Comparative Study of Load Balancing Algorithms on Cloud

Shivangi Mayur, Nidhi Chaudhary
M-Tech (CSE) Student, Assistant Professor
Department of Computer Science and Engineering,
Meerut Institute of Engineering and Technology, Meerut, India,

Abstract: In modern times, cloud computing is one of the huge achievements. Cloud computing refers to the many different types of applications being delivered to the internet. Cloud computing is a variable technology that can support a broad spectrum of applications. The essential idea of cloud computing is to give a stage of sharing of assets which incorporates programming and foundation with the assistance of virtualization. Present paper deals with an extensive study on various literatures to conceptualize the process of load balancing under cloud environment where the main concept behind the cloud load balancing is to increase or decrease some parameters according to algorithms to increases the overall system performance. To conduct this study multiple research papers in the time frame of 2012 to 2016 has been collected and reviewed.

I. INTRODUCTION

Load balancing is the technique of distribution of workload among nodes in a distributed computing environment. An efficient load balancing algorithm will make sure that every node in the system does more or less same volume of work such that it ensures no node in the system is overloaded or sits idle for any instant of time. Cloud load balancing reduces cost and maximizes resource availability. Cloud load balancing is the way toward dispersing workloads over numerous figuring assets. Cloud load balancing decreases costs related with record the board frameworks and augments accessibility of assets. It is a kind of load balancing and not to be mistaken for Domain Name System (DNS) load balancing. While DNS load balancing utilizes programming or equipment to play out the function, cloud load balancing utilizes administrations offered by different PC organize companies. Figure indicates the cloud load balancing.

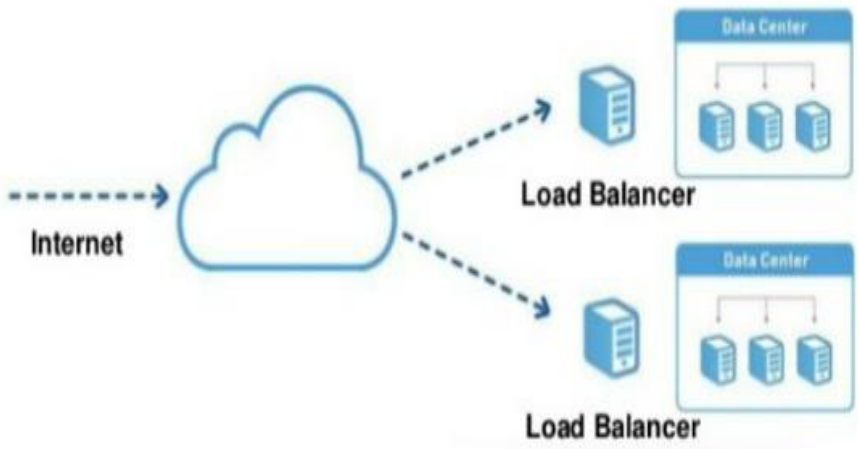

Figure 1: Cloud Load Balancing

II. LITERATURE REVIEW

An algorithm was developed by (Wei et al. 2014) [Ref. 18] where they focused on renewable energy sources in central. Different key factors like generation models, prediction methods of renewable power, capacity planning of green data centers, intra

datacenter tasks arrangement and load balancing across geologically spread data centers are employed in applying renewable energy in cloud data centers. The main objective was to diminish the energy costs incurred and to decrease the carbon discharge. The workload is distributed geographically across different data centers in a cloud environment.

A power optimization algorithm to balance the workload among the virtual machines effectively in order to show a trade – off between the performance and the energy consumption is developed by (Ankita and Upinder 2014) [Ref. 2]. Different types of algorithms like round robin, throttled load balancing, equally spread current execution algorithm are taken for analysis. A power consumption calculator is designed to monitor the energy utilization.

A load balancing algorithm is implemented to work in a heterogeneous cloud radio access network so as to diminish the peak transmission load in the remote radio heads of the mobile networks (Chen et al. 2015) [Ref. 5] . Balancing the workload in the backhauls is an important task in the mobile network. The load is balanced based on the nearest neighbor location. The experiments are conducted to prove that the load balanced in the backhauls is better than the existing compression methodologies prevailing in the mobile network.

INS - the index name servers was proposed by (Tin-Yu et al. 2015) [Ref. 17] to lessen the load because of the occurrence of duplicate files. The system focused on administering file storage, data de-duplication, selection of the best node and task allocation among servers. By carefully monitoring the presence of the files in the storage systems, the replications are avoided and the storage efficiency is improved thereby. The performance of such system is enhanced with efficient storage and access and the workload is balanced across the server in an effective manner. Data duplication rate and the loading rate across the clients are observed and the delay time in access is considerably reduced.

An algorithm where the multimedia workload is balanced across the different servers of a cloud network is proposed by (Chun-Cheng et al. 2014) [Ref. 6]. Based on the clients" requests the services are redirected to the appropriate servers to enhance the system performance by reducing the workload of the servers. Genetic algorithm is employed for the process of tasks assignment. Dynamic scenarios are generated for the purpose of experimentation and the results present a significant enhancement in the performance and throughput of the system.

A resource provision algorithm for Forwarding and Control Element Separation – ForCES network is developed by (Zhuge et al. 2014) [Ref. 19]. The algorithm is designed in such a way that the proper efficiency can be achieved if there source scheduling is executed based on the cost and the priority of the user applications. The experiment is conducted in a Cloudsim environment and the tasks completion time and costs are analyzed to present a better result.

A Stratus system developed by (Joseph et al. 2013) [Ref. 9] operates on Voronoi partitions to find out where to route the data center requirements based on the comparative precedence of the cloud operators. The system is designed to minimize the energy consumption thereby reducing the carbon emission and the cost as well by locating the nearest cloud servers to post the requests. Voronoi partitions are the formation of subsets and the graph model called stratus is designed to monitor and transfer the workloads. A pair wise partitioning rule is designated to choose the appropriate partitions. Experiments are investigated to analyze the cooling cost and the carbon emission ratio.

A traffic supervision technique described by (Chadi et al. 2014) [Ref. 4] managed load balancing using a useful VLAN mapping scheme. Column generation method is employed to determine the problem of mapping. The researchers have affirmed to produce an effective throughput by discovering the search space in less than 1% with an optimality gap of 4% when compared to the existing methodologies.

Two approaches named Case Based Reasoning (CBR) and rule based approach for proper resource scheduling to meet the quality of services requirements of the clients are developed (Michael et al. 2012) [Ref. 11]. The workload taken for the experimentation is the synthetic workload and the real world scientific workflow from the field of bioinformatics. The objective of the algorithms developed is to prevail over the SLA violations and to elicit a suitable resource allocation scheme to meet the QoS necessities.

A hybrid algorithm is implemented by (Moganarangan et al. 2016) [Ref. 12] integrating the logic of ACO and cuckoo search algorithms. This research work focuses on reducing the energy consumption in job scheduling task by including a makespan restraint. A heuristic approach of power aware task allocation is performed by using the logic of relative superiority (RS) of the processors. This heuristic approach refines a local optimum result. A Voltage Frequency Scaling (VFS) is used to estimate the processor energy thereby making a decision on the CPU utilization. Makespan model contemplates on effective job scheduling to the virtual machines in order to reduce the completion time of the tasks. The results show a significant improvement in the energy when compared to the ACO algorithm.

A research work called multi – objective nested Particle Swarm Optimization Technology for workload balancing is proposed by (Jena 2015) [Ref. 8]. The central idea is to perform an effective job scheduling so as to reduce the processing time in the cloud data centers thereby optimizing the energy. Multiple objectives like mutation operators, particles etc are used in the PSO optimization by taking various parameters of data centers, tasks arrival, length of the tasks, task completion time.

An effective load balancing where they have implicated fuzzy logic in the process is proposed by (Srinivas et al. 2012) [Ref. 16]. The main motive of this algorithm developed is to maintain a stable state across the virtual machines. The parameters considered are speed of the processor, number of virtual machines and the numbers of tasks allocated to each virtual machine. The internal parameters of the virtual machine and the data center are utilized for the experimental set up. The result proves to show a substantial progress in the data center processing time the overall execution time.

A load balancing algorithm using game theory is developed by (Mangal et al. 2014) [Ref. 10] to maintain a proper task scheduling across different virtual machines. The basic principle of cloud partition under various categories like normal, idle and heavy load conditions are done primarily. Different algorithms like random algorithm, weighted round robin, dynamic round robin is implemented for the idle mode partition. The concept of game theory which involves conflict and cooperation is utilized in the normal mode cloud partition to perform load balancing.

A flexible and energy aware structure for distribution of virtual machines in the data centers is proposed by (Corentin et al. 2012) [Ref. 3]. Technical constraints of different categories like Hardware, QoS, Availability etc are taken into account while performing the allocation of VM‟s in data centers. A power calculator is employed to predict the power consumption of every parts of the data center. The proposed system is assured to save the energy 18% more when compared to the existing methodologies.

TRACON - a Task and Resource Allocation Control Framework is proposed by (Ron and Howie 2014) [Ref. 15] for improving the system performance in a virtualized environment. The researchers have proclaimed that 25% improvement is proved in the throughput of the proposed system when compared to the existing approaches. Experiments are investigated to prove a significant improvement in terms of the number of tasks, execution time and the normalized throughput.

I – Aware (Interference Aware) strategy is developed by (Fei et al. 2014) [Ref. 7] to conduct virtual migration in a cloud computing environment. The main idea is to minimize the number of migrations while performing load balancing. The I/O and network throughput, CPU utilization and scalability are treated as parameters while conducting the experimental analysis. The results show a consistent improvement in the VM migration level and reduced VM interference.

A system is implemented to influence the NDN (named data networking) (Ruitao et al. 2015) [Ref. 14] to support VM migration in the cloud data centers in a flawless manner. The load balancing algorithm is intended to reduce the number of service interruptions. The load on different VM‟s is analyzed proportional to the hop counts of the request and the response. The system proves to optimize the performance by reducing the number of migrations and also the service disruptions.

An extensive research is conducted by (Abdulhussein et al. 2015) [Ref. 1] to apply genetic algorithm to perform load balancing in a cloud computing environment. By careful resource utilization the energy consumption is reduced to a proportional ratio. The goal is to reduce the VM migration cost. The experiment is conducted to analyze the performance of the virtual machines across the ESX servers in terms of the memory and CPU utilization. The yield of energy efficient load balancing when compared with the normal load balancing technique has proven to show a considerable improvement in the throughput of the system.

Dispersing the virtual machine resources proficiently (Mohammad and Atif 2014) [Ref. 13] with respect to the applications requirements is discussed which concentrated on cost cutting scheme in terms of energy and optimizing the number of servers in use. The researchers have shown that the proposed method can trim down the cost of servers on the whole, simultaneously pledge the fullest resource utilization and QoS demand in various proportions of server resources.

III. TABLE: Results & Description

REFERENCE NO.	TITLE	DESCRIPTION
18	Harnessing Renewable Energy	The main objective was to diminish the energy costs incurred and to decrease the carbon discharge.
2	Energy Efficiency	A power consumption calculator is designed to monitor the energy utilization.
17	De-Duplication and Feedback Schemes	Data duplication rate and the loading rate are observed and the delay time in access is considerably reduced.
6	Dynamic Multi-service LB in Cloud-Based Multimedia System	The results present a significant enhancement in the performance & throughput of the system.
4	Scalable Traffic Management in Cloud Data Center	The researchers have affirmed to produce an effective throughput by discovering the search space in less than 1% with an optimality gap of 4%.

12	Novel Algorithm for Reducing Energy Consumption	The results show a significant improvement in the energy when compared to the ACO algorithm.
8	Multi-Objective Nested PSO Technology for Workload	This technology is used to reduce the processing time in the cloud data centers.
16	Efficient Load Balancing Using Fuzzy Logic	The result shows a progress in data center processing time & overall execution time.
10	Partitioning Method & Game Theory	To maintain a proper task scheduling across different virtual machines.
3	Energy Aware Framework for Virtual Machine Placement	The proposed system is assured to save the energy more than18%.
1	Energy Efficient Load Balancing	The goal is to reduce the VM migration cost & improvement in the throughput.
13	Virtual Machine Resource Allocation	The proposed method can trim the cost of servers on the whole.

IV. CONCLUSION

Review of papers on Cloud load balancing shows that it can be implied in the distribution of the data resources across various virtual machines located at different locations. The main concept behind the cloud load balancing is to increase the throughput of the data transmission, sufficiently utilizing the available resources among the VMs and increasing the efficiency of the network. The performance of the cloud distributed system is gained through load balancing as the extended load is distributed among the underlying VMs. The virtualization technology in the cloud based systems has a scheduling system that act on the onset priorities. Cloud load balancing ensures the security of the cloud interconnected systems as the VMs provide services based on Software as Service and Product as Service as and when demanded. Therefore, the distributed network in the cloud ensures a security check before data transmissions.

V. FUTURE SCOPE

The results of the algorithms that are highlighted in this review paper may be enhanced by the researchers in future. Load balancing and security are the major research topics as Cloud is most frequently used now these days.

REFERENCES

[1] Abdulhussein Abdulmohson, Sudha Pelluri & Ramachandram Sirandas,2015, "Energy Efficient Load Balancing of Virtual Machines in Cloud Environments", International Journal of Cloud-Computing and Super- Computing Vol.2, No.1 (2015), pp 21-34.

[2] Ankita Sharma, Upinder Pal Singh, 2014, "Energy Efficiency in Cloud Data Centers Using Load Balancing", International Journal of Computer Trends and Technology, Volume.11, No.4, May 2014 pp: 174 – 181.

[3] Corentin Dupont, Giovanni Giuliani, Fabien Hermenier, Thomas Schulze & Andrey Somov, 2012, "An Energy Aware Framework for Virtual Machine Placement in Cloud Federated Data Centers", e-Energy 2012, May 9-11 2012, ACM.

[4] Chadi Assi, Sara Ayoubi, Samir Sebbah & Khaled Shaban, 2014,"Towards Scalable Traffic Management in Cloud Data Centers", IEEE Transactions On Communications, Vol. 62, No. 3, March 2014, pp 1033 – 1045.

[5] Chen Ran, Shaowei Wang, & Chonggang Wang, 2015, „Balancing Backhaul Load In Heterogeneous Cloud Radio Access Networks", IEEE Wireless Communications, June 2015, pp 42 – 48.

[6] Chun-Cheng Lin, Hui-Hsin Chin & Der-Jiunn Deng, 2014, "Dynamic Multi service Load Balancing in Cloud-Based Multimedia System", IEEE Systems Journal, Vol 8, No.1, March 2014, pp 225 – 234.

[7] Fei Xu, Fangming Liu, Linghui Liu, Hai Jin, Bo Li & Baochun Li, 2014, "IAware: Making Live Migration of Virtual Machines Interference-Aware in the Cloud", IEEE Transactions on Computers, Vol 63, No.12, December 2014, pp 3012 – 3025.

[8] Jena, R. K. (2015). Multi objective task scheduling in cloud environment using nested PSO framework. Procedia Computer Science, 57, 1219-1227.

[9] Joseph Doyle, Robert Shorten, & Donal O"Mahony , 2013, "Stratus: Load Balancing the Cloud for Carbon Emissions Control", IEEE Transactions on Cloud Computing, Vol.1, No.1, January 2013, pp 116– 128.

[10] Mangal Nath Tiwari , Kamalendra Kumar Gautam , Dr Rakesh KumarKatare , 2014, "Analysis of Public Cloud Load Balancing using Partitioning Method and Game Theory", International Journal of Advanced Research in Computer Science and Software Engineering,Volume 4, Issue 2, February 2014, pp: 807 – 812.

[11] Michael Maurer, Ivona Brandic, Rizos Sakellariou, 2013, "Adaptive resource configuration for Cloud infrastructure management", Future Generation Computer Systems 29 (2013), pp: 472 – 487.

[12] N, Moganarangan, R,G, Babukarthik, S, Bhuvaneswari, M,S, Saleem Basha, P, Dhavachelvan, "A novel algorithm for reducing energy consumption in cloud computing environment: Web service computing approach", Journal of King Saud University – Computer and Information Sciences (2016) 28, pp: 55–67.

[13] Mohammad Mehedi Hassan, Atif Alamri, 2014, "Virtual Machine Resource Allocation for Multimedia Cloud: A Nash Bargaining Approach", International Symposium on Emerging Inter-networks, Communication and Mobility (EICM-2014), Procedia Computer Science 34 (2014) pp: 571 – 576.

[14] Ruitao Xie, Yonggang Wen, Xiaohua Jia & Haiyong Xie, 2015, "Supporting Seamless Virtual Machine Migration via Named Data Networking in Cloud Data Center", IEEE Transactions on Parallel and Distributed Systems, Vol.26, No.12, December 2015, pp 3485 – 3497.

[15] Ron C, Chiang & H. Howie Huang, 2014, "TRACON: Interference-Aware Scheduling for Data-Intensive Applications in Virtualized Environments", IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 5, May 2014, pp 1349 – 1357.

[16] Srinivas Sethi, Anupama Sahu, Suvendu Kumar Jena, 2012, „Efficient load balancing in cloud computing using fuzzy logic", IOSR Journal of Engineering (IOSRJEN) ISSN: 2250-3021 Volume 2, Issue 7(July2012), PP 65-71.

[17] Tin-Yu Wu, Jeng-Shyang Pan & Chia-Fan Lin, 2014, "Improving Accessing Efficiency of Cloud Storage Using De-Duplication and Feedback Schemes", IEEE Systems Journal Vol 8, No.1, March 2014pp 208 – 218.

[18] Wei Deng, Fangming Liu & Hai Jin, 2014, "Harnessing Renewable Energy in cloud data centers: opportunities and challenges" , IEEE Network, Jan / Feb 2014, pp: 48 – 55.

[19] Zhuge Bin, Deng Li, Dai Guowei, Wan Lei, Wang Weiming, Lan Julong, 2014, "Resource Scheduling Algorithm and Economic Model in ForCES Networks", Network Technology and Applications, China Communications, March 2014, pp 91 – 103.