A Critical Study for Managing Applications Efficiently Using Load BalancingApproach In context with Cloud ComputingEnvironment

[1]Shashikant Raghunathrao Deshmukh, [2] Dr. S.K Yadav, [3] Dr. D.N Kyatanvar

[1]Research Scholar, Shri JJT University, Chudela, JhunJhunu, Rajasthan, India shashikantraghunathrao@gmail.com
[2]Professor,Shri JJT University Chudela, JhunJhunu, Rajasthan, India
[3]Professor,Shri JJT University Chudela, JhunJhunu, Rajasthan, India

Article History: Received: 10 December 2019; Revised: 12 January 2020; Accepted: 27 Feb 2020; Published online: 28 April 2020

Abstract
The process of distribution of workloads as well as computational resource(s) in a cloud computing environment is known as cloud load balancing. Load balancing assists businesses to accomplishrequests or assignmentburdens by distributingavailable resources through numerous systems, networks, or servers. The crucial objective of load harmonising is to keep any single server from becoming burdened and probably failing. In supplementary arguments, load balancing increases provision convenience and aids in the prevention of outages. Load balancing is well-defined as the systematic and well-organizeddissemination of network(s) or application(s)loadthroughseveral servers in the server farm. Every load balancers presence among client devices and background servers. Various incoming requests are received and later distributedsame to whicheverobtainable server having capability of accomplishing them.Different categories of load balancing algorithms in cloud computing such as static algorithms, dynamic algorithms, round-robin algorithms, weighted round robin load balancing algorithms, opportunistic load balancing algorithms, min to min load balancing algorithms, max to min load balancing algorithms.

Keywords: Cloud Computing, Load Balancing, Workload.

1. Introduction

The cloud computing is the distribution of dissimilar services through the internet. The resources includes tool(s) and application(s) such as data storage, servers, databases, networking, and softwares. As long as an electronic devices has access permission to the internet, it has also getting access to the informationas well as the software's program to execute it. In simple way one can explain about cloud computing that , it is the provision of computing (execution) services which includes servers, storage, databases, networking, software, analytics, and intelligence over the internet ("the cloud") for offeringquicker innovation, flexible resources, and financial prudence of measure. Cloud computing is an application based software infrastructure which stores data on serves remotely.The servers can be accessed with the help of the internet with access permissions by the administrator. The front end applications enables any authentic user to get access for the data which is stored on the cloud using secure internet browser or any valid cloud computing application software.

One can describe cloud load balancing as the method of distributingavailable workloads along with the computing properties in a cloud computing environments. It also responsible for enablingthe respective organizationfor managing the available not only the workload demands but also the demands for applications by distributing existing available and free resources among numerous systems, networks or servers. Hence the load balancing in cloud which includes holding the rotation of workloads traffic and demands that exist over the internet by various applications as well as clients' requirements.

One can also describe about the concept of cloud computing as it is one of the innovativestandard of demonstratingcapabilities for computing as a service. Two of its important facilities such as sharing of the resources and cost effective as well. Itoccurs in eachfield of life, augmenting their functionality and accumulating new opportunities to it. Consequently, the emphasisis on resolving its dilemmas such as balancing of the load becomingadded challenging and the exploration in swarm-based algorithms to discoverthe finestoutcomeswhich has been intensifying exponentially because of increase in demands of cloud based applications and services.

As the load on the internet increasingexponentially, which is about 100% yearly of the current traffics load. Henceforward, the workload on the server increasing very fast leading towards the overfilling of servers primarily for standard and widely used web servers. There are couple offundamentalresolutions to get out of the problems of overfillinghappeningwith the servers in action. In case of the single server determination in which the server is improved to a greater performance server. Though, the new server might also be get overloaded with increase in demands which is mere a challengingsituation for addition of resources for balancing the load that means upgrading is very much essential and required for better results. Additionally, the advancementprocedure is difficult and costly.There is a multiple server resolution in which a scalability service systems on a collection of servers is constructed. This results into theadditional expenses incurredand get involved as well as further scalability for building a server constellation system which will be required for the service(s) which are based on existing network.
In terms of productivity and investment, cloud computing delivers severalbenefits for both cloud handlers and cloud service providers. Though, the presence of different services and a large number of applicants complicates the management of workload in cloud servers. Furthermore, task scheduling and resource allocation are established in cloud systems to manage the work load; however, determining the resources left in the physical or VM node at the time of task assignment makes it more complex. As a result, load balancing in cloud systems has become critical for achieving reliable performance, and numerous methodologies have been developed to address this issue.The development of optimization models aided in the resolution of load balancing issues. Load balancing is typically used to accelerate the computation of resources on physical or virtual nodes that vary unconditionally over time.
This paper describes a critical study of cloud load balancing insights not only in the context of the cloud but also reported various algorithmic schemes. The paper also focused on various types of cloud load balancing along with the load balancers. The significance of load balancing in clod computing is also thoroughly discussed.

In the context of cloud load balancing, itpermits for creation of a suitable balance of the workload performed on various hardware equipment and devices. The devicesworkloads balancing across servers or within a single cloud server between the central processing units (CPUs) and hard drives. The intention behind the introducing the cloud load balancing was to increase the execution speedas well as the improvement in the performance of each devices, as well as to prevent devices individuallyfrom increasing respective performances as per the expected threshold of parameters.

1.1 Types of Load Balancing

Various types of load balancing that one should have awarenessregarding network. Server load balance concept is for relational databases. The global server load balance is recommended to troubleshoot across numeroustopographicallocalities, also the Domain Name Systems (DNS) load balancing make sure about the functionalities of domain name. The Load balancers which are available in the cloud might also be preferred for load balancing approach.

a. Load Balancing for Networks

The cloud load balancing influences network layer data and leaves it to resolvewherever network traffic should be routed. As per as Transmission Control Protocols (TCP) / User Datagram Protocols (UDP) traffic is taken care by Layer 4 load balancing. Thesolution is one of the best with respect to local balancing, but it is not able to balance traffic distribution across servers.

b. Load Balancing at Layer 7

One of the very old type of load balancing is Hypertext Transfer Protocol (HTTPs) load balancing. This type of load balancing is devised based on layer seven of protocol stack. This means that load balancing takes place at the layer of operation. This is the most adaptable type of load balancing

because it permits for making distribution decisions based on information gained from the address i.e. HTTP address.

c. Internal Cloud Load Balancing

It is analogous to network load balancing, this type of load balancing is used for balancing the internal infrastructure.

1.2 Load Balancers Insights

Load balancers are further classified as hardware, software, or virtual.

a. Hardware Load Balancer

This is reliant on the premise and the physical hardware to allocate network and applications traffic. The devices arein capable of handling high traffic volumes. Unfortunately this kind of approach is expensive as compared to the limited flexibility which degrades the speed in turn decrease in performance.

b. Software Load Balancer

As compared to hardware solutions, the load balancing based on software's are cheaper. The nature of the balancer is open source that means one can customise the balancing as per the requirements from the clients. To get assured reliability, the form is also available commercially. One pre requisite is that before use, installation must be done.

c. Virtual Load Balancer

These kind of load balancers are different than that of software based load balancers. In this kind of approach, the custom installation software for hardware on the virtual machines.

1.3 Cloud load balancing Importance In Cloud Computing

In terms of balancing of load in the context of cloud computing, the following are some of the issues that should put some lights on the significance of balancing the load.

a. Offerings of enhanced performance

Load balancing approach is a low-cost and simple for implementations techniques which is enablingthe businesses to work on client submissionsconsiderably more rapidlyalong with delivering the far better results at an attractive lesser investment.

b. Supports for maintaining the traffic flow of a website

It can provide scalability to control traffic on a website. High-end traffic can be managed with effective load balancers, which are used in conjunction with network devices and servers. Cloud load balancing is used by ecommerce businesses that must manage and distribute a large number of visitors per second.

c. Capability of handling of rapid bursts in traffic

Load balancers can handle any sudden traffic bursts that occur at the same time. For example, in the case of a university result, the website may be forced to close due to a high volume of requests. When using load balancers, there is no need to be concerned about traffic flow. Whatever the volume of

traffic, load balancers will distribute the entire load of the website evenly across multiple servers and provide maximum results in the shortest amount of time.

d. Better flexibility

One of the primary cause for using a load balancer is to guard the website from any type of unpredictedcalamity. When a workload is distributed across multiple network servers or units, if one of the node gets failed, the load is moved to alternative nodes. This offersfurther flexibility, scalability, and the capability to handle added traffics.

Load balancers are useful in a cloud environment because of these characteristics. It is done to avoid a large workload from overwhelming a single server [1].

2. Literature Survey:

The major goals of load balancing include establishing a fault tolerance system, maintaining system stability, improving performance and efficiency, minimising job execution time and waiting time in queue, increasing user satisfaction, and improving the resource utilisation ratio.

Mainak Adhikari and Tarachand Amgoth (2017) discussed various algorithms for balancing the load among the VMs and minimising the make span of the tasks. However, they are unable to locate potential information about resources and tasks, which may result in the incorrect assignment of tasks to virtual machines. As a result of their research, they proposed a new load balancing algorithm for the Infrastructure as a Service (IaaS) cloud. They also devised an effective strategy for configuring the servers based on the number of incoming tasks and their sizes in order to find suitable VMs for assignment and maximise computing resource utilisation. HBLBA is a new heuristic-based load-balancing algorithm for IaaS cloud. The proposed algorithm consists of two phases: server configuration and task-VM mapping. The goal of the server configuration is to host the appropriate number of VM instances and types to serve the incoming tasks while minimising the make span and maximising resource utilisation. The maximum queue length of each VM instance must be determined by the Virtual Machine Mapping based on its performance. The tasks assigned to the VM instances are stored in logical queues. The efficient strategy introduced is to configure the servers based on the number of tasks and their sizes in order to find suitable VMs for assignment. HBLBA's proposed algorithm is divided into two stages: server configuration and task-VM mapping. The admission controller's job is to determine whether or not a set of tasks can be assigned to a server. This decision is based on the computing resources available on the host server.Each server has a bounded queue for tasks received from customers via the data centre broker. The lengths of the server queues are fixed, whereas the lengths of the VMs are dynamic. [2].

Tianyi Chen et al. (2016) reported a systematic approach to designing energy-aware traffic-efficient geographical load balancing schemes for data-center networks that are not only optimal, but also computationally efficient and amenable to distributed implementation in their research article Distributed Stochastic Geographical Load Balancing over Cloud Networks. The proposed research described optimal schemes for real-time geographical load balancing tailored to the upcoming sustainable cloud networks. A stochastic optimization problem was developed to minimise the long-term aggregate cost of the MN-to-DC network while accounting for the spatiotemporal variability of workloads, renewables, and electricity prices. They also focused on stochastic load balancing and real-time distributed load balancing. [3].

Qi Liu, Member et. al (2016) and Z. Fu, X. Sun (2015) discussed , an adaptive scheme is proposed to achieve time and space efficiency in a heterogeneous cloud environment. To optimise the execution time of the Map phase, a dynamic speculative execution strategy on real-time management of cluster resources is presented, and a prediction model is used for fast prediction of task execution time. By combining the prediction model with a multi-objective optimization algorithm, an adaptive solution

for optimising space-time performance is obtained. The concept is based on Pay-as-you-consume, which is becoming more popular as one of the cloud consuming models due to its benefits, such as a large number of convenient services, reducing the burden of storage and flexible data access, and minimising the cost of hardware and software. To improve the performance of the Map phase, a dynamic speculative execution strategy is proposed. To predict the execution time of each reducer, a prediction model called PMK-ELM is presented. Better load balancing is achieved when combined with the DNSGA-II, which is designed to facilitate the selection of a suitable sequence for disperse variables. A real heterogeneous cloud environment has been set up in the research laboratory for testing the performance and benefits of the proposed scheme. The server also has 288 GB of memory and a 10 TB hard drive. On the server, eight virtual machines with varying amounts of memory and processors are established, and they are linked to a physical switch via the bridge mode. The research describes the algorithm for finding the task when there are enough resources, as well as execution time prediction using K-ELM. The heterogeneous environment makes it difficult to obtain a balanced load due to the different performance of each node. In the proposed method for allocating more data to a node with superior performance. Hadoop ensures that all tasks running on each node can substantially complete by monitoring running map and reduce tasks. This method is capable of detecting a wide range of load skew. However, Hadoop's complex implementation had a significant impact. A general method for measuring performance and load efficiency in cloud systems has been tested and presented [4][5][6][7].

Hung-Chang Hsiao et. al. (2012) described in their paper Load Rebalancing for Distributed File Systems in Clouds, they discuss load balancing. To address the load imbalance problem, the proposed research presents a fully distributed load rebalancing algorithm. Our algorithm is evaluated in comparison to a centralised approach in a production system and a competing distributed solution presented in the literature. The simulation results show that the proposed approach is comparable to the existing centralised approach and outperforms the prior distributed algorithm in terms of load imbalance factor, movement cost, and algorithmic overhead. In a cluster environment, the performance of the proposed Hadoop distributed file system implementation is investigated and reported. The problem of load balancing is also revisited and explained using a large-scale distributed file system. The contributions are divided into several sections, including a load balancing algorithm for distributing file chunks as uniformly as possible while minimising movement costs as much as possible, as well as an extensive study of load balancing algorithms based on DHTs. The simulation results show that, while each node performs our load balancing algorithm independently without acquiring global knowledge, the proposed solution is also compared with the centralized approach in Hadoop HDFS [9] and remarkably outperforms the competing distributed algorithm in [10] in terms of the load imbalance factor, the cost of movement, and the algorithmic overhead Furthermore, the load balancing algorithm demonstrated a high rate of convergence. Analytical models were also developed by the researchers to validate the efficiency and effectiveness of our design. Furthermore, the load balancing algorithm was implemented in HDFS and its performance in a cluster environment was investigated.

Amanpreet Kaur and Bikrampal Kaur (2019) proposed two hybrid approaches for the HDD-PLB framework: the Hybrid Predict Earliest Finish Time (PEFT) Heuristic with Ant Colony Optimization (ACO) met heuristic (HPA) and the Hybrid Heterogeneous Earliest Finish Time (HEFT) heuristic with ACO (HHA). The two proposed load balancing approaches were also critically examined and compared to determine which is superior for the proposed HDD-PLB framework. To fill the gaps identified, a framework for VM load balancing known as the ‘‘Hybrid approach based Deadline-constrained, Dynamic VM Provisioning and Load Balancing (HDD-PLB)”framework for Workflow execution was proposed and implemented. The reported and proposed framework is based on two hybrid approaches: PEFT-ACO and HEFT-ACO. In the proposed HDDPLBFW, Cybershake, Genome, and Ligo Workflows have been executed. The framework's performance has been described and analysed using two key metrics: Makespan and Cost.The simulation results of makespan and cost metrics for the two approaches have been analysed and compared, and the same have been reported. [11][12][13][14].

Ranesh Kumar Naha and Mohamed Othman (2016), described cost-aware service brokering and performance sentient load balancing algorithms in the cloud in their research presentation. They proposed three distinct cloud brokering algorithms as well as a load balancing algorithm. The cloud broker is in charge of managing provider resources as well as user requests. We employ two distinct policies, each with its own algorithm, to discover the available resources for users. The regionalList is a list of regions where data centres can be found. The dcCostLoadList contains a cost and load-by-load list for future allocation.This algorithm keeps a list of data centres organised by data centre location. All requests are distributed among all available data centres by the load-aware algorithm. A load balancing technique is used in the resource selection. A separate table for data centre costs is kept by the load-aware over cost algorithm. During simulation, it compares the list of lowest-cost data centres to another list involving the sender region [14] [15] [16].

MARWA GAMAL (2017) described An osmotic hybrid artificial bee and ant colony (OH BAC) optimization load balancing algorithm is formed using a hybrid meta heuristics technique that combines osmotic behaviour with bio-inspired load balancing algorithms as well as exploitation of the advantages of bio-inspired algorithms. The simulation results show that the drawbacks of existing bio-inspired algorithms in achieving load balancing between physical machines are overcome.The study also aimed to reduce energy consumption, the number of VM migrations, and the number of shutdown hosts when compared to existing algorithms. The improvement in service quality (QoSs) as measured by service level agreement violation (SLAV) and performance degradation due to migrations (PDMs). The OH BAC algorithm proposes using the osmosis theory from chemistry to form osmotic computing and find load balancing for VM placement[17][18] [19][20].

Jia Zhao. et. al (2014) , Bays theorem was used in the cloud environment in the research article on heuristic clustering-based task deployment approach for load balancing. The paper shed light on the problem of selecting physical hosts for deploying requested tasks and proposed a novel heuristic approach called LB-BC that combines the Bayes theorem with the clustering process to obtain the optimal clustering set of physical hosts (Load Balancing based on Bayes and Clustering). Clearly, the proposed approach helped to reduce the number of task deployment failure events, improve throughput, and optimise the external service performance of cloud data centres [21][22][23][24].

V. Priya , C. Sathiya Kumar and Ramani Kannan (2018) , in the published research article discussed and reported on a resource scheduling algorithm with load balancing for cloud service provisioning, as well as the introduction of a resource scheduling and load balancing algorithm for efficient cloud service provisioning. To achieve resource scheduling efficiency in cloud infrastructure, the method builds a Fuzzy-based Multidimensional Resource Scheduling model. Increasing Virtual Machine utilisation through effective and equitable load balancing is then accomplished by dynamically selecting a request from a class using the Multidimensional Queuing Load Optimization algorithm.In a cloud environment, the use of an effective integrated scheduling and load balancing algorithm subject to maximum resource utilisation and minimum processing time. After that, a load balancing algorithm is implemented to avoid resource underutilization and overutilization, thereby improving latency time for each type of request. When compared to state-of-the-art works, simulation analysis shows that the method improves resource scheduling efficiency by 7% and reduces response time by 35.5 percent. The F-MRSQN method's effectiveness is estimated by obtaining simulation results for testing the average success rate, resource scheduling efficiency, and response time [25][26][27][28].

Chun-Cheng Lin et. al (2014) , in Their proposed research article, Dynamic Multiservice Load Balancing in Cloud-Based Multimedia System, focused on a more practical dynamic multiservice scenario in which each server cluster only handles one type of multimedia task and each client requests a different type of multimedia service at a different time. An integer linear programming problem, which is computationally intractable in general, can be used to model such a scenario. The assumption is that in the CMS, each server cluster can only handle one type of multimedia service task at a time, and that each client requests a different type of multimedia service at a different time.

Such a problem can be modelled as an integer linear programming formulation at each time step, which is computationally intractable in general. [29][30][31][32].

3. Activities Involved In Load Balancing

Load balancing involves numerous activities such as identifying a VM's resource details, task scheduling, resource allocation, and migration.

a. Identification of resource details of a VM

This checks the status of a VM's resource details. It displays the current resource utilisation of the VM as well as the unallocated resources. Based on this phase, the VM's status can be classified as balanced, overloaded, or under loaded in relation to a threshold.

b. Task scheduling

A scheduling algorithm schedules tasks to appropriate resources on appropriate VMs once the resource details of a VM are identified.

c. Resource allocation

The resources are allotted to scheduled tasks to be completed. To accomplish this, a resource allocation policy is being used. There are numerous scheduling and allocation policies proposed in the literature. While scheduling is necessary to expedite execution, allocation policy is used to ensure proper resource management and improve resource performance. The efficacy of the scheduling algorithm and the allocation policy determine the strength of the load balancing algorithm.

d. Migration

Migration is an important stage in the cloud load balancing process, and the latter is insufficient without the former. In the cloud, migration is classified into two types based on the entity being considered: VM migration and task migration. The movement of a VM from one physical host to another to alleviate overloading is known as VM migration, and it is classified into two types: live VM migration and non-live migration. Similarly, task migration is the movement of tasks between VMs and is classified into two types: intra VM task migration and inter VM task migration. In the literature, a large number of migration approaches have been proposed. An effective migration technique results in effective load balancing. According to the results of a comprehensive survey, the task migration process is more time and cost effective than VM migration, and the trend has shifted from VM to task migration.

4. Comparative Analysis of Various Load Balancing For Cloud Computing

Table 4.1 compares various load balancing scenarios in a cloud computing environment It describes the knowledge base, the benefits and drawbacks of each type of algorithm, as well as the issues addressed by these algorithms. Comparison of Various Load Balancing Scenarios in a Cloud Computing Environment.

Table 4.1 Comparative Analysis of Various Load Balancing For Cloud Computing

Sr.	TypeofAlgorithm	Knowledge Base	Focused Key Issue(s)	Application Areas	Drawbacks
1	Static (Fixed in Nature)	The requirement of Priorknowledgebaseabout every node statisticsandusers requirement(s).	The Utilization of resources, Scaling along with time taken for responding. Electrical parameters such as consumptionof power as well as energyUtilizationMakespan. Actual output (throughput and performance).	Applications areas in homogeneousenvironment.	Rigid and lack of scalability. Incompatibility With changes in users(s) expectations along with the load.

Sr.	TypeofAlgorithm	Knowledge Base	Focused Key Issue(s)	Application Areas	Drawbacks
2	Dynamic	For adaptation of the dynamic load requirement(s), the monitoring of every node is carried out in real-time mode.	A processor's location to which load is transferred by an overloaded processor. Task delegation to a remote machine. Obtaining Information Estimated load. Keeping the number of migrations to a minimum. Throughput	In a heterogeneous environment, it is used.	Time Consuming Complexity
3	Centralized	A single node or server is in charge of maintaining and updating network statistics on a regular basis.	Policies based on thresholds Intensity of Throughput(s) failure ness. Communication between the central server and the network's processors. Overhead Accompanied	Preferred in the application areas where network size is limited with negligible or no load.	It is not faults tolerant(s). The central decision-making node is overburdened.
4	Distributed	To make efficient balancing decision(s), every load balancing processor(s) in the network stores the respective own local database (e.g. MIB).	Processors that participate in load balancing are chosen. It's migration time! Communication between processors Criteria for information exchange Throughput Tolerant to faults.	Usefulness is in a big and diverse environment.	Communication(s) overhead and complexity in Algorithm(s)
5	Hierarchical	To obtain information about network performance, nodes at distinct level(s) of hierarchy communicates with node(s) lower to self.	Threshold policies criteria's for exchanging of the info. Selection of nodes at different level(s) of network failures intensities and duration for migrating the performance.	Useful for medium or large-scale network work in a heterogeneous environment. The workflow is single.	Tolerance of faults are less as well as more complicated.

Sr.	TypeofAlgorithm	Knowledge Base	Focused Key Issue(s)	Application Areas	Drawbacks
6	Dependency in Workflow	Directed Acyclic Graphs are followed for modelling the dependencies between tasks and also it can be used for making decisions regarding scheduling(s).	Transaction incentivisation workflows and multiple workflows Work flows for data incentives Execution time Fault tolerance	In any type of environment, it is used to model task dependencies.	Modelling is difficult, and knowledge base maintenance is complicated. Greater complexity.

In addition to Table 4.1, the critical analysis of various algorithms is carried out for load balancing algorithms to achieve efficient throughput(s) as well as for reducing response time and avoiding resources overloading (statically and dynamically). The summary is represented in Table 4.2 these algorithms are compared in tabular form with theiradvantagesanddisadvantages.

Table 4.2 ComparisonofCurrentExistingLoadBalancingAlgorithm

Sr.	Algorithms	Static/Dynamic	Enlightenment	Pros(Advantages)	Cons. (Disadvantages)
1	Round robinapproach andRandomized concept.	Static	Processes are distributed evenly among all processors. Theprocessallocationorderis retained locallyoneach processor. In this fashionthe requests of user(s)beingprocessedin circularwaybyusingthisaid lgorithm(s).	Works well when the number of processes is greater than the number of processors. R/R does notdemands forcommunications among processes.	There are no expectations for improved performance in this (R/R) approach.
2	CentralManager	Static	The central processor is in charge of selecting the host for all new processes. The minimum loaded processor is determined by the total load selected when the process is established.	Load balancing decision(s) are through the load scheduler based on systems load info.	A high level of inter-process communication results in a bottleneck state.
4	Thresholding	Static	Processes are allocated to hosts without waiting time when they are created. Hosts for fresh processes	Threshold has little interposescommunication(s) Anumberoflocalpr	When all remote processes are overburdened, all processes are assigned locally.

Sr.	Algorithms	Static/Dynamic	Enlightenment	Pros(Advantages)	Cons. (Disadvantages)
			are assigned provincially fairly than distantly.	ocessallocationsaredone.	
5	Min-Min	Static	For all of the tasks, the lowestprobableaccomplishmentperiod is hunted. The lowest value (s) is/are taken from the minimum times, which is the shortest time among all tasks on any type of resource(s) in the system. The task is allocated to the respective machine(s) in the system founded on that smallestperiod.	Performs well with the least amount of resources.	It can take on the form of starvations.
6	Max-Min	Static	Max-Minisalmostcomparable tothemin-minalgorithmbutonethingisdiverse that is as after getting minimum executiontimes,themaximumvalueispreferredthatis themaximumtimeamongstallthetasksondiversetypeofresources.	Performance is well with a lesser number of resources.	It can furthermore leads towards the starvations situation..

Conclusion

The paper discussed cloud computing insights as well as challenges such as server failures, loss of confidentiality, and improper workloads in the context of cloud computing. This paper also described a critical study of load balancing concepts in the context of the cloud, as well as various types of load balancing techniques with numerous load balancers. The significance of load balancing in cloud computing is also thoroughly discussed. In addition, comparisons of various types of load balancing scenarios in a cloud computing environment are described.. It describes the knowledge base, the benefits and drawbacks of each type of algorithm, as well as the issues addressed by these algorithms. Different Types of Load Balancing Scenarios in a Cloud Computing Environment are compared. Furthermore, the paper focused on the different types of load balancing, as well as the significance of cloud load balancing for providing performance, maintaining the traffic of a website(s), handling sudden bursts in traffic, and flexibility. A literature review is conducted and summarised in tabular form as a ready reference for selecting the load balancing algorithm.The advantages and disadvantages are also included for comparative analysis, which will assist the researcher or designer in selecting an appropriate algorithm based on the requirements. Load balancing activities such as identifying a VM's resource details, task scheduling, resource allocation, and migration are also mentioned. As a result, a novel approach in the form of a model for optimal load balancing is required, such as minimum makespan, priority, and load balancing.

References

1. Afzal, S., Kavitha, G. Load balancing in cloud computing – A hierarchical taxonomical classification. J Cloud Comp **8,** 22 (2019). https://doi.org/10.1186/s13677-019-0146-7
2. M. Adhikari, T. Amgoth, Heuristic-based load-balancing algorithm for IaaS cloud, Future Generation Computer Systems (2017), https://doi.org/10.1016/j.future.2017.10.035
3. T. Chen, A. G. Marques and G. B. Giannakis, "DGLB: Distributed Stochastic Geographical Load Balancing over Cloud Networks," in IEEE Transactions on Parallel and Distributed Systems, vol. 28, no. 7, pp. 1866-1880, 1 July 2017, doi: 10.1109/TPDS.2016.2636210.
4. Q. Liu, W. Cai, J. Shen, X. Liu and N. Linge, "An adaptive approach to better load balancing in a consumer-centric cloud environment," in IEEE Transactions on Consumer Electronics, vol. 62, no. 3, pp. 243-250, August 2016, doi: 10.1109/TCE.2016.7613190.
5. M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, and M. Zaharia, "A view of cloud computing," Communications of the ACM, vol. 53, no. 4, pp.50-58, 2010.
6. Z. Fu, X. Sun, Q. Liu, L. Zhou, and J. Shu, "Achieving Efficient Cloud Search Services: Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Parallel Computing," IEICE Trans. Commun., vol. E98-B, no. 1, pp.190-200, 2015.
7. B. Palanisamy, A. Singh, and L. Liu, "Cost-effective resource provisioning for mapreduce in a cloud," IEEE Trans. Parallel Distrib. Syst., vol. 26, no. 5, pp. 1265-1279, 2015.
8. H. Hsiao, H. Chung, H. Shen and Y. Chao, "Load Rebalancing for Distributed File Systems in Clouds," in IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 5, pp. 951-962, May 2013, doi: 10.1109/TPDS.2012.196.
9. url VMware, http://www.vmware.com/ accessed on 9th December 2019.
10. I. Raicu, I. T. Foster, and P. Beckman, "Making a Case for Distributed File systems at Exascale," in Proc. 3rd Int'l Workshop Large-Scale System and Application Performance (LSAP'11), June 2011, pp. 11–18.
11. Kaur, Amanpreet & Kaur, Bikrampal. (2019). Load Balancing Optimization based on Hybrid Heuristic-Metaheuristic Techniques in Cloud Environment. Journal of King Saud University - Computer and Information Sciences. 10.1016/j.jksuci.2019.02.010.
12. Masdari, M., Salehi, F., Jalali, M., Bidaki, M., 2017. A survey of PSO-based scheduling algorithms in cloud computing. J. Network Syst. Manage. 25 (1), 122–158.
13. Yakhchi, M., Ghafari, S. M., Yakhchi, S., Fazeli, M., & Patooghi, A., 2015, ''Proposing a Load Balancing Method based on Cuckoo Optimization Algorithm for Energy Management in Cloud Computing Infrastructures. In: Proceedings of 6th International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO).
14. Naha, Ranesh & Othman, Mohamed. (2016). Cost-aware service brokering and performance sentient load balancing algorithms in the cloud. Journal of Network and Computer Applications. 75. 47-57. 10.1016/j.jnca.2016.08.018.
15. Domanal, S. G., Reddy, G.R.M., 2014. Optimal load balancing in cloud computing by efficient utilization of virtual machines.In:2014SixthInternationalConference on Communication Systems and Networks (COMSNETS), pp.1–4.
16. Kessaci,Y.,Melab,N.,Talbi,E-G.,2013.APareto-basedgeneticalgorithmforopti- mized assignment of vm requests on a cloud brokering environment.In:2013 IEEE Congress on Evolutionary Computation (CEC),pp.2496–2503.
17. M. Gamal, R. Rizk, H. Mahdi and B. E. Elnaghi, "Osmotic Bio-Inspired Load Balancing Algorithm in Cloud Computing," in IEEE Access, vol. 7, pp. 42735-42744, 2019, doi: 10.1109/ACCESS.2019.2907615.
18. A. Shawish and M. Salama, ``Cloud computing: Paradigms and technologies,"in Inter-cooperative Collective Intelligence: Techniques and Applications, vol. 495. Berlin, Germany: Springer, 2014, pp. 39_67.
19. M. Villari, M. Fazio, S. Dustdar, O. Rana, and R. Ranjan, ``Osmotic computing: A new paradigm for edge/cloud integration," IEEE Cloud Comput., vol. 3, no. 6, pp. 76_83, Nov./Dec. 2016.
20. B. Balusamy, J. Sridhar, D. Dhamodaran, and P. V. Krishna, ``Bio-inspired algorithms for cloud

computing: A review," Int. J. Innov. Comput. Appl., vol. 6, nos. 3_4, pp. 182_202, 2015.

21. J. Zhao, K. Yang, X. Wei, Y. Ding, L. Hu and G. Xu, "A Heuristic Clustering-Based Task Deployment Approach for Load Balancing Using Bayes Theorem in Cloud Environment," in IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 2, pp. 305-316, 1 Feb. 2016, doi: 10.1109/TPDS.2015.2402655.
22. T. Erl, R. Puttini, and Z. Mahmood, Cloud Computing: Concepts,Technology & Architecture, first ed. Indiana, U.S.: Prentice Hall,2013.
23. T. You, W. Li, Z. Fang, H. Wang, and G. Qu, ”Performance Evaluation of Dynamic Load Balancing Algorithms,” TELKOMNIKA Indonesian Journal of Electrical Engineering, vol. 12, no.4, 2014.
24. S. Song, T. Lv, and X. Chen,”A Static Load Balancing algorithm for Future Internet,” Journal of Electrical Engineering,vol. 12, no. 6, 2014.
25. V. Priya, C. Sathiya Kumar, Ramani Kannan, Resource scheduling algorithm with load balancing for cloud service provisioning,Applied Soft Computing,Volume 76,2019,Pages 416-424,ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2018.12.021.
26. D. Chitra Devi, V. Rhymend Uthariaraj, Load balancing in cloud computing environment using improved weighted round robin algorithm for nonpreemptive dependent tasks, Sci. World J. 2016 (2016) 1–14.
27. Qi Zhang, Mohamed Faten Zhani, Raouf Boutaba, Joseph L. Heller stein, Dynamic heterogeneity-aware resource provisioning in the cloud, IEEE Transaction Cloud Computing. 2 (1) (2014) 14–28.
28. Aarti Singh, Dimple Junejab, Manisha Malhotraa, Autonomous agent based load balancing algorithm in cloud computing, in: International Conference on Advanced Computing Technologies and Applications, ICACTA-2015, 45,Elsevier, 2015, pp. 832–841.
29. C. Lin, H. Chin and D. Deng, "Dynamic Multiservice Load Balancing in Cloud-Based Multimedia System," in IEEE Systems Journal, vol. 8, no. 1, pp. 225-234, March 2014, doi: 10.1109/JSYST.2013.2256320.
30. J. Sun, X. Wu, and X. Sha, “Load balancing algorithm with multiservice in heterogeneous wireless networks,” in Proc. 6th Int. ICST Conf. Commun. Networking China (China Com), 2011, pp. 703–707
31. X. Nan, Y. He, and L. Guan, “Optimal resource allocation for multimedia cloud based on queuing model,” in Proc. 13th IEEE Int. Workshop Multimedia Signal Process., Oct. 2011, pp. 1–6.
32. M. Garey and D. Johnson, Computers and Intractability—A Guide to the Theory of NP-Completeness. San Francisco, CA, USA: Freeman, 1979.[14] S. Kirkpatrik, C. Gelatt, and M. Vecchi, “Optimization by simulated annealing,” Science, vol. 220, pp. 671–680, May 1983.
33. Noshy M, Ibrahim A, Ali HA (2018) Optimization of live virtual machine migration in cloud computing: a survey and future directions. J Netw Comput Appl:1–10
34. Gkatzikis L, Koutsopoulos I (2013) Migrate or not? Exploiting dynamic task migration in mobile cloud computing systems. IEEE Wirel Commun 20(3):24–32
35. Jamshidi P, Ahmad A, Pahl C (2013) Cloud migration research: a systematic review. IEEE Trans Cloud Comp 1(2):142–157
36. Shamsinezhad E, Shahbahrami A, Hedayati A, Zadeh AK, Banirostam H (2013) Presentation methods for task migration in cloud computing by combination of Yu router and post-copy. Int J Comp Sci Iss 10(4):98
37. Katyal, Mayanka & Mishra, Atul. (2014). A Comparative Study of Load BalancingAlgorithms in Cloud Computing Environment. Sourcehttps://arxiv.org/ftp/arxiv/papers/1403/1403.6918.pdf