International Journal of Scientific Research in Computer Science, Engineering and Information Technology

© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

Survey of Load Balancing Methods in Cloud Computing

Sachin Shingne, Dr. Umesh K Lilhore, Phiza AmbreenKhan

PG Department of Computer Science & Engineering, NRI-IIST Bhopal, Maharashtra, India

ABSTRACT

Now these days cloud computing technology is widely used technology in IT world. Cloud computing serves computing resources as services to cloud users. Cloud computing provides PaaS, IaaS and SaaS services. Cloud uses can be divided in to four major group's private, public, community and hybrid. Cloud computing reduce cost of ownership. This quality of cloud computing makes it more popular among users. A user can stores their private data over cloud and can access any time. Cloud commuting service are based on "Pay and use based". Day by day size of cloud users and cloud services are getting increases rapidly. So it is quite challenging for cloud service providers to satisfy cloud user requirement with out and failure in optimum cost. To avoid failure in service and achieved optimum cost various load balancing methods are used. A load balancing is a technique which migrates a job from over loaded machine to an under loaded machine without disturbing current running jobs. In this survey paper we are presenting a review and comparison of various load balancing method suggested by various researchers' for cloud computing.

Keywords- Cloud computing, load balancing, Performance, optimal utilization, Computing resources

I. INTRODUCTION

Cloud is the group of appropriated PCs that gives on-request computational assets over a organize infrastructure. A Cloud registering is turning into a propelled innovation as of late. It is thoughtfully conveyed framework where processing assets circulated through the system (Cloud) and administrations pooled together to give the clients on pay-as-required premise [2].

Distributed computing gives everything as an administration and are sent as open, private, group, and half and half mists. The three fundamental administration layers of distributed computing are: Software as a Service (SaaS), where the client does not have to deal with the establishment and setup of any equipment or programming, for example, Google Online office, Google Docs, Email cloud, and so forth. Stage as a Service (PaaS), where an administration is a conveyance of a processing stage over the web where clients can make and introduce their own application as they require [1, 7].

Setup of processing stage and server is overseen by the seller or cloud supplier. Case of PaaS is Google App Engine. Foundation as a Service (IaaS), where servers, programming, and system gear is given as an on-request benefit by the cloud supplier. The primary capacity of sharing assets, programming, data through the web are the principle enthusiasm for distributed computing with an intend to diminish capital and operational cost, better execution as far as reaction time and information handling time, keep up the framework steadiness and to suit future adjustment of the framework. So there are different specialized difficulties that should be tended to like Virtual Machine (VM) movement, server combination, adaptation to internal failure, high accessibility and adaptability. In any case, the focal issue is the heap adjusting [4].

It is the component of spreading the heap among different hubs of a circulated framework to enhance both asset sending and occupation reaction time while additionally maintaining a strategic distance from a circumstance where a portion of the hubs are having an immense measure of load while different hubs are doing nothing or sit without moving with next to no work. It too guarantees that all the processor in the framework or every hub in the system does roughly the equivalent measure of work at any moment of time [6].

The objective of load adjusting is to enhance the execution by adjusting the heap among the different assets (organize joins, focal handling units, plate drives, and so forth.) to accomplish ideal asset use, greatest throughput, most extreme reaction time, and to evade over-burden.

II. CLOUD COMPUTING

It is a data innovation (IT) worldview that empowers omnipresent access to shared pools of configurable framework assets and more elevated amount benefits that can be quickly provisioned with insignificant administration exertion, frequently finished the Internet. Distributed computing depends on sharing of assets to accomplish cognizance and economies of scale, like an open utility [10].

Outsider mists empower associations to center around their center organizations as opposed to exhausting assets on PC foundation and maintenance. Advocates take note of that distributed computing enables organizations to evade or limit in advance IT framework costs. Defenders likewise assert that distributed computing enables ventures to get their applications up and running speedier, with enhanced reasonability and less support, and that it empowers IT groups to all the more quickly change assets to meet fluctuating and capricious demand. Cloud suppliers normally utilize a "pay-as-you-go" demonstrate, which can prompt sudden working costs if overseers are not acquainted with cloud-valuing models [14].

2.1 Model in cloud computing-Following deployment models are used-

- **Private cloud**-Private cloud is cloud infrastructure operated totally for a dedicated enterprise, whether managed internally or with the aid of a third-party, and hosted either internally or externally.
- **Public cloud**-A cloud is called a "public cloud" when the offerings are rendered over a community this is open for public use. Public cloud services may be unfastened.[89] Technically there can be very little distinction between public and private cloud structure, but, protection consideration may be notably different for services which might be made to be had through a service company for a public target market and whilst communique is effected over a non-relied on community. Services (AWS), Oracle, Microsoft and Google.
- **Hybrid cloud**-Hybrid cloud is a composition of two or extra clouds (personal, community or public) that continue to be distinct entities however are sure collectively, imparting the blessings of a couple of deployment fashions.
- **Community cloud**-Community cloud stocks infrastructure among several groups from a specific network with commonplace worries (protection, compliance, jurisdiction, etc.), whether managed internally or by using a third-celebration, and either hosted internally or externally.

2.2 Cloud services-Though provider-orientated structure advocates "everything as a provider" (with the acronyms EaaS or XaaS,[57] or actually aas), cloud-computing vendors provide their "offerings" according to one of a kind fashions, of which the 3 general fashions per NIST are Infrastructure as a-

Service (IaaS)
Platform as a Service (PaaS)
Software as a Service (SaaS)

Figure 1. Cloud Services

These fashions offer increasing abstraction; they're as a result regularly portrayed as a layers in a stack: infrastructure-, platform- and software program-as-a-provider, however these want no longer be related. For instance, you'll provide SaaS implemented on physical machines (bare metal), without using underlying PaaS or IaaS layers, and conversely you can actually run a program on IaaS and get admission to it at once, without wrapping it as SaaS.

III. LOAD BALANING IN CLOUD

Cloud load balancing is the process of distributing workloads and computing resources in a cloud computing surroundings. Load balancing allows establishments to manage software or workload demands by way of allocating resources amongst a couple of computers, networks or servers. Cloud load balancing entails hosting the distribution of workload traffic and needs that reside over the Internet. Figure 3.1 shows cloud load balancing.

3.1 Types of Load Balancing- Load balancing algorithms can be broadly categorized into 2 sorts:

- **Static algorithms**-In Static scheduling the mission of duties to processors is performed before application execution starts off evolved e.g. in compile time. Scheduling decision is based totally on facts approximately mission execution times, processing sources, and so forth. Static scheduling strategies are no preemptive. The goal of static scheduling methods is to decrease the overall execution time. These algorithms cannot adapt to load adjustments in the course of run-time [7].
- Dynamic scheduling (often referred to as dynamic load balancing) is based on the redistribution of approaches among the processors during execution time.

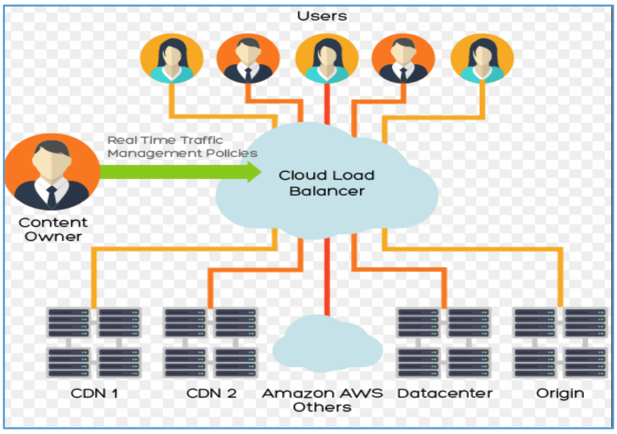

Figure 2. Cloud Load Balancing [10]

3.2 Why Load balancing?

Load balancing provides following benefits:

• **Optimum utilization of resources:** it's far one of the paramount pursuits of load balancing as right useful resource usage is essential for performance of the cloud version.

• **High Throughput:** high throughput is a favored characteristic required for an excessive overall performance machine that's handiest viable if the workload and sources are disbursed to the distinctive nodes frivolously.

• **Short response time:** Load balancing methods attempt to less time in response.

• **Avoiding bottlenecks:** To avoid any sort of congestion or bottlenecks inside the cloud environment both in community or records availability.

IV. EXISITNG WORK

Due to the current emergence of cloud computing research on this region is inside the initial level. A useful resource allocation mechanism with pre-empt in a position project execution which increases the usage of clouds proposed [1]. Cloud offer diverse services to the used via various element that it's represent. Cloud is the concept that evolve distribution of assets in digital environment so that every consumer can get admission to sources, allows utilization of assets.

Load balancing in cloud computing device mentioned on fundamental ideas of Cloud

Computing [1, 7]. Load balancing and studied a few present load balancing algorithms which can be implemented to clouds. In Load Balancing method idea of virtual community and fuzzy concept play a major role as virtual community is the key idea to distribute resources on various node in order that every node has equivalent get right of entry to resources [9].

For even distribution of sources there need to be some means to analyses availability of resources as well. For implementation evaluation of load balancing method numerous simulator are used.

Different algorithms distinct attitude so that simulator may be used to analyses the precise algorithm that can be applied in a particular environment. Simulator analyses the performance of a method for execution in real existence utility [6].

Load balancing is one of the paramount responsibilities undertaken to enhance the overall performance of a cloud and to attain superior aid utilization, it's miles an inherent part of managing a cloud computing surroundings load balancing strategies strive to obtain stronger throughput, shortened response time, heading off congestion, brief response time and so on. Load may additionally contain of quantity of memory used, CPU load, community load, postpone load etc. [9].

The intention of load balancing methods is to distribute the weight calmly amongst distinct nodes in order that no node gets crushed. Proper load balancing can help in using to be had assets optimally, thereby minimizing the useful resource intake. It additionally helps in imposing fail-over, permitting scalability, averting bottlenecks and over-provisioning, decreasing reaction time and many others. [13].

As cloud carrier carriers dispense unique offerings and assets via servers load balancing techniques make certain efficient working of the servers through taking suitable movement if any server is overloaded or if some is in an idle kingdom due to loss of resources [8].

V. LOAD BALACNING METHODS

Following Load balancing methods are widely used in cloud computing:

- **Round Robin Method**-It is one of the simplest scheduling techniques that utilize the principle of time slices. The time is divided into multiple slices and each node is given a particular time interval i.e. it employs the principle of time scheduling. Each node is given a time slice and in this time slice the node will perform its operations [4].This algorithm works on random selection of the virtual machines.
- **Honey Bee Foraging Algorithm:** This set of rules emulates the conduct of honey bees to discover food so its named honey bee foraging set of rules this set of rules is derived from an in depth evaluation of the behavior that honey bees adopt to discover and acquire food. In bee hives, there's a class of bees referred to as the scout bees which forage for meals sources, upon locating one, they come back to the beehive to advertise this the use of a dance called waggle/tremble/vibration dance the show of this dance, offers the concept of the quality and/or amount of meals and also its distance from the beehive. Forager bees then comply with the Scout Bees to the location of meals after which begin to reap it they then return to the beehive and do a waggle or tremble or vibration dance to different bees in the hive giving an idea of the way lots food is left and as a result resulting in both more exploitation or abandonment of the meals supply [10]
- **Weighted Round Robin**-Another way to define round robin algorithm is a better allocation concept known as Weighted Round Robin. Allocation in which one can assign a weight to each virtual machine so that if one virtual

machine is capable of handling twice as much load as the other, the more powerful server gets a weight of 2. In this cases, the Data Center Controller will assign two requests to the powerful virtual machine for each request assigned to a weaker one. The key issue in this allocation is this that it does not consider the advanced load balancing requirement such as processing times for each individual requests [7].

- **Ant Colony Load Balancing:** [1] proposed a load balancing algorithm based totally on ant colony and complex network concept (ACLB) in an open cloud federation this method brings into utility small international and scale free traits of a complicated community to attain stronger load balancing this technique overcomes heterogeneity it adapts to dynamic environments, has excellent fault tolerance and may be very solid and enhances the general overall performance of the device
- **Equally Spread Current Execution Algorithm-** Equally spread current execution algorithm process handle with priorities. ESCE algorithm distribute the load randomly by checking the size and transfer the load to that virtual machine which is lightly loaded or handle that task easy and take less time to accomplish the task and maximize throughput. ESCE algorithm a spread spectrum technique in which the load balancer spread the load of the job in hand into multiple virtual machines [16].
- **Active Clustering:** [5] investigated a self-aggregation load balancing set of rules it businesses like (Similar carrier type) times collectively it consists of iterative execution by using every node in the community at a random factor a node becomes initiator and chooses a in shape maker node randomly from its modern-day neighbors ,the healthy maker makes a hyperlink between one of the fit makers neighbors this is just like initiator node sooner or later match maker eliminates the hyperlink among itself and the initiator node the principle intention of grouping similar nodes is to optimize process assignments by using connecting comparable offerings [8].

VI. CONCLUSIONS & FUTURE WORKS REFERENCES

Cloud computing has revolutionized the way resources and offerings are availed by using users over the Internet however it has its challenges efficient load balancing is one of the key issues concerning any cloud service company as a good distribution of workload throughout extraordinary nodes is a pivotal requirement for high useful resource usage and user pleasure there are distinctive classifications of load balancing algorithms every algorithm gives finest effects in a particular condition and scenario, depending on goals of the cloud environment and given resources an set of rules is chosen. In this survey paper we are presenting a review and comparison of various load balancing method suggested by various researchers' for cloud computing.

In future work we will developed an efficient load balancing method for cloud computing and will checked its performance with various existing methods to shows the robustness and efficiency of proposed method.

VII. REFERENCES

[1]. Ashish Gupta, Ritu Garg, "Load Balancing Based Task Scheduling with ACO in Cloud Computing", IEEE international Conference on Computer Applications (ICCA) June 2017, pp 174-180.

[2]. K.Sutha, Dr.G.M.Kadhar Nawaz, "Research Perspective of Job Scheduling in Cloud Computing", 2016 IEEE Eighth International Conference on Advanced Computing (ICoAC), April-2016, pp 61-67.

[3]. AV. Karthick, Dr.E.Ramaraj, R.Ganapathy Subramanian, An Efficient Multi Queue Job Scheduling for Cloud Computing, IEEE 2014

World Congress on Computing and Communication Technologies, pp. 164-166.

[4]. Manisha Patel, Umesh Lilhore," A Survey on Efficient Data Retrieval for Cloud Computing", International Journal of Research in Advent Technology, Vol.5, No.1, January 2017, PP 1-5.

[5]. Huankai Chen, Frank Wang, Dr Na Helian, Gbola Akanmu, User-

[6]. Priority Guided Min-Min Scheduling Algorithm For Load Balancing in Cloud Computing, IEEE February 2013

[7]. Rahul Upadhyay, Umesh lilhore," Review of Various Load Distribution Methods for Cloud Computing, to Improve Cloud Performance", IJCSE, Volume-4, Issue 12, 2016, PP 61-64.

[8]. Yichao Yang, Yanbo Zhou, Zhili Sun, Haitham Cruickshank, Heuristic Scheduling Algorithms for Allocation of Virtualized Network and Computing Resources, Journal of Software Engineering and Applications, January 2013, 6, pp. 1-13

[9]. Umesh Lilhore, Santosh Kumar, "Advance Anticipatory Performance Improvement Model, for Cloud Computing", International Journal of Recent Trends in Engineering &amp; Research (IJRTER), Volume 02, Issue 08, 2016, PP 210-216.

[10]. Ke Liu1, Hai Jin , Jinjun Chen , Xiao Liu , Dong Yuan and Yun Yang, A Compromised-Time-Cost Scheduling Algorithm in SwinDeW-C for Instance-Intensive Cost-Constrained Workflows on a Cloud Computing Platform, The International Journal of High Performance Computing Applications, Volume 24(4) pp. 445-456, May 2010

[11]. Umesh Lilhore, Dr. Santosh Kumar," Anticipatory Data Replication Strategy with Dynamic Distributed Model for Cloud Computing", International Journal of Research in Applied Science &amp; Engineering Technology (IJRASET), Volume 4 Issue VIII, August 2016, PP 554-559.

[12]. Arash Ghorbannia Delavar, Mahdi Javanmard, Mehrdad Barzegar Shabestari and Marjan Khosravi Talebi, RSDC (Reliable Scheduling Distributed In Cloud Computing), International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.2, No.3, June 2012

[13]. Umesh Lilhore and Dr. Santosh Kumar, "Modified fuzzy logic and advanced particle swarm optimization model for cloud computing", International Journal of Modern Trends in Engineering and Research (IJMTER), Volume 03, Issue 08, August- 2016, PP 230-235.

[14]. Nitish Chopra, Sarbjeet Singh, HEFT based Workflow Scheduling Algorithm for Cost Optimization within Deadline in Hybrid Clouds, IEEE - 31661, 4th Computing, Communications and Networking Technologies(ICCCNT) July 4-6, 2013

[15]. Umesh Lilhore and Santosh Kumar, "A Novel Performance Improvement Model for Cloud Computing", IJSDR, Volume 1, Issue 8, 2016, 410-412.

[16]. Wei Wang, Guosun Zeng, Daizhong Tang, Jing Yao, Cloud-DLS: Dynamic trusted scheduling for Cloud computing, Expert Systems with Applications 39 (2012) pp. 2321-2329 2011 Elsevier Ltd.