

Reducing Load At Data Center Using Load Balancing Algorithms

¹Maney Saketh Naidu, ²P Ramakrishna

^{1,2}Dept. of CSE, KIET, Kakinada, AP, India

Abstract

Despite the many past research conducted in the Cloud Computing field, some challenges still exist related to workload balancing in cloud-based applications and specifically in the Infrastructure as service (IaaS) cloud model. Efficient allocation of tasks is a crucial process in cloud computing due to the restricted number of resources/virtual machines. IaaS is one of the models of this technology that handles the backend where servers, data centers, and virtual machines are managed. Cloud Service Providers should ensure high service delivery performance in such models, avoiding situations such as hosts being overloaded or underloaded as this will result in higher execution time or machine failure, etc. Task Scheduling highly contributes to load balancing, and scheduling tasks much adheres to the requirements of the Service Level Agreement (SLA), a document offered by cloud developers to users. Important SLA parameters such as Deadline are addressed in the LB algorithm. The proposed algorithm is aimed to optimize resources and improve Load Balancing in view of the Quality of Service (QoS) task parameters, the priority of VMs, and resource allocation. The proposed LB algorithm addresses the stated issues and the current research gap based on the literature's findings. Results showed that the proposed LB algorithm results in an average of 78% resource utilization compared to the existing Dynamic LBA algorithm. It also achieves good performance in terms of less Execution time and Make.

Keywords

Cloud Computing, Virtual Machines, Load Balancing, Broker Policy, Performance Evaluation.

I. Introduction

Cloud computing is the on-demand delivery of applications, compute power, database storage, and other IT resources through a cloud services platform via the internet with pay-as-you-go pricing [1]. It is an emerging field for research and study. It is a pool of multiple configurable computing resources available on demand to user. It has evolved from past technologies like web services, hardware virtualization, grid and utility computing, system management. Cloud computing mainly focuses to give maximum numbers of shared resources and support for user requests in actual time. The cloud services are of three classes: Infrastructure as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). Cloud computing has many issues such as load balancing, cloud security, management of energy, privacy which hinder its growth. Load balancing issue is handled using algorithms which helps to distribute load between all the nodes. It also ensures that every computing resource is distributed efficiently and fairly. It helps in preventing bottlenecks of the system which may occur due to load imbalance. Load balancing provides better response time and high resource utilization [2]. This paper compares performance of round robin, equally spread current execution and throttled load balancing algorithms along with closest datacenter, optimize response time and reconfigure dynamically with load service broker policies and evaluate them on the basis of overall response time and data processing time.

Load Balancing is a method which redistributes the workload with system nodes, to improve resource consumption and system performance. Load Balancing is taken in to account so that all virtual machine gets equal quantity of workload which increases throughput and reduces response time. Load balancing is a technique that helped systems and assets by giving a maximum throughput and less response time. Load balancing separates the traffic between all servers, so information can be sent and get back immediately with load balancing [3].

The organization of the paper is as follows: In section II review of some related work is done. In section III, load balancing algorithms are discussed. In section IV Service Broker policies are discussed. Section V consists of implementation details. Section VI discusses results of the implementation and Section VII, concludes the paper.

II. Literature Survey

Jadeja, Y. and K. Modi [4] reviews the architecture of cloud computing, its benefits and issues such as security, privacy etc. and some of its major applications.

Jing Yao & Ju-hou He [5] discusses about architecture plan of cloud computing where cloud computing framework are divided in to two parts that is front-end & back-end. Both are connected through the internet. Front end is visible to users and back end is for cloud framework. Front end consist of client's computer accessed by the cloud, where as back end gives the 'cloud computing services' like storage, computers etc. It also discusses about the services and layers provided by cloud computing design which are Software as a Service, Platform as a Service, and Infrastructure as a Service and some issues related to privacy, security, reliability etc.

Khiyaita et al. [6] provides the definition and taxonomy of load balancing. They described the different implementations of load balancing in most used distributed systems. They also mentioned the major challenges of load balancing in cloud computing. Shreya Purohit [7] provides an in depth study of the factors favoring cloud computing, reviewing various cloud deployment and service models. It considers security, privacy, and internet dependency and availability as challenges of cloud computing. The author inspects certain benefits of cloud computing over traditional IT service environment including adaptability, higher resource usage, reduced capital, and scalability which are considered as reasons for switching to cloud computing environment. It considers vertical scalability as technical challenge in cloud computing.

Bhathiya, Wickremasinghe [8] has discussed the detailed functioning of GUI based tool called as Cloud Analyst which was developed to simulate large-scale Cloud applications for studying the behaviour of such applications under various deployment configurations. Developers uses Cloud Analyst which helps in understanding how to spread applications among Cloud infrastructures and use value added services such as performance optimization of applications and providers incoming with the use of Service Brokers.

S. Mohapatra et al. [9] discussed a performance comparison for different load balancing algorithms of virtual machine and policies in cloud computing. In this study performance of four well

known load balancing algorithms namely First Come First Serve, Execution Load, Round Robin and Throttled Load Balancing Algorithms have been analyzed based on the average response time, average datacenter request servicing time and total cost. The simulation results according to the CloudAnalyst simulator show that round robin has the best integration performance.

IsamAzawiMohialdeen [10] discusses about various scheduling policies. Author does a comparative study of scheduling algorithms in cloud computing and explains their requirement in cloud environment.

Singh, A. et al. [11] develops an alternative method for round robin scheduling which improves the CPU efficiency in real time and time sharing operating system. The algorithm proposed by author improves all the snags of simple round robin architecture. He has also done a comparative analysis of simple round robin scheduling algorithm with proposed algorithm. The proposed algorithm increases the system throughput and solves the problem faced in simple round robin architecture by decreasing the performance parameters to desirable extent.

V. Behaland A. Kumar [12] provide a comparative study of Round Robin and Throttled virtual machine load balancing algorithms has been proposed. Both the algorithms are used with optimized response time service broker policy and simulation is performed to calculate overall response time, datacenters hourly average processing times, response time according to region, datacenters request servicing time, user base hourly response times and total cost which has significant effect on performance. According to the simulation results, the combination of the proposed strategy of throttled and optimized response time service broker policy has the better performance than round robin load balancing algorithm in heterogeneous cloud computing environment.

M. Nitika et al. [13] addressed execution of three load balancing algorithms examined the inadequacies and researched why it is unrealistic to have Centralized Scheduling policy during the cloud condition. Author inspected three possible solutions which are Honeybee Foraging Behaviour algorithm, Random Sampling algorithm and Active Clustering algorithm proposed for load balancing.

K Nishant et al. [14] have proposed an algorithm which is a modified approach of ant colony optimization that has been applied from the view of cloud network systems with the purpose of load balancing of nodes. It is different from the original approach in which each ant builds own result set and later builds a complete result set. However in their approach the ants update a single result set rather than their individual result set. This approach detects overloaded and under loaded nodes and thereby perform operations based on the identified nodes. The task of each ant is specialized rather than being general and the task depends on the type of first node which was encountered whether it was overloaded or under loaded.

III. Load Balancing Algorithms

A. Round Robin Algorithm (RR)

Round Robin is one of the traditional widely used algorithms. In round robin policy, the time slices are allotted to each task in uniform proportion and in circular fashion. Each task is allotted to available virtual machine in circular order. This policy is not considered as priority intended scheduling policy. In it, situation occurs where some nodes are massively loaded and some are slightly loaded. This leads to situation where system load gets imbalance [15].

B. Equally Spread Current Execution Algorithm (ESCE)

ESCE algorithm balances the tasks among available Virtual machines in a way to even out the number of active tasks at any given time on each Virtual Machine. ESCE algorithm handles the system workload with priorities [13]. ESCE distributes the datacenters workload randomly by checking the size and transfer the load to that virtual machine which is lightly loaded. This algorithm finds the VM with least number of allocations and in a way that the number of active tasks on each VM is kept evenly distributed among the VMs. C. Throttled Algorithm Throttled algorithm initiates by assigning favourable virtual machine when customer sends request to load balancer. The role of load balancer is to look after an index table of all virtual machine together with their states depicting busy and available mode. At start, all virtual machines are set to available mode. The datacenter controller consults balancer for next virtual machine allocation, when it receives a new request. The balancer start checking table thoroughly until a relevant match of virtual machine is found. If favourable virtual machine is found then the balancer returns id of that particular virtual machine to datacenters controller. At that instant, datacenters controller sends request to virtual machine identified by that particular id. After that, datacenters controller sends notification to the balancer of new allocation so that it can update the table. If there's a case, when virtual machine is not found, then the balancer returns -1 value and datacenters queues the request. As soon as virtual machine finishes with the processing of the assigned request, later the datacenters controller receives a response cloudlet and it sends the notification to balancer to virtual machine de-allocation [8] [16].

IV. Service Broker Policies

A. Closest Datacenter Policy

In closest datacenter policy the datacenter which is having least proximity from the user is selected. If more than one Datacenters having same proximity then it will select datacenter randomly to balance the load, here proximity is in term of least network latency.

B. Optimize Response Time Policy

First, it identifies the closest datacenter using previous policy but when Closest Datacenter's performance (considers response time) starts degrading it estimates current response time for each datacenter then searches for the datacenter which has a least estimated response time. But there may be 50:50 chances for the selection of closest and fastest datacenter.

C. Dynamically re-configurable routing with load

This is an extension to Closest Datacenter Policy where the routing logic is similar. But it has one more responsibility of scaling the application deployment based on the load it is facing. It also increases or decreases the no. of VMs accordingly. This will be done taking into consideration the current processing times and best processing time ever achieved. This policy is under research so it gives vague results.

V. Implementation Details

CloudAnalyst is used for constructing and analysis of large-scale cloud computing environment. It comprises of three vital menus namely, Configure Simulation, Define Internet Characteristics and Run Simulation [17]. These menus are used for framing the whole simulation procedure. Cloud Analyst v1.0 beta is used to

evaluate three algorithms which are Round Robin, ESCE and Throttled load balancing algorithm. Results may vary with some other version of CloudAnalyst. Simulation is done for six user base identified as UB1, UB2, UB3, UB4, UB5, UB6 with different regions for each user. In this simulation four datacenters are taken into addition which is named as DC1, DC2, DC3, and DC4. Cloud Analyst is open resource toolkit which simulates and evaluates various cloud services. Cloud analyst allows modeler to simulate simple experiments with a little variation in parameters in a swift way. Cloud analyst is equipped with a feature called GUI. It is an extended version of CloudSim.

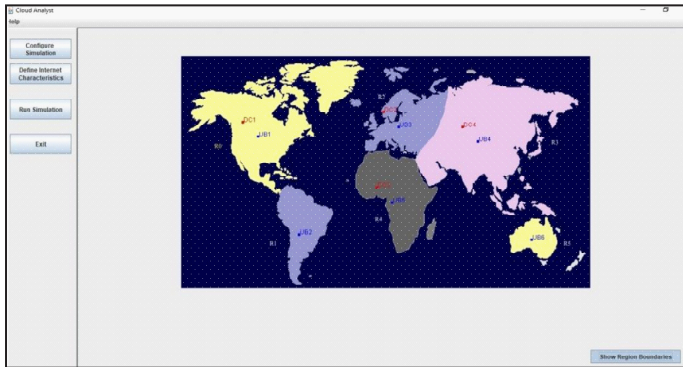


Fig. 1: Cloud Analyst

Figure 1 shows the simulated picture in CloudAnalyst simulator. Datacenter DC1 is placed in region R0 and has user base UB1. There is no datacenter placed in R1 but has one user base UB2. Region R2 has datacenter DC2 and user base UB3. Region R3 has datacenter DC4 and user base UB4. Region R4 has datacenter DC3 and user base UB5. Region R5 has no datacenter but has user base UB6. By this type of scenario it can simulate all possible ways for simulation process. The four different datacenters identified as DC1, DC2, DC3, DC4 having 25, 50 and 75,100 numbers of VMs respectively.

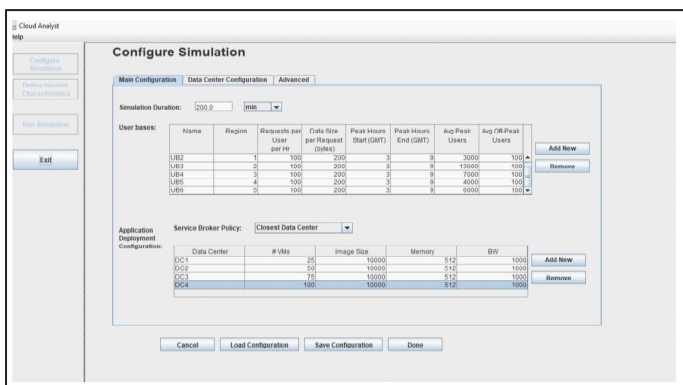


Fig. 2 shows the configuration simulation consisting of simulation duration, user bases and application deployment configuration. In Simulation Duration, it defines the interval taken by the simulation which is in minutes, hours and days. User bases table define a no of users along with their regions, request per user, data size per request etc. In Application deployment configuration, service broker policy provides 3 different policies namely closest datacenter, optimize response time and reconfigure with dynamically with load to be chosen. It also contains the details of the data like storage size, memory, bandwidth and virtual machines owed to each datacenter.

Figure 3 shows the datacenter configuration which defines a list of datacenters with the information like regions, OS, VMM, Cost

per VMM, Storage cost etc. It also provides the details of the physical hardware of the datacenters with the information like Id, Memory, Storage, No of processors etc.

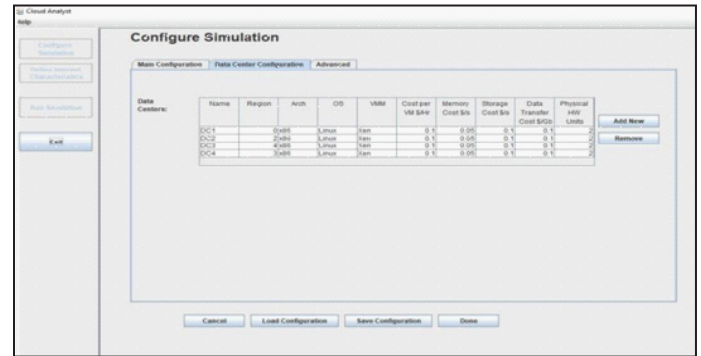


Figure 3: datacenter configuration

The other parameters are fixed as shown in Table 1.

Table 1 parameters and their values

Parameter	Value Passed
VM-Image size	10000 Mb
VM-Memory	512
VM-bandwidth	1000 Mb
Service Broker Policy	Closest datacenter Optimized response time Dynamically re-configurable routing with load policy
Datacenter architecture	x86
Datacenter-OS	Linux
Datacenter-VMM	Xen
Datacenter-No. of VMs	DC1-25,DC2-50,DC3-75 DC4-100
Datacenter-memory per machine	2GB
Datacenter-storage per machine	1TB
Datacenter-available bandwidth per machine	1000000 bit/s
Datacenter-processor speed	10000 MIPS
Datacenter-VM Policy	Time shared
User grouping factor	
(Equivalent to number of simultaneous users from a single user base)	100
Request grouping factor	
(Equivalent to number of simultaneous requests a single application server instance can support)	100
Executable instruction length	1000 bytes

VI. Conclusion

This section concludes the paper by highlighting the findings and obtained results from the proposed LB algorithm. As we saw from the literature, task scheduling highly contributes to balancing the load in a cloud environment. Improving the Load Balancing process through Task Scheduling can result in efficient utilization

of cloud resources. The objective of this paper was to provide an enhanced Load Balancing algorithm. Results proved that our algorithm reduces Makespan and provide efficient resource utilization of 78% compared to existing Dynamic LBA. It also shows that the proposed algorithm can function in a dynamic cloud environment where user requests arrive in random order and where there are many changes in the length of the user requests. The algorithm is also able to handle large size requests compared to the existing approach. The algorithm address SLA violation of VMs by reallocating resources to execute tasks efficiently. In the future, authors will work to optimize the cloud resources further and enhance cloud-based application performance, such as considering more SLA parameters. For example, the algorithm will be tested based on the number of violations and the migration count for better performance. Also, the algorithm will be comprehensively compared to other existing algorithms in the literature.

References

- [1] "What is Cloud Computing?", Amazon Web Services. Retrieved 2018-04-13.
- [2] YogeshBanyal and Jawahar Thakur, "Load Balancing Algorithms in Cloud Computing: Review", International Journal of Advance Engineering and Research Development (IJAERD), Volume-05 Issue-03 pp 427-433, 2018.
- [3] Martin Randles, David Lamb, A. Taleb-Bendiab, "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing", IEEE 24th International Conference on Advanced Information Networking and Applications Workshops, 2010.
- [4] Y. Jadeja and K. Modi, "Cloud Computing - Concepts, Architecture and Challenges", International Conference on Computing, Electronics and Electrical Technologies (ICCEET), 2012.
- [5] Jing Yao and Ju-hou He, "Load Balancing Strategy of Cloud Computing based on Artificial Bee Algorithm", ISBN: 9781-4673-0893-9, Pages: 185-189, IEEE, April 2012.
- [6] A. Khiyaita, M. Zbakh, H. El Bakkali and Dafir El Kettani, "Load Balancing Cloud Computing : State Of Art", IEEE, 2012.
- [7] Shreya Purohit, "An Exhaustive Study on Cloud Computing," Volume II, Issue I, ISSN NO. 2454-5678, 2016.
- [8] Bhathiya, Wickremasinghe,"Cloud Analyst: A Cloud Sim-based Visual Modeller for Analysing Cloud Computing Environments and Applications", IEEE, 2010.
- [9] S. Mohapatra, K. SmrutiRekha, and S. Mohanty, "A Comparison of Four Popular Heuristics for Load Balancing of Virtual Machines in Cloud Computing", International Journal of Computer Applications, Volume 68-Issue 6, pp. 33-38, April 2013.
- [10] IsamAzawiMohialdeen, "Comparative Study of Scheduling Algorithms in Cloud Computing Environment", Journal of Computer Science, Volume 9 Issue 2, 2013.
- [11] A. Singh, P. Goyal, and S. Batra, "An Optimized Round Robin Scheduling Algorithm for CPU Scheduling", International Journal on Computer Science and Engineering (IJCSE), Volume 2 Issue 7,pp. 2383-2385, 2010.
- [12] V. Behal and A. Kumar "Cloud computing: Performance analysis of load balancing algorithms in cloud heterogeneous environment", Confluence the Next Generation Information Technology Summit (Confluence), 5th International Conference, 2014. [13] M. Nitika, M. Shaveta, and M.G. Raj, "Comparative analysis of load balancing algorithms in cloud computing", International Journal of Advanced Research in Computer Engineering & Technology, 2012. 1(3): p.120-124.
- [14] Nishant Kumar, Pratik Sharma, Vishal Krishna, Chhavi Gupta, KuwarPratap Singh, Nitin, and Ravi Rastogi, "Load Balancing of Nodes in Cloud Using Ant Colony Optimization", 14th International Conference on Modeling and Simulation, UKSim, pp. 3–8, IEEE,2012.
- [15] Ajit Singh, PriyankaGoyal and SahilBatra, "An optimized round robin scheduling algorithm for CPU scheduling", International journal of computer and electrical engineering (IJCEE), vol. 2, No. 7, pp. 2383-2385, 2010.
- [16] Ram Prasad Padhy and P Goutam Prasad Rao , "Load balancing in cloud computing system", Department of Computer Science and Engineering National Institute of Technology, Rourkela-769 008, Orissa, India May,2011.
- [17] Bhasker Prasad Rimal, Eummi Choi and Lan Lump (2009) "A Taxonomy and Survey if Cloud Computing System" 5Th International Joint Conference on INC, IMS and IDC, NCM '09, 25-27 Aug. 2009, pp. 170-175.