A SURVEY ON CLOUD COMPUTING LOAD BALANCING

Sachin¹ and Dr rahul Kumar Mishra²
¹,²IFTM University, Moradabad

Abstract- Load Balancing is the standout amongst the most noteworthy parts in circulated situations. As Cloud Computing is outstanding amongst other stage that gives stockpiling of information in extremely insignificant cost and available forever finished the internet, load balancing for the cloud computing has transformed into an exceptionally intriguing and imperative investigation region. Load balancing backings to get a high client fulfillment and utilization of asset proportion by ensuring a capable and sensible portion of each computing resource. There are various challenges in the load balancing systems such as security, adaptation to internal failure and so on in cloudcomputing environments. Many researchers have been proposed a few techniques to improve the load balancing. This paper depicts a diagram on load balancing plans in cloud environments. We investigate the differing sorts of algorithms that is proposed by many researchers to take care of the issue of load balancing in cloud computing.

Keywords- Cloud Computing, Load Balancing.

I. INTRODUCTION

Load balancing is the process of reassigning the aggregate burdens to the individual hubs of the aggregate framework to make the best reaction time and furthermore great use of the assets. Cloud computing is an internet computing in which the load balancing is the one of the testing errand. Different strategies are to be utilized to improve a framework by designating the heaps to the hubs in a balancing manner yet because of network congestion; transmission capacity utilization and so forth, there were issues are happened. These problems were settled by a portion of the current strategies [Ram Prasad Padhy 2011]. A load balancing algorithm which is dynamic in nature does not consider the past state or conduct of the framework, that is, it relies upon the current behavior of the system. There were various goals that identified with the load balancing, for example, to enhance the performance substantially, to keep up the framework solidness and so on.

Contingent upon the present condition of the framework, load balancing algorithms can be sorted into two types they are static algorithms and dynamic algorithms. In the static algorithm there was earlier information of the framework is required and not rely upon the current system [Ratan Mishra 2012]. On account of dynamic algorithm it depends on the current system and it is preferable execution over the static algorithm.

There are mainly two types of load balancing algorithms:

A. Static Algorithm: In static algorithm the traffic is partitioned equally among the servers. This algorithm requires earlier information of framework assets; with the goal that the choice of moving of the heap does not rely upon the present state of system Static algorithm is legitimate in the framework which has low variety in load.

B. Dynamic Algorithm: In dynamic algorithm the lightest server in the entire network or system is looked and favored for balancing a load. For this genuine time communication with network is required which can increment the traffic in the framework? Here current condition of the framework is utilized to settle on choices to manage the load.
The goal of load balancing is to enhance the execution by balancing the load among the different assets (network links, central processing units, disk drives, and so forth.) to accomplish optimal resource usage, most extreme throughput, greatest reaction time, and to evade over-burden. Below figure 2 demonstrates the block diagram of cloud architecture.

D. Zhang et al.[2009] proposed a binary treestructure that is utilized to segment the reproduction district into sub-spaces. The qualities of this quick versatile balancing method are to be balanced the workload between the processors from local areas to global areas. According to the distinction of workload, the courses of action of the cells are gotten. In any case, the primary workload concentrates on specific cells with the goal that the strategy of altering the vertices of the framework can be long a direct result of the local workload can be considered. This problem can be stayed away from by the quick load balancingadaptive method.

Markus et al.[2010] addressed the concept of overlay systems for the interconnection of machines that makes the foundation of an online environment. So the proposed network that improves practicality and
load balancing to the dynamic virtual situations. This proposed framework created hyper verse engineering that can be in charge of the correct facilitating of the virtual world. There were self-sorted out load balancing techniques by which the world surface is subdivided in to little cells, and it is overseen by a public server. In this cells various hotspots with the goal that unquestionably the mass of the question in the cell can be ascertained by the public server Hotspot accuracy is better while expanding the network load. The proposed algorithm can't stay away from the over-burden hubs however discover the quantity of connections that doled out to every hub while joining the network. The points of interest are the network becomes versatility, effective directing, and blame tolerant. The inconvenience is the over-burden proportion toward the start is higher with the goal that public servers are at first put haphazardly so some time is utilized for balancing the load.

Yunhua Deng et al. [2010] proposed an algorithm for load balancing in powerful distributed framework and different hybrid environments In most shared framework the non-uniform of items in the space and furthermore the load of the hub can be changed ceaselessly because of the addition, cancellation and different various operations. This will prompts diminish the execution of the framework. So the idea of virtual server can be presented. In this proposed load balancing algorithm, the load information of the associate hubs is put away in different directories. These indexes help to plan reassignment of the virtual servers to build up a better balance. Insatiable heuristic algorithm used to discover a better solution for the best possible usage of the hubs. The tremendous number of virtual servers in the framework builds the use. The different load information in to the relating pool and after that the virtual server assignments are to be finished. This proposed algorithm ought to be connected to various sorts of assets like stockpiling, transmission capacity and so on. It was intended to deal with the different circumstances like changing heap of the hub, hub limit, entering and leaving of hubs and furthermore inclusion and erasure of the hubs. Points of interest are high hub use and increasing scalability. Drawback is the reassignment of the virtual server is difficult.

B. Dong et al.[2012] proposed dynamic record movement load balancing algorithm based on distributed architecture. Considered the substantial file system there were various problems like dynamic document movement, algorithm based just on incorporated framework and so on. So these problems are to be kept away from by the presentation of the algorithm called self-acting load balancing algorithm (SALB). In the parallel file system the information are exchanged amongst the memory and the storage devices with the goal that the data management is a vital part of the parallel file system. There were various challenges that are confronted amid load balancing in the parallel file system. They are versatility and the accessibility of the framework, network transmission and the heap movement. Considered the dynamic load balancing algorithms, the heap in every I/O servers are diverse in light of the fact that the workload moves toward becoming fluctuates ceaselessly. So there were some basic leadership algorithms are needed.

Tejinder Sharma et al, [2013] proposed an effective and upgraded scheduling algorithm that can keep up the load balancing and gives better enhanced methodologies through productive job scheduling and adjusted asset allocation techniques. Load Balancing is finished with the assistance of load balancers where every approaching solicitation is diverted and is straightforward to customer who makes the demand. In light of parameters, for example, accessibility or current load, the load balancer utilizes different scheduling algorithms to figure out which server should deal with and advances the demand on to the chose server. Load balancing guarantees that every one of the processors in the framework and additionally in the network does around the equivalent measure of work at any moment of time. The general reaction time and server farm preparing time is enhanced and taken a toll is reduced.
PriyankSinghal et al.[2014],the Authors have concentrated on a two level undertaking distribution system over a three level cloud architecture. They examined the utilization of a half and half assignment scheduling algorithm which joins two regularly utilized planning strategies, the MM (Min-Min) and OLB (OpportunisticLoad Balancing) to make a cross breed Balanced Load Min-Min calculation (BLMM) algorithm. The idea of BLMMscheduling algorithm is to disseminate undertaking among each administration supervisor into some subtasks to be executed in a reasonable administration hub. BLMMconsiders the execution time of each subtask on each administration hub. Each subtask will be made sense of the execution time on various administration hubs through specialist. As per the data assembling by operator, each administration chief picks the administration hub of most brief execution time to execute diverse subtasks and records it into the Min-time cluster. At long last, the Min-time cluster of each subtask is recorded that is an arrangement of insignificant execution time on certain administration hubs. This prompts more effective execution and keeps up load balancing of the system nodes.

Suriya Begum et.al.[2014] proposed a Mathematical model solely considering virtual machine for performing load balancing. The framework together tends to the steering and assignment planning and furthermore concentrates on the issues relating to asset allotment. A novel mathematical model considering stochastic model for load balancing and planning for cloud computing groups has been created. A cloud system comprises of various networked servers. Each of the servers may have different Virtual Machines. Each Virtual Machine requires an arrangement of assets, including CPU, memory, and storage space, considered a stochastic model for load balancing and planning for cloud computing groups. An essential commitment is the advancement of edge based non-pre-emptive VMconfiguration strategies. These approaches are made almost throughput-ideal by picking adequately long casing terms, though the broadly utilized best fit strategy was appeared to be not throughput optimal.

Anjali et al.[2015] proposed Dynamic Load Balancing in Cloud Computing utilizing specialists. The utilization of mobile has indicated preferable outcomes over existing loadbalancing algorithms. The entire working is finished with the assistance of specialists known as a consistent operator and a Mobile agent is a program which relocates starting with one machine then onto the next. With the assistance of this, an executable code is moved to another host. It runs freely as indicated by the enthusiasm of customer. Mobile agent adds to consistent specialist. The diverse elements of mobile agent which makes it remarkable are the capacity of learning and mobility.

ReenaPanwar et al. [2015] have introduced a dynamic load management algorithm for conveyance of the whole approaching solicitation among the virtual machines (VMs) adequately. The heap is overseen by the server by considering the present status of present VMs for ask for task strongly for all free VMs to be utilized at ask for task and will take more demands that are dynamic in nature Therefore, diminishment accordingly time when contrasted with VM Assign Algorithm. The test comes about has demonstrated that this algorithm have least reaction time and appropriate asset use by utilizing Cloud Analyst apparatus and checked its execution on different distinctive load distributions.

Surbhi Kapoor et al. [2015] considered a Cluster based load balancing which functions admirably in heterogeneous hubs condition, considers asset particular requests of the undertakings and decreases filtering overhead by separating the machines into groups. At the point when client presents the errand, the load balancer matches undertaking assets particular prerequisites with limit scope of bunch keeping in mind the end goal to allot the assignment to fitting group. At that point among the bunches, load balancer matches the appropriate accessible VM to which errand must be doled out, examining of records is separated into two levels. This will decrease the overhead required in checking list and can
allocate better VM to the assignment. Reasonable for heterogeneous condition and it likewise considers asset specific demands of the tasks.

Milani and Navimipour [2016] have introduced an orderly audit of the current load balancing systems. They grouped the current strategies in view of various parameters. They looked at some prevalent load-adjusting calculations and introduced their primary properties, including their favorable circumstances and drawbacks. They additionally tended to the difficulties of these calculations and said the open issues. In any case, their work does not have a dialog with respect to the heap adjusting and undertaking planning strategies in Hadoop MapReduce that is an issue these days.

Gholi and Rahmani [2017] have focused in paper on task scheduling and load balancing algorithms and introduced the latest classification of such algorithms like Hadoop MapReduce load balancing category, Natural Phenomena-based load balancing category, Agent-based load balancing category, General load balancing category, application-oriented category, network-aware category, and workflow specific category.

LOAD BALANCING: Load balancing in mists is a technique that circulates the abundance dynamic nearby workload equally over every one of the hubs. It is utilized for accomplishing a superior administration provisioning and asset use proportion, henceforth enhancing the general execution of the framework. Incoming errands are originating from various area are gotten by the load balancer and afterward dispersed to the datacenter for the best possible load distribution [P. B. Soundarabai 2014, M. R. Sumalatha 2014].

The aim of load balancing is as follows:
1. to improve the performance
2. Build fault tolerance system
3. To maximize resource utilization
4. To increase the availability of services
5. Maintain system stability
6. To reduce the execution time and waiting time of task coming from different location
7. Accommodate future modification
8. To increase the user satisfaction.

III. BASIC TYPES OF LOAD BALANCING ALGORITHMS

There is a to a great degree huge requirement for load balancing in intricate and substantial disseminated frameworks. Load balancer takes a choice to exchange the job to the remote server for load balancing. Load balancer can works in two ways: one is cooperative and no cooperative. In cooperative way, to achieve the ideal reaction time, every one of the hubs works to assemble [Venubabu Kunamneni 2012]. In no cooperative way, reaction time is increment by the autonomously running the assignments. Some of the algorithms for load balancing are considered in this paper.
IV. CHALLENGES OF LOAD BALANCING

**Performance:** It can be defined as the effectiveness of the framework. It must be improved.

**Response Time:** can be defined as the measure of time taken to respond by a load balancing algorithm in a conveyed framework. For better execution, this parameter ought to be reduced.

**Resource Utilization:** this is utilized to test the usage of assets. It ought to be most extreme for a productive load balancing system.

**Overhead Associated:** determines the measure of overhead included while implementing a load balancing system. It is made out of overhead because of development of errands, between process correspondences. Overhead ought to be lessened with the goal that a load balancing algorithm performs well.

**Scalability:** the quality of service should be same if the number of users increases. The more number of nodes can be added without affecting the service.

**Point of Failure:** designed the system in such a path, to the point that the single point disappointment does not influence the provisioning of services. Like in centralized system, on the off chance that one central node is flop, at that point the entire system would bomb, so load balancing framework must be composed keeping in mind the end goal to overcome this problem.

**Fault Tolerance:** In spite of the node failure, the ability of a system to perform uniform load balancing. The load balancing is the best fault-tolerant technique.

**Throughput:** It is the number of task executed in the fixed interval of time. To improve the performance of the system, throughput should be high.

**GOALS OF LOAD BALANCING:** The load balancing of an application has an immediate contact on the speedup. The reworking of balanced workload by implies of jobs and minimizing the entomb process communication require with ideal asset usage and assignment reaction time are the essential streamlining goal of load balancing [Che-Lun Hung 2012, RIMA P 2014, S. Kapoor 2015]. A portion of the main goals of a load balancing algorithm, as called attention to by [3] are to:

**Job Equality:** To treat all jobs in the system equally regardless of their origin;
**System stability:** The ability to account for emergency situations such as sudden surge of arrivals so that system performance does not deteriorate beyond a certain threshold while preventing nodes of the distributed system from spending too much time passing up jobs among themselves instead of executing these jobs.

**Performance Improvement:** Achieve a greater overall improvement in system performance at a reasonable cost, e.g., reduce task response time while keeping acceptable delays;

**Modifiability:** Have the ability to modify it in accordance with any changes or expand in the distributed system configuration.

**V. CONCLUSION**

The load balancing of the present framework is one of the greatest issues. Various techniques and algorithms are utilized to solve the problem. In this paper we study different existing load balancing techniques in different environments. Countless and diverse sorts of soft computing techniques can be incorporated into the future for the better use and needs of the client. The different load balancing techniques are likewise being looked at here. The principle reason for load balancing is to fulfill the client prerequisite by distributing load powerfully among the hubs and to make most extreme asset use by reassigning the total load to individual node. This guarantees each asset is disseminated productively and equitably. So the performance of the framework is expanded. We have likewise examined virtualization of cloud and required subjective lattice for load balancing.

**VI. FUTURE WORK**

The load balancing is one of the greatest issues in Cloud Computing condition. Load balancing is the necessity of a cloud environment and how well this prerequisite is met relies upon the algorithm chosen. The different load balancing techniques are likewise being looked at The performance of the load balancing algorithms is assessed by various parameters like throughput, reaction time, execution time, add up to cost et cetera. Different parameters like adaptation to non-critical failure can be incorporated into the future for the better usage and necessities of the user.

**REFERENCES**

[1] Dongliang Zhang, Changjun Jiang, Shu Li, “A fast adaptive load balancing method for parallel particle-based simulations”, Simulation Modelling Practice and Theory 17 (2009) 1032–1042.

[2] Bin Dong, Xiuqiao Li, Qimeng Wu, Limin Xiao, Li Ruan, “A dynamic and adaptive load balancing strategy for parallel file system with large-scale I/O servers”, J. Parallel Distribution Computing. 72 (2012) 1254–1268.

[3] Markus Esch, Eric Tobias, “Decentralized scale-free network construction and load balancing in Massive Multiuser Virtual Environments”, in: Collaborative Computing: Networking, Applications and Worksharing, CollaborateCom, 2010, 6th International Conference on, IEEE, 2010, pp. 1–10.

[4] Yunhua Deng, Rynson W.H. Lau, “Heat diffusion based dynamic load balancing for distributed virtual environments”, in: Proceedings of the17th ACM Symposium on Virtual Reality Software and Technology, ACM, 2010, pp. 203–210

[5] Ram Prasad Padhy , P Goutam Prasad Rao, —Load Balancing in Cloud Computing Systems, National Institute of Technology, Rourkela, India, 2011.

[6] Ratan Mishra and Anant Jaiswal, —Ant Colony Optimization: A solution of Load Balancing in Cloudl, International Journal of Web & Semantic Technology (IJWeST), April 2012

[7] Venubabu Kunamneni, “Dynamic Load Balancing for the cloud”, International Journal of Computer Science and Electrical Engineering, 2012.

[8] Che-Lun Hung1, Hsiao-hsi Wang2 and Yu-Chen Hu2, —Efficient Load Balancing Algorithm for Cloud Computing Networkl, IEEE Vol. 9, pp: 70-78, 2012

[9] RIMA P. LINGAWAR , MILIND V. SRODE ,MANGESH M. GHONGE ,”Survey on Load-Balancing Techniques in Cloud Computing “,International Journal of Advent Research in Computer & Electronics, Vol.1, No.3, E-ISSN: 2348-5523, May 2014.
[10] S. Kapoor, and C. Dabas, Cluster based load balancing in cloud computing, Proc. Eighth International Conference in Contemporary Computing (IC3), 2015, 76-81.

[11] P. B. Soundarabai, A. S. Rani, R. K. Sahai, J. Thriveni, K. R. Venugopal, and L. M. Patnalk, Load balancing with availability checker and load reporters (LB-ACLRs) for improved performance in distributed systems, Proc. 2nd International Conference on Devices, Circuits and Systems (ICDCS), 2014, 1-5.

[12] M. R. Sumalatha, C. Selvakumar, T. Priya, R. T. Azariah, and P. M. Manohar, CLBC-Cost effective load balanced resource allocation for partitioned cloud system, Proc. International Conference on Recent Trends in Information Technology (ICRTIT), 2014, 1-5.

[13] Anjali, Jitender Grover, Anjali, Jitender Grover. 2015 A New Approach for Dynamic Load Balancing in Cloud Computing.

[14] Reena Panwar, Prof. Dr. Bhawna Mallick. 2015 Load Balancing in Cloud Computing Using Dynamic Load Management Algorithm.

[15] Surbhi Kapoor, Dr. Chetna Dabas. 2015 Cluster Based Load Balancing in Cloud Computing.

[16] Tejinder Sharma, Vijay Kumar Banga. 2013 Efficient and Enhanced Algorithm in Cloud.

[17] Priyank Singhal, Sumiran Shah, Sumiran Shah. 2014 Load Balancing Algorithm over a Distributed Cloud Network">

[18] Suriya Begum, Prashanth CSR. 2014 Mathematical Modelling of Joint Routing and Scheduling for an Effective Load Balancing in Cloud.

[19] Milani, A.S., Navimipour, N.J., 2016. Load balancing mechanisms and techniques in the cloud environments: systematic literature review and future trends. J. Netw. Comput. Appl. 71, 86–98.

[20] Ghomi, Rahmani, Qader, 2017. Load-balancing algorithms in cloud computing: A survey. Journal of Network and Computer Applications Volume 88 Issue C, June 2017.