IJCSEInternational Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-8, Issue-1, Jan 2020 E-ISSN: 2347-2693

Load Balancing Issues and Techniques In Cloud Computing

Roopali Gupta[1*], Meenu Dave[2]

[1,2]Dept. of Computer Science, Jagannath University, Jaipur, India

**Corresponding Author: roopaligoyal.gupta@gmail.com, Tel.: +91-9999437355*

DOI: https://doi.org/10.26438/ijcse/v8i1.137140 | Available online at: www.ijcseonline.org

Accepted: 20/Jan/2020, Published: 31/Jan/2020

Abstract— Cloud computing is growing rapidly due to its attractive features. Load on cloud is escalating rapidly due to the increase of new applications. An efficient use of cloud depends on several aspects such as security, speed, privacy etc. Load balancing ensures that every device and processor perform equal number of tasks in equal amount of time. Load balancing contribute to a reduction of resource consumption, set failover, enable expandability, avoid bottleneck and over-provisioning etc. So, this paper discusses basics of load balancing, issues, challenges, benefits, metrics, types and various techniques of load balancing in cloud computing.

Keywords— Resource Allocation, Cloud Computing, Load Balancing, Virtual Machines, CPU, QoS.

I. INTRODUCTION

Cloud computing defined by NIST is a model to facilitate ubiquitous, convenient, on demand network to admittance computing resources [1]. Cloud computing [2] is a set of resources provided to the various users as per their requirement. Cloud users [3] can access the services from anywhere and at any time and they need to pay as per their usage. Nowadays most of the companies shifting towards cloud services [4] dues to its effectiveness. AWS [5][6][7] cloud provider Amazon, Google cloud provider Google and AZURE cloud provider Microsoft all are providing powerful and reliable platform to users.

The major objective of Cloud computing [8] is to reduce the cost, improve response time and provide enhanced performance. It allows users to access [9][10] any type of hardware or software resources with the help of internet. Cloud computing offer services such SaaS (Software as a Service), IaaS (Infrastructure as a Service) and PaaS (Platform as a Service [10][11].

Cloud Computing system are depended on term virtualization. Virtualization [7] is process by which we can access number of different tasks on a single machine. The various services provided to cloud users via datacentres are based on the concept of virtualization.

The main issues of cloud computing [4] are Load Balancing, Virtual Machine immigration, Server Consolidation, Energy supervision, proficient resource consumption etc. as they are not completely addressed.

The most important issue among all the issues mentioned above is load balancing that is the mechanism which divides the workload evenly to all the nodes available in cloud. It helps in preventing bottleneck which happens due to load imbalance in the system. It facilitates by continuing service by implementing fail-over. So, this paper will discuss various aspects of load balancing by focusing on various issues and techniques in rest of the sections.

II. LOAD BALANCING IN CLOUD COMPUTING

Load balancing[7][12][13][14] is a method that redistributes the load among the diverse nodes within the network while not eliminating any of the running task to ensure equal distribution of load so that no 2 nodes is overloaded, under loaded or idle. Load balancing [15][16][17] may be a concern as it becomes difficult to accept various requests at a time. This problem referred as NP complete.

In cloud computing surroundings, load balancing offered in restricted servers at datacentre, if the request is more than the capability of datacentre than the general performance gets degraded. In such cases load balancer is employed for improve the performance of datacentre. The load may be any sort like network load, memory load, C.P.U load and delay load etc. [7]. Load balancing [18][19] [20] helps in speed up the execution of the application, to sustain QoS matrix delivered to the end customers and additionally ensures system stability.

Several characteristics of Load balancing as follows [10][21][22]: - equal distribution of work among all the nodes, help in achieving user satisfaction, enhance overall

performance of system, minimize the response time and provide services to realize complete resource utilization. Figure1 shows working of load balancing in cloud computing.

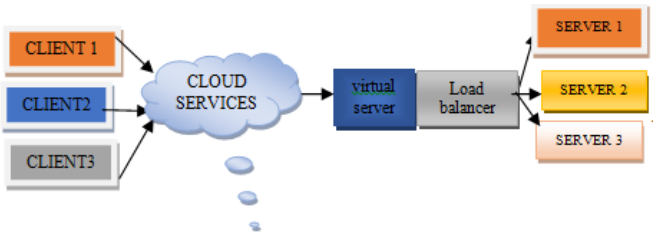

Figure 1. Load Balancing in cloud Computing

A load balancer [23] is a hardware that acts like a proxy to allocate network and application traffic across diverse servers. It works among server and client. It distributes the load among the servers based on different algorithms and a load balancer helps to develop the concurrent user capacity and overall consistency of applications.

III. LOAD BALANCNG BENEFITS

The various benefits of load balancing in cloud computing are [24]: -

1) **Scalability:** In load balancing algorithm we can add any number of servers at any time without any disturbance and still application runs smoothly.

2) **Performance**: An effective load balancing technique helps cloud services and application to respond faster than the regular completion time.

3) **Availability:** Load balancing ensures the service availability. In case of unavailability of small number of servers; the load is further distributed efficiently.

4) **Reliability:** The redundancy of the server which helps in keeping cloud services reliability protected. In case of failure also the cloud serving recourse will transfer services to other location.

IV. LOAD BALANCING METRICS

Various metrics are considered in existing load balancing techniques in cloud computing is discussed below [24][25][26][27]:

1) **Throughput:** It calculates maximum number of tasks whose execution completed. It must be maximum which is directly proportional to system performance.

2) **Overhead:** It calculates overhead incurred during load balancing technique implementation. It involves overhead due to progress of tasks, inter-process and inter-processor communication. The overhead cost should be least so that load balancing algorithm works efficiently.

3) **Fault Tolerance:** In case of failure of node or link still load balancing mechanism should not affect to all other nodes or link. The algorithm must be capable of handling good fault tolerance.

4) **Migration Time:** The time taken to switch from one job or resources to other one. It must be minimum to enhance the performance of the system.

5) **Turnaround Time:** The overall time taken by the system from request compliance to the response from the server is termed as turnaround time.

6) **Response time:** Total time system takes to serve the request submitted.

7) **Scalability:** It finds out how system is proficient of accomplishing load balancing algorithm with a limited number of machines and hosts.

8) **Make span:** It specifies the completion time once the resources assigned to the users.

9) **Degree of Imbalance:** It determines the imbalance between the virtual machines.

V. TYPES OF LOAD BALANCING ALGORITHM

The types of load balancing algorithm are defined below [28]:

1) **Static Algorithms:** They are non-pre-emptive in nature that is once the load is owed to the node it cannot be transferred to another node. It requires the former knowledge of the systems resources and not depends on the current system state. Load in the system cannot be changed once it starts execution. The major drawback of this approach is that this algorithm is appropriate for simulation and homogeneous system environment.

2) **Dynamic Algorithms:** They don't require former information or knowledge about the system resources as load distribution or job is assign based on current system state. It observes changes on the system and redistributes the load among the processors if the server is heavily loaded at runtime. In this communication overhead increases as the number of processors increases. It is suitable for heterogeneous cloud system.

3) **Sender Initiated Algorithm:** In this sender initiates the request message to transfer the data if the receiver is ready to accept the workload than sender will be sending the data to the receiver.

4) **Receiver Initiated Algorithm:** In this receiver sends the request message to transfer the data if the sender is ready to accept the workload than receiver will be sending the data to the sender.

VI. TECHNIQUES OF LOAD BALANCING

The various load balancing techniques or algorithm are discussed below [24][29][30][31]:

1) ***Round Robin Scheduling Algorithm:***

In round robin algorithm requests/jobs generated by the client and is being forwarded to the group of servers with the help of load balancer in round robin fashion that is time stamp is allotted to each request/job. Load balancer allocates equal jobs/request to the group of servers in circular fashion. If new requests come than the same process will be followed repeatedly. In case of failover request is forwarded to the other active servers. It is easy to implement and simple to understand.

2) ***Weighted Round Robin Algorithm:***

Weighted Round Robin algorithm is the extension of Round robin algorithm. In this we assign the weight to the servers and the server with greater weight is having more capacity. The capacity means CPU capacity, memory capacity and the other capacity of server in all respect. This means server with greater capacity is more powerful as its processing capability will be higher. According to the capability of server the incoming request will be allocated for processing and the load balancer is responsible for allocating the incoming client request to the server. So, the servers can process the request according to their capabilities. In case of new incoming request/job coming from the client will be allocated according to the weight of servers.

3) ***Least Connection Algorithm:***

In least connection algorithm requests from the client is allocated to the load balancer and the load balancer will forward the request to the server according to the existing connection on it. In case of new incoming request from the client will be allocated to the server with least connection.

4) ***Weighted Least Connection Algorithm:***

Weighted least connection algorithm is same as least connection algorithm with little difference. In this we assign weights to the server and assign the job alternatively based on their weights. In case of new request first job will be allocated to their server with least connection**.**

5) ***Random Algorithm:***

In random algorithm requests are connected according to the random number generated. The random number will be allocated to any server. In case of new request generated from the client the same process will be followed.

VII. CONCLUSION AND FUTURE SCOPE

In cloud computing load balancing is the major issues which is under consideration. With the help of load balancing we can calculate actual status of the cloud system and its effectiveness. In this paper we have discussed about the basics of load balancing its benefits, performance metrics, types of load balancing and its techniques. In future we will come up with an algorithm for enhanced behaviour of load balancing which will help the user in improved utilization by taking various parameters under consideration.

REFERENCES

[1] P. Mell and T. Grance." The NIST definition of cloud computing," National Institute of Standards and Technology, Information Technology Laboratory, Version **15**, **Oct. 7, 2009.**

[2] I. Foster, Y. Zha, I. Raicu and S. Lu," Cloud computing and grid computing 360-degree compared," 2008 Grid computing environments workshop, Austin, TX, **2008, pp. 1-10.**

[3] M. Ahmed, A. Sina, R. Chowdhury et all," An advanced survey on cloud computing and state-of-the-art research," 2012 International Journal of Computer Science,**2012 Volume 9.**

[4] R. Panwar, B. Mallick," Load balancing in cloud computing using dynamic load management algorithm",2015 International Conference on Green Computing and Internet of Things (ICGCIoT),**2015, pp.773-778.**

[5] S. Katoch and J. Thakur," Load balancing algorithms in cloud computing environment: a review," 2014 International Journal on Recent and Innovation Trends in Computing and Communication, Volume **2**, Issue 8, **Aug 2014**, pp.2151-2156.

[6] L. Mukati, A. Upadhyay," A survey on static and dynamic load balancing algorithms in cloud computing," 2019 International conference on Recent Advances in Interdisciplinary Trends in Engineering & Applications (RAITEA), Indore,**2019.**

[7] S. Joshi and U. Kumari," Load balancing in cloud computing: Challenges & issuies,"2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), Noida,**2016, pp.120-125.**

[8] K. Li, G. Xu, G. Zhao,Y. Dong and D. Wang ," Cloud task scheduling based on load balancing ant colony optimization," 2011 Sixth Annual Chinagrid Conference, Liaoning, **2011,pp. 3-9.**

[9] A. T. Velte, T. J. Velte, R. Elsenpeter, "Cloud Computing: A Practical Approach, "Tata McGraw Hill ,**2010.**

[10] M. Randles, D. Lamb and A. Bendiab, "A comparative Study into distributed load balancing algorithms for cloud computing,"2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops (WALNA), Perth,WA, **April 2010, pp.551-556.**

[11] HC. Hseih, ML. Chiang,,"The incremental load balance cloud algorithm by using dynamic data deployment," 2019 Journal of Grid Computing ,Volume **17**,Issue **3**, **September 2019**,pp.553-575.

[12] H. Shoja, H. Nahid and R. Azizi,"A comparative survey on load balancing algorithms in cloud computing,"2014 Fifth International Conference on Computing Communications and Networking Technologies(ICCCNT), Hefei,**2014, pp. 1-5.**

[13] R. Achar, P.S. Thilagam,N. Soans ,P.V. Vikyath, S. Rao and A.M. Vijeth," Load balancing in cloud based on live migration of virtual machines," 2013 Annual IEEE India Conference(INDICON),Mumbai,**2013,pp. 1-5.**

[14] E. Gupta, V. Deshpande," A technique based on ant colony optimization for load balancing in cloud data centre," 2014 International Conference on Information Technology (ICIT), Bhubaneswar,**2014, pp. 12-17.**

[15] S. Aslanzadeh and Z. Chaczko," Load balancing optimization in cloud computing: applying endocrine-particle swarm optimization,"2015

IEEE International Conference on Electro/Information Technology (EIT), Dekalb, IL, **2015, pp. 165-169.**

[16] F. Vhansure ,A. Deshmukh, and S.Sumathy,"Load balancing in multi cloud computing environment with genetic algorithm",2017 IOP Conference Series: Materials Science and Engineering, **2017,pp.1-10.**

[17] W. Sun, Z. Ji,J. Sun ,N. Zhang and Y. Hu, "SAACO: A self-adaptive ant colony optimization in cloud computing," 2015 IEEE Fifth International Conference On Big Data and Cloud Computing, Dalian, **2015,pp. 148-153.**

[18] S. G. Domanal and G.R.M. Reddy," Load balancing in cloud computing using modified throttled algorithm," 2013 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), Bangalore, **2013, pp.1-5.**

[19] K.R. Babu and P. Samuel," Enhanced bee colony algorithm for efficient load balancing and scheduling in cloud,"2016 6th International Conference on Innovations in Bio-Inspired Computing and Applications (IBICA), Kochi, **2016,pp.67-78.**

[20] K.R.R Babu and P. Samuel, "Service-level agreement-aware scheduling and load balancing of tasks in cloud,"2019 Software: Practice and Experience, **2019, pp.1-18.**

[21] R.G. Rajan and V. Jeyakrishnam , "A survey on load balancing in cloud computing environments," 2013 International Journal of Advanced Research in Computer and Communication Engineering,vol.2,Issue no. 12, **2013,pp.4726-4728.**

[22] S. Aslam and M. A. Shah," Load balancing Algorithms in Cloud Computing: A survey of modern techniques," 2015 National Software Engineering Conference (NSEC), Rawalpindi, **2015, pp.30-35.**

[23] G. Fan, L. Chen, H. Yu and D. Liu," Formally modelling and analysing cost-aware job scheduling for cloud data center,"2018 Software: Practice and Experience, Volume **48**, Issue ***9***, **2018**, pp.1536-1559.

[24] J. M. Shah, S. Pandya, N. Joshi, K. Kotecha and D.B. Choksi," Load balancing in cloud computing: Methodological survey on different types of algorithm," 2017 International Conference on Trends in Electronics and Informatics (ICEI), Tirunelveli, **2017, pp. 100-107.**

[25] Z. N. Rashid,S. R. M.Zebari , K.H. Sharif and K . Jacksi," Distributed Cloud Computing and Distributed parallel computing: A Review," 2018 International Conference on Advance Science and Engineering (ICOASE), Duhok, **2018, pp.167-172.**

[26] N. AI. Sallami, S. AI. Alousi," Load balancing with neural network",2014 International Journal of Advanced Computer Science and Applications, **2014, pp.138-145.**

[27] A.S. Milani, N.J. Navimipour," Load balancing mechanisms and techniques in the cloud environments: Systematic literature review and future trends",2016 Journal of Network and Computer Applications, **2016, pp.86-98.**

[28] J. Acharya, M. Mehta and B. Saini," Particle swarm optimization Based load balancing in Cloud Computing," 2016 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, **2016, pp.1-4.**

[29] V.N. Volkova, L.V. Chemenkaya, E.N. Desyatirikova, M. Hajali and A. Osama," Load balancing in cloud computing,"2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), Moscow, **2018, pp.387-390.**

[30] A. Vig , R.S. Kushwah and S.S. Kushwah ,"An Efficient Distributed Approach for Load Balancing in Cloud Computing,"2015 International conference on Computational Intelligence and Communication Networks(CICN),Jabalpur, **2015,pp.751-755.**

[31] C. Fancy, M. Pushpalatha and Pushpa," Experimentation of traditional load balancing algorithms in software defined network," 2019 International journal of Recent Technology and Engineering, **2019, pp.527-532.**

Authors Profile

Ms Roopali Gupta is a Ph.D scholar in Computer Science at Jagannath University Jaipur,Rajasthan.She received her M.Tech. in Computer Science degree from Jamia Hamdard University, Delhi in 2013.Her area of interest is load balancing in Cloud Computing .She published three papers in her area of interest.

Prof. (Dr.) Meenu Dave, M.Tech.., Ph.D.(Computer Science) has taught Computer Science in different capacities at a number of Engineering Colleges and Institutes. She has extensive experience in teaching Cloud Computing, Artificial Intelligence ,Knowledge Management and Data Mining at the post graduate level.She has also authored several research papers in the specified areas which have been published in leading journals.

