and basic idea in distributed computing and it uses the assets ideally, along these lines limiting the use of assets. Subsequently burden will be appropriated over the hubs in cloud-based Model, with the goal that every asset does likewise measure of work anytime of time which is designated by a heap balancer. The heap balancer administers the different solicitation assignment to various servers. The heap balancer utilizes different calculation to manage the server which needs to deal with the demand.[1] The proposes load adjusting calculations are to accomplish high throughput and limiting the reaction time.

III- RELATED WORKS

Round Robin Algorithm

It is a static load balancing algorithm, which does not take into account the previous load state of a node at the time of allocating jobs. It uses the round robin scheduling algorithm for allocating jobs. It selects the first node randomly and then, allocates jobs to all other nodes in a round robin manner. This algorithm will not be suitable for cloud computing because some nodes might be heavily loaded and some are not. Since the running time of any process is not known prior to execution, there is a possibility that nodes may get heavily loaded. Hence, weighted round-robin algorithm [3] was proposed to solve this problem. In this algorithm, each node is assigned a specific weight. Depending on the weight assigned to the node, it would receive appropriate number of requests. If the weights assigned to all the nodes are equal, then each node will receive same traffic. In cloud computing system, precise prediction of execution time is not possible therefore, this algorithm is not preferred.

Throttled Load Balancing Algorithm

Throttled load balancer [2] is a dynamic load balancing algorithm. In this algorithm, the client first requests the load

balancer to find a suitable Virtual machine to perform the required operation. In Cloud computing, there may be multiple instances of virtual machine. These virtual machines can be grouped based on the type of requests they can handle. Whenever a client sends a request, the load balancer will first look for that group, which can handle this request and allocate the process to the lightly loaded instance of that group.

Equally spread current execution algorithm

Equally Spread current execution [2] is a dynamic load balancing algorithm, which handles the process with priority. It determines the priority by checking the size of the process. This algorithm distributes the load randomly by first checking the size of the process and then transferring the load to a Virtual Machine, which is lightly loaded. The load balancer spreads the load on to different nodes, and hence, it is known as spread spectrum technique.

An enhanced priority based HTV load balancing algorithm

This algorithm accomplishes an effective and trustworthy resource distribution of the tasks on the servers in data centers. This algorithm focuses on three parameters such as server load, stats of server and time constraint of the tasks. This algorithm calculated the load and performance factor of each virtual machine and then allocates the arriving task to various virtual machines according to their completion time and stand-by time to rise the throughput and performance. [3]

Virtual Machine and Physical Machine Categorization Algorithm

In general, load balancing approach that amplify the physical resource utilization and curtail the energy consumption. To calculate the performance of the approach it is compared with the existing load balancing approach and judged against the number of migration and energy consumption. Experiment results say that this approach gives better result while it is compared with the existing load balancing methods. [4]

Composite procedure

Load Balancing is required to appropriately undertake the resources of the service contributor. Load balancer is a component of cloud fabric to dispense the workload among many virtual machines in a Server over the network to accomplish optimal resource consumption, reduction in data processing time, and decline in average response time, and eliminate the overload. This composite approach is effectively utilized to handle the load distribution techniques very well. This method makes use of the advent of Equally Spread Current Execution (ESCE) and Throttled algorithms. [5]

Distributed dynamic priority based algorithm

It is used for balancing the load on instances effectively, improves the system consistency, minimizes response time and increases the throughput. Allocating the resources on virtual machines based on priority achieves the better response time and processing time. Load balancing guarantees all occurrences in a node in the networks to do the identical amount of work at any instant of time. Priority based resource provision to improve the utilization of resources and reducing response time of cloud services. [6]

Procedure : Burstness-aware load balancing

This procedure takes two load balancing algorithms. In which, RR is used for burst and Random is used in non-burst state. In order to handle the user request, Fuzzy logic based validation is used and the request is forward to a well-balanced VM. Experimental results demonstrate that the algorithm advances the average response time and average processing time in contrast with other algorithms. [7]

Honey Behavior Load Balancing method

The high priority tasks are detached from overloaded virtual machine and they are fixed to under loaded virtual machine by bearing in mind least numbers of similar priorities to those tasks, cost effective virtual machine, and least probable completion time which also balances the loads of reliant on tasks in pre-emptive manner. The minimum expected finishing time, cost and priority at submission time of that task aids to yield minimum completion time, declines waiting time of the tasks and eventually reaches better resource utilization. [8]

Procedure: Weighted based optimized load balancing

The loads are distributed among the servers which spawns VMs by considering weight as a parameter. The results are compared with existing Round Robin and EIPR algorithms. Simulation results have affirmed that this algorithm has disseminated the work load evenly among virtual machines. [9]

New enhanced load balancing procedure

Firstly, an adaptive strategy has been developed for load balancing according to the quality of the solutions discovered by Genetic. Secondly, the improved load balancing strategy takes other constraints like fitness and the choice of the initial resource pool. These constraints offer the significant impact on the result of the algorithm. This enhanced load balancing algorithm provides the better result than the existing genetic approach. [10]

Tow –level global load balancing Framework

A framework existing to represent the global server load balancing of the Web sites in a cloud with the help of two-level load balancing architecture. This framework depicts the load balancing mechanism on different data centers managed by cloud service providers.[11] Further, the method of handling the request with or without priority under the give work load also shown as client request and load balancer.

Procedure: Dynamic load balancing

Load balancing algorithms play important role in equalizing load among data centers and in efficient use of computing resources. In this paper, performance of a dynamic load balancing algorithm has been evaluated by dividing data-

centers in different zones. It has been shown that the proposed algorithm improves the computing efficiency of data-centers and minimizes the response time of user's applications. [12]

IV- PERFORMANCE ANALYSIS OF LOAD BALANCING ALGORITHMS

The forthcoming part discusses various load balancing algorithms and shows the results comparatively.

Min- min LB procedure

Min-Min is a basic, simple and fastest algorithm which is capable of giving better performance. Min-Min uses the ideal tasks at initial stage, which results in best and attains the progress of the overall, make span. The assignment of small task first is considered as an issue. This method gives focus to minimum completion tasks rather than longer completion time. This phenomenon may lead to underutilization of resources. Min-Min displays minimum completion time for jobs which are not allocated and later depute the jobs with minimum completion time (hence min-min) to a node that is talented to manage the task. [13]

Max-min LB procedure

At first for all the existing tasks are produced to the system and minimum accomplishment time for all of them are calculated, then among these tasks the one which is having the completion time, the maximum is selected and that is assigned to the corresponding machine. If the task is a single long time consuming task then, Max-Min algorithm runs small tasks parallel along with the long time consuming task. It further considers the task with maximum completion time whereas Min-Min considers the minimum completion time at first. This algorithm performance is affected due to starvation where the tasks are having the extreme finishing time will get performed first while leaving behind the tasks having the minimum completion time. [13]

RASA procedure

Resource Aware Scheduling Algorithm takes the results of Min-Min and Max-Min algorithm alternatively in order to give improved results. The resource finishing time is measured followed by Min-Min and Max –Min algorithms are utilized alternatively. [13]

Minimum Make span procedure

This algorithm first verifies the completion of all tasks with minimum makes span and it relates with the task and migrated. The two tasks give the same make span time it selects the node with higher performance resources. [13]

PA-LBIMM Algorithm

PA-LBIMM priority aware load balancing improved min min algorithm separate the tasks into G1 and G2groups. The tasks submitted by VIP user's or high priority users are considered as group G1 and tasks submitted by low priority users are considered as group G2. Tasks are planned to the resources on the basis of priority. At first, for all the tasks in G1, each task is allocated to the VIP category resource with the help Min- Min, and then each task in G2 group is given to all the resources by using Min-Min. [14]

RPA-LBIMM Algorithm

This algorithm RPA-LBIMM recovery priority aware load balancing improved min-min algorithm is also similar to the priority aware load balancing improved min-min algorithm. Here use the recovery policy which helps the cloud scheduler to reschedule the tasks if a resource fails at the time of execution to achieve the minimum make span. According to this policy, First of all, scheduler looks for the failed resource. All the tasks that were scheduled by PA-LBIMM to execute on will be considered as a task set. [14]

TABLE 1. Cloud Computing Load balancing Mechanisms

S. No.	Load balancing alg.	Parameters Used
1	**Min Min**	Make span, Resource utilization
2	**MaxMin**	Maekspan, Resource utilization
3	**RASA**	Maekspan , Resource utilization
4	**Minimum Makespan**	Maekspan , Resource utilization
5	**PA-LBIMM**	Maekspan
6	**RPA-LBIMM**	Maekspan

Table 1 lists out the various load balancing algorithms and the parameters used in the algorithms. All these algorithms use the parameters completion time and resources utilization in cloud environment. These algorithms distribute the load based on the number of user request and number of available resource in cloud and also considering the proper utilization of resource.

TABLE 2. Task Completion Time makespoan (sec.) of the given algorithms

S. No.	Load balancing alg.	Completion time Makespan (sec.)
1	**Min Min**	30.35
2	**MaxMin**	20.03
3	**RASA**	11.45
4	**Minimum Makespan**	10.8
5	**PA-LBIMM**	38.37
6	**RPA-LBIMM**	35

Table 2 represents the completion time of the task based on the number of nodes in the simulation area. The makespan of the different algorithms are given.

TABLE 3. Utilization of resources by various algorithms

S. No.	Load balancing alg.	Resource utilization (%)
1	**Min Min**	43.09
2	**Max Min**	79.32
3	**RASA**	88.72
4	**Minimum Make span**	91.01
5	**PA-LBIMM**	--
6	**RPA-LBIMM**	--

Table 3 represents the resource utilization of various load balancing algorithms. The average resource utilization of the different algorithms is given.

V - CONCLUSION

This study deals with various load balancing algorithms. The existing algorithms are static, dynamic, composite, and prioritized. The ultimate purpose of those algorithms is to reduce the response time and maximize the resource utilization. The results of the previous algorithms are limited to give improved result. Still there is plenty of space to improve the results to extract best service from cloud service providers. This study also shows the comparative results of the existing load balancing algorithms based on the parameters such as make span and resource utilization.

REFERENCES

[1] Reena Panwar Bhawna Mallick, "A Comparative Study of Load Balancing Algorithms in Cloud Computing", International Journal of Computer Applications (0975 – 8887) Volume 117 – No. 24, May 2015.

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

[3] Divya Garg, Urvashi Saxena," Dynamic Queue Based Enhanced HTV Dynamic Load Balancing Algorithm in Cloud Computing", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 5, Issue 1, January 2016.

[4] Sachin Soni, Praveen Yadav, "A Load Balancing Approach to Minimize the Resource Wastage in Cloud Computing", International Advanced Research Journal in Science, Engineering and Technology, Vol. 3, Issue 3, March 2016.

[5] Navtej Singh Ghumman, Rajesh Sachdeva," an efficient approach for load balancing in cloud computing using composite techniques", International Journal of Research in Engineering and Applied Sciences, volume 6, issue 2 February, 2016.

[6] G.Suryadevi, D.Vijayakumar, R.SabariMuthuKumar, Dr. K .G. Srinivasagan, "An Efficient Distributed Dynamic Load Balancing Algorithm for Private Cloud Environment", International Journal of Innovative Research in Science, Engineering and Technology, Volume 3, Special Issue 3, March 2014.

[7] Sally F. Issawi, Alaa Al Halees, Alaa Al Halees, "An Efficient Adaptive Load Balancing Algorithm for Cloud Computing Under Bursty Workloads", Engineering, Technology & Applied Science Research Vol. 5, No. 3, 2015.

[8] Khushbu Zalavadiya,Dinesh Vaghela , "Honey Bee Behavior Load Balancing of Tasks in Cloud Computing", International Journal of Computer Applications (0975 – 8887), Volume 139 – No.1, April 2016.

[9] B. Nithya Nandhalakshmi , Mahalingam, "Efficient Load Balancing in Cloud Computing Using Weighted Throttled Algorithm", International Journal of Innovative Research in Computer and Communication Engineering ,Vol. 3, Issue 6, June 2015.

[10] Er. Pooja Er. Vivek Thapar,"An Enhanced Virtual Machine Load Balancing Algorithm for Cloud Environment", , International Journal of Emerging Research in Management &Technology, ISSN: 2278-9359 (Volume-5, Issue-5),May 2016.

[11] Po-Huei Liang1 and Jiann-Min Yang. "Evaluation of two-level Global Load Balancing Framework in cloud environment" International Journal of Computer Science & Information Technology (IJCSIT) Vol 7, No 2, April 2015.

[12] Navtej Singh Ghumman, Rajesh Sachdeva," an efficient approach for load balancing in cloud computing using composite techniques", International Journal of Research in Engineering and Applied Sciences, volume 6, issue 2 February, 2016.

[13] Er. Rajeev Mangla, Er. Harpreet Singh," Recovery and user priority based load balancing in cloud computing", International Journal of Engineering and Science and Research, February 2015.

[14] Harish Chandra, Pradeep Semwal, Sandeep Chopra," load balancing in cloud computing using a novel minimum makespan algorithm", International Journal of Advanced Research in Computer Engineering & Technology ,Volume 5, Issue 4, April 2016.