International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

A Review on Various Load Balancing Algorithms in Cloud Computing

Shyam Singh Rajput
M.Tech Scholar, Department of CSE,
ITM Universe, Gwalior, M.P., India

Virendra Singh Kushwah
Assistant Professor, Department of CSE
ITM Universe, Gwalior, M.P., India

Abstract— Cloud computing is a mannequin that provides resources like services, applications over the web. Load balancing is likely probably the most essential challenges in cloud computing which is requisite to distributing the dynamic workload over multiple machines to make certain that no single machine is overloaded. It helps in ideal use of resource and as a consequence enhancing the performance of the system. There are many current scheduling algorithms that may preserve load balancing and offer higher strategies via efficient job scheduling and useful resource allocation methods as well. In an effort to reap highest profits with enhanced load balancing algorithms, it's essential to utilize resources efficiently. This paper describes a brief overview on various load balancing algorithms in cloud computing environment. All other existing algorithms have different working manner with respect to cloud computing and also support different QoSs parameters.

Keywords—Cloud computing, Cloud service models, Virtual machine, Load balancing.

I. INTRODUCTION

Cloud computing is new technology which is based on internet in which internet can represented as a cloud. Cloud is a platform that provides resources like services, applications and storage network to a computers and devices based on pay-per- model [1].Many Cloud computing providers have setup a few data centers at various geographical places over the web as a way to serve wants of their buyers around world [2].Today's this technology rising at a quick rate. It has also enabled to effortlessly access to associated data and applications from anywhere around the world. The fundamental purpose of using this technology is to growth the performance and efficiency and reduces the cost [3]. The increases amount of information storage quickly in cloud computing environment. The increase data storage very fast so the load balancing is a primary concern in cloud computing. When a numbers of jobs occur equal time then load balancing is predominant issue. Load balancing helps to work distribute between all to be had nodes to be certain that no node is overloaded and no want is free. Load balancing benefits to use of resources which too assistance in bettering the performance of the method in cloud system [4].

A. Cloud Computing Characteristics
1) *Services provide on demand:* - When user want to resources then the cloud provide a services on demand.
2) *Rapid Elasticity:* - Number of various resources in cloud is increase or decrease easily.
3) *Resource Pooling:* - In cloud structure resources are allocated consistent with consumer requirement. The each resource is collective to serve finish users utilizing model of cloud.
4) *Broad Access to Network:* - The cloud resources access is possible throughout the network and used standard methods for the users to right to use the network.
5) *Pay per use:* - Every consumers pay charges when it is usage of computing resources [4], [5].

B. Services of CloudCloud computing provide many services to the end users:

Fig.1. Cloud Service Models

1) *Software as a Service:* This service model is a software conveyance model in which every single applications are gotten to through web browser. In this SaaS model the users are not concerned with the cloud structure and also not

concern with network, servers, storage, operating systems, and platform. Exclusive cloud consumer's applications are simply organized in a particular logical atmosphere on the SaaS, to achieve optimization in terms of security, speed, availability, disaster recovery and maintenance. There are various services are provide as SaaS like salesforce.com. Google docs, Google mail.

2) *Platform as a Service:* This Service model is a development platform which permits for cloud customers to progress cloud application and services (example SaaS) directly on PaaS cloud. In this Platform as a Service model the consumers does not control the basic cloud foundation including servers, network, storage and operating system. Example of PaaS is Aneka and Google App Engine.

3) *Infrastructure as a Service:* This service model provides Infrastructure of servers and different software to the customer on their demand. By using this users now not have to buy any specified hardware and software. Cloud users directly use of IT infrastructure (storage, networks, processing and other computing resources). IaaS provideservices like Amazons EC2 [4],[6].

C. *Cloud Deployment Models*
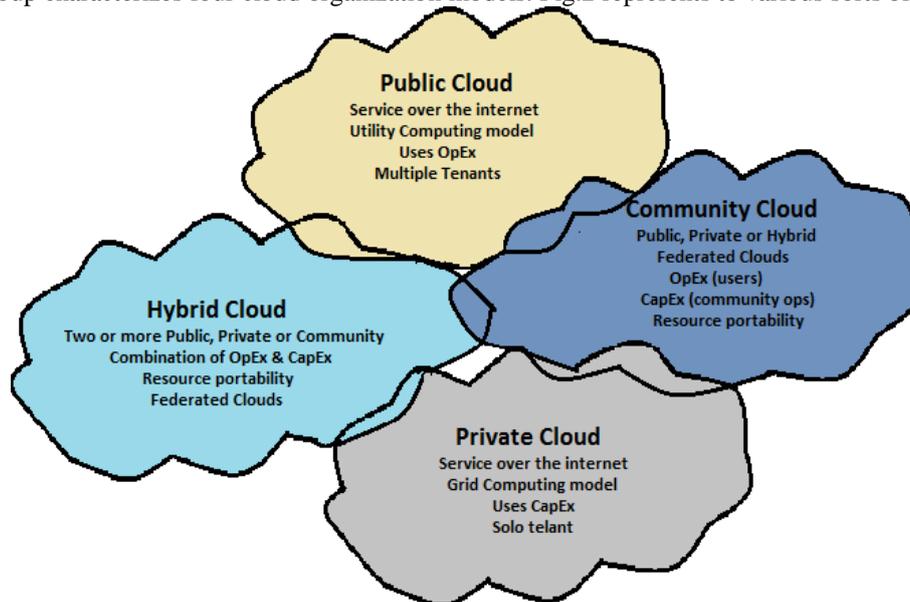The cloud group characterizes four cloud organization models. Fig.2 represents to various sorts of cloud.

Fig.2. Cloud Deployment Models

1) *Public Cloud: -* This cloud is utilized by the general public users and the cloud service provider has the full responsibility for public cloud with its own qualities, policy, costing, profit, and charging model. Many popular cloud services are Google App Engine, Amazon EC2and salesforce.com.

2) *Private Cloud: -* This cloud will be cloud bases worked for a solitary association and give security to its resources.

3) *Community Cloud: -* In community cloud, cloud infrastructure which can be used through several organizations in a private community. This cloud is shared amongst many associations that have comparative cloud prerequisites.

4) *Hybrid Cloud: -* This cloud it utilizes a combination of no under two clouds where the clouds incorporate a blend of private cloud, public cloud or community cloud[6], [7].

II. VIRTUALIZATION

Cloud computing uses virtualization for provisioning services to the client. On virtualization many running systems can be keep running on the single computer so resource utilization is increased. The hardware resources are combined for enhanced the productivity of server. Computer architecture is uses software for the proper resource utilization called Hypervisor. It is too named the VMM (Virtual Machine Monitor) for running the more than one operating systems on the single host. There are two types of virtualization.

A. *Full Virtualization:-* In Full Virtualization, the entire installation of one machine is done on another machine. So all the real machine functionality may too be available in virtual machine [5].

B. *Para Virtualization: -*In Para Virtualization, on a solitary desktop different operating systems can be run. Here entire functionalities are not fully available; services are delivered in a partial manner [5].

III. LOAD BALANCING

In cloud environment, load balancing is a method that circulates the dynamic nearby workload similarly across entire available nodes. Load balancing is used for achieving anenhanced resource utilization ration and service provisioning, therefore enhancing the entiresystem performance. In load balancing, incoming tasks are coming from the different location are received by the load balancer and then tasks are circulated to the data farm for the appropriate load distribution. A perfect load balancing algorithm ought to evade under loading or overloading particular node. The

important goal of a load balancing in an environment of cloud computing remains togrowththe reply time of job with the aid of job by the whole load of procedure. Load balancers can operate in dual exclusive approaches: one is the cooperative and other is Non- cooperative. In non-cooperative mode, the duties run independently as a way to enhance the reply time of nearby tasks. In cooperative, the nodes work even as to be able to acquire the common purpose of optimizing the overall response time [4], [5] and [8].
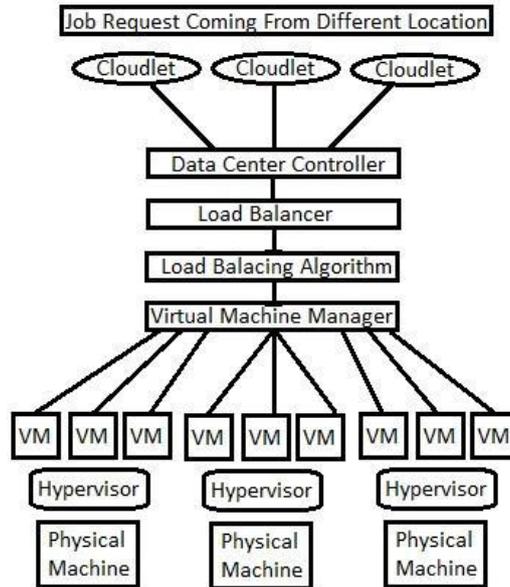

Fig.3. Cloud Load Balancing [5]

In the algorithms of cloud load balancing, is mostly alienated into two different groups: static and dynamic load balancing algorithm:

A. *Static Approach:* This static approach is commonly described in implementation or design of system. This algorithm divisions the traffic alike between all of the users. This algorithm requires a previous knowledge of approach resources, so that no longer rely upon the existing state of system for decision of shifting of the load. These are much simpler and ignore the existing state or the load on the node within system.

B. *Dynamic Approach:* In this dynamic approach considered only the present condition of the system during load balancing decision. This dynamic methodology is extraapposite for broadly distributed system such a cloud computing. Dynamic approach has two parts:

1) *Centralized Approach:* **--** In this centralized approach Simplest as only node is in charge for distribution and managing within the whole system.

2) *Distributed Approach:***--** in this dynamic methodology each node freely constructs own load vector. Vector accumulating the load know-how of other nodes. In this all selections are made locally utilizing local load vectors. This procedure is more compatible for generally allotted systems comparable to cloud computing [9], [10], [11].

IV.   EXISTING ALGORITHMS OF LOAD BALANCING IN CLOUD COMPUTING

In a cloud computing atmosphere there are a lot of algorithm of load balancing which attain higher throughput, toughen the reaction time, high resource use and better efficiency [9], [10].

A. *Task Scheduling founded on load balancing:***--**This algorithm mainly consists two level task scheduling device which is in light of load balancing to meet element requirements of users. This algorithm obtains high resource utilization, and this algorithm attains load balancing via first mapping jobs to virtual machines after which all virtual machines towards host resources .This improving the task response time, also provide well useful resource utilization [10].

B. *Opportunistic Load Balancing Algorithm (OLB):***--**This algorithm does not contemplate the existing workload of the Virtual machine (VM). It makes every node to be occupied. On this algorithm each and every unexecuted challenge can be finished in random order so that each task can be allotted to the node randomly. These processes slow manner because it will not calculate the existingoperation time of the node [12].

C. *Round-Robin Load Balancer:***--**Round-robin load balancer its' a static load balancing algorithm, on the time of allocating jobs which does now not remember the earlier load state of a node. This algorithm uses for allocating the jobs. In this algorithm it chooses the nodes randomly and after which jobs allot to all nodes in round robin manner. This algorithm will not be suitable for cloud computing because how much operating time of any process is not known prior to execution; there's a risk that nodes may earn heavily loaded. Then weighted round-robin algorithm used to be proposed to resolve this type problem. In this weighted round-robin algorithm, every node is appointed a specific weight. The weight assign to the nodes is helpful to control the network traffic [8],[13].

D. *Min-Min algorithm:***--**Min-Min Job Scheduling Algorithm it's a static scheduling algorithm. Min-Min algorithm begins with a suite of un-scheduling jobs. On this algorithm the jobs having minimum execution time first identifies and these jobs are scheduled first in this algorithm. Then it will calculate the expected completion time intended for

each tasks according to available virtual machines then the resource that has the least completion time for selected task is scheduled on that resource. The resource ready time is updated and except the entire unexecuted tasks are scheduled the procedure is repeated. Main problem of this algorithm it's chooses small tasks to be finished firstly, which in turn long task delays for very long time. Min-Min algorithm is did not utilize resources competently which lead to a load imbalance [1], [9].

E. *Max-Min Algorithm:*--Max-Min algorithm is close to equal as the min-min algorithm. The core difference among Min-Min and Max-Min algorithm is following: in this algorithm first finding out minimum execution times, then first the most extreme value is choose Then the performance time for all tasks is updated on that machine, this is done by adding the performance time of the assigned task to the performance times of other tasks on that machine. Then all assigned task is erased from the list that executed the system [9].

F. *Randomized:*--This is a static algorithm in nature. In this randomized algorithm a procedure may also be care of by a specific node n with a likelihood p. This algorithm functions admirably when every singleprocedure are of equivalent loaded. Issue emerges at the point when burdens are of various computational complexities. This randomized algorithm is not keeping up deterministic methodology [9].

G. *First Come First Serve Algorithm:*--In this algorithm jobs are served in the direction wherein they arrive i.e. Jobs are queued and served in structure of FIFO. This algorithm is discreet and really quick but doesn't provide so much effectively to job and resource optimization [3].

H. *Shortest Response Time First:*--In this algorithm every procedure is assigned a need which is permitted to run and equivalent need procedures are scheduled in FCFS demand. The SJF algorithm chooses the occupation with the most limited preparing time first. In this SJF algorithm shorter jobs are executed before long jobs. In this algorithm, it is significant to know or evaluation processing time of every job which is large SJF problem [9].

I. *Equally Spread Current Execution:*--This is an algorithm of dynamic load balancing, which handles the method with priority. It chooses the need by checking the scale of the system. This similarly spread present execution algorithm disseminates the load randomly by checking the extent of the system after then transferring load to a VM (Virtual Machine). On this algorithm load balancer spreads the load on to various nodes, so it's knows spread spectrum methodology [13].

J. *Resource Awareness Scheduling Algorithm*:--Resource Awareness Algorithm it's a blend of Min-Min and Max-Min algorithm and has no time consuming instruction. Thetime complexity of this algorithm is O (mn2) where n is number of jobs and m is number of resources [10].

V. LITERATURE REVIEW

Patel et al [14] provided a study on special mission scheduling algorithms and modification of min-min load balancing algorithm. The altered algorithm is founded on min-min process and venture rescheduling to use unutilized assets. It selects challenge with minimal completion time and assigns it to right assets for utilizing assets easily.SrinivasSethi et al [15] proposed an algorithm of load balancing creation fuzzy good judgment use in a cloud computing atmosphere. This presented a new fuzzy logic centered load balancing algorithm with additional parameters- memory usage, bandwidth utilization, disk house utilization and digital computerdevice position and named it as Fuzzy Active Monitoring Load Balancer (FAMLB).

Shridhar G. Domanal and G. Ram Mohana Reddy et al. [16] have been proposed a neighborhood enhanced load balancing methodology for distributing incoming job request uniformly between the servers or digital machines. They analyzed the performance of their algorithm making utilizing the Cloud Analyst simulator.N. J. Kansalet al. [17] have mentioned a number of the prevailing approaches which can be geared toward reducing the related overhead, provider response time and bettering efficiency of the method. The paper additionally presents important points about quite a lot of parameters, used to estimate the prevailing approaches.

Soumya Ray and A. D. Sarkar et al. [18] have mentioned particularly lots algorithms to load balancing like circular robin algorithm, imperative queuing algorithm and Randomized algorithm, their analysis is applied on MIPS vs. VM and MIPS vs. HOST groundwork. Their outcome show that these algorithms can in all likelihood reinforce the response time in order of magnitude, with admire to quantity of VMs in Datacenter. Execution judgment of the imitation indicates that the change of MIPS will have an effect on the response time. Increasing MIPS vs. VM reducing the response time. With the intention to control the random selection established load distribution main issue, the algorithm may also be carried out as the future course of labor to evaluate quite a lot of parameters.

Ankush P. Deshmukh and Prof. KumarswamyPamu et al. [19] discussed on one-of-a-kind load balancing strategies, algorithms and approaches. The research additionally suggests that the dynamic LB is more efficient than other static LB approaches.AlexandruIosup et al. [20] Analyzed the effectively of cloud computing facilities for scientific computing workloads and evaluated the vicinity in real experimental computing workloads of Many-assignment Computing (MTC) users, that's, of customers who rent loosely coupled functions includes many tasks to gain their scientific ambitions. In addition they perform an empirical evaluation of the performance of four business cloud computing offerings.

Zhang Bo et al. [21], proposed an algorithm which adds capability to the dynamic steadiness mechanism for the cloud. The experiments exhibit that the algorithm has received animproved load balancing degree and used less time in loading all duties.D. A. Menasce et al. [22] mentioned the cloud computing, that's advantages, and risks described several present cloud computing platforms. He additionally discussed the results of quantitative experiments applied using Planet Lab, cloud computing capacity planning and platform ways for cloud users and cloud carrier providers. To preserve the load balancing within the cloud computing method.

VI. CONCLUSION

In this paper, we have studied load balancing algorithms under cloud environment. In cloud computing, load balancing is a predominant task among the various defined approach based activity. It's requisite to allocate the dynamic local workload equally through the entire nodes to acquire an excessive user gratification and resource utilization ratio by making definite that every computing resource is distributed effectually and justly. In this paper, we have done the survey on the various and numerous load balancing algorithms within the cloud environment which are already proposed by various researchers and authors.

ACKNOWLEDGMENT

REFERENCES

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

[2]   Rajkumar Buyya, Rajiv Ranjan and Rodrigo N. Calheiros on "InterCloud: Utility-Oriented Federation of Cloud Computing Environments for Scaling of Application Services", ICA3PP 2010, Part I, LNCS 6081, pp. 13–31, 2010.

[3]   Sumit Dhaka and Mrs.Radhika Garg on"Review of resource scheduling in cloud computing", International journal of Enhanced Research in Management &ComputerApplications.ISSN:2319-7471, Vol.4Issue 7, July-2015.

[4]   Rajwinder Kaur and Pawan Luthra on "Load Balancing in Cloud System using Max Min and Min-Min Algorithm", International Journal of Computer Applications (0975 – 8887) National Conference on Emerging Trends in Computer Technology (NCETCT-2014).

[5]   Dharmesh Kashyap and Jaydeep Viradiya on "A Survey Of Various Load Balancing Algorithms In Cloud Computing",INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 3, ISSUE 11, NOVEMBER 2014.ISSN 2277-8616.

[6]   PS Yoganandani, Rahul Johari, Kunal Krishna, Rahul Kumar and Sumit Maurya on "Clearing The Clouds On Cloud Computing: Survey Paper", International Journal of Recent Development in Engineering and Technology Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)) Volume No.-1, Issue No.-1, July 2014)

[7]   Md. Imran Alam, Manjusha Pandey and Siddharth S Rautaray on "A Comprehensive Survey on Cloud Computing", I.J. Information Technology and Computer Science, 2015, 02, 68-79 Published Online January 2015 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijitcs.2015.02.09

[8]   Biju Ramesh on "Load balancing in Cloud Computing – An Analysis", International Conference on Security and Authentication - SAPIENCE14, ISBN 978-93-83459-32-2 © 2014 Bonfring.

[9]   Rajwinder Kaur and Pawan Luthra on "Load Balancing in Cloud Computing", Proc. of Int. Conf. on Recent Trends in Information, Telecommunication and Computing, ITC.

[10]   Amandeep, Vandana Yadav and Faz Mohammad on "Different Strategies for Load Balancing in Cloud Computing Environment: a critical Study", International Journal of Scientific Research Engineering & Technology (IJSRET), ISSN 2278 – 0882 Volume 3 Issue 1, April 2014.

[11]   Tushar Desai and Jignesh Prajapati on "A Survey Of Various Load Balancing Techniques And Challenges In Cloud Computing", INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 2, ISSUE 11, NOVEMBER 2013,ISSN 2277-8616

[12]   Dr. V. Venkatesa Kumar, Ms. R. Revathi and Mr. M. Newlin Rajkumar on "AN ASSESSMENT ON VARIOUS LOAD BALANCING TECHNIQUES IN CLOUD COMPUTING", IJAICT Volume 1, Issue 8, December 2014 ISSN 2348 – 9928, Doi:01.0401/ijaict.2014.08.01 Published on 05 (01) 2015

[13]   Komal Purba and Nitin Bhagat on "A Review on Load Balancing Algorithm in Cloud Computing", SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE) – volume1 issue10 Dec 2014, ISSN: 2348 – 8387

[14]   Gaurang Patel, Rutvik Mehta and Upendra Bhoi on "Enhanced Load Balanced Min-Min algorithm for Static Meta Task Scheduling in Cloud Computing", 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015) Available online at www.sciencedirect.com.

[15]   S. Sethi, A. Sahu, and S. K. Jena, on "Efficient load balancing in cloud computing using fuzzy logic,"IOSR Journal of Engineering, vol. 2, no. 7, pp. 65–71, 2012.

[16]   S. G. Domanal and G. R. M. Reddy, on "Load balancing in cloud computing using modified throttled algorithm," 2013.

[17]   N. J. Kansal and I. Chana, "Existing load balancing techniques in cloud computing: A system atic review.," Journal of Information Systems & Communication, vol. 3, no. 1, 2012.

[18]   S. Ray and A. De Sarkar, on "Execution analysis of load balancing algorithms in cloud computing environment.," International Journal on Cloud Computing: Services & Architecture, vol. 2, no. 5,2012.

[19]   A. P. Deshmukh and K. Pamu, on "Applying load balancing: A dynamic approach," International Journal, vol. 2, no. 6, 2012.

[20]   AlexandruIosup, S. Ostermann, M. N. Yigitbasi, R. Prodan, T. Fahringer, and D. H. Epema, on "Performance analysis of cloud computing services for many-tasks scientific computing," Parallel and Distributed Systems, IEEE Transactions on, vol. 22, no. 6, pp. 931–945, 2011.

[21]   Z. Bo, G. Ji, and A. Jieqing, on "Cloud loading balance algorithm," in Information Science and Engineering (ICISE), 2010 2nd International Conference on, pp. 5001–5004, IEEE, 2010.

[22]   D. A. Menasc´e and P. Ngo, on "Understanding cloud computing: Experimentation and capacity planning," in Computer Measurement Group Conference, 2009.

[23]   Kushwah, V.S. et al. (2014), "A Survey on Various Fault Tolerant Approaches for Cloud Environment During Load Balancing". Inter. Journal of Computer Networking, Wireless & Mobile Communications, ISSN (Print): 2250-1568; E-ISSN: 22789448, Vol. 4, Issue 6, December 2014, pp. 25-34.