Journal of Science and Technology
ISSN: 2456-5660 Volume 8, Issue 04 (April -2023)
www.jst.org.in *DOI:https://doi.org/10.46243/jst.2023.v8.i04.pp72 - 79*

A LOAD BALANCING ALGORITHM FOR THE DATA CENTRES TO OPTIMIZE CLOUD COMPUTING

Dr.D.V Manjula, [1],Vinnakota Naga Teja Sravanthi [2],Maddula Kavya Sri Lakshmi Manisri Priyanka [3],Tatipakala Gayatri Priya [4],Jalli Ravi Kiran [5],Rajana Durgesh [6]
ASSOCIATE PROFESSOR
DEPT OF COMPUTER SCIENCE AND ENGINEERING
PRAGATI ENGINEERING COLLEGE(A),SURAMPALEM(EAST GODAVARI)A.P,INDIA

To Cite this Article
Dr.D.V Manjula, ,Vinnakota Naga Teja Sravanthi ,Maddula Kavya Sri Lakshmi Manisri Priyanka ,Tatipakala Gayatri Priya ,Jalli Ravi Kiran ,Rajana Durgesh**, "A LOAD BALANCING ALGORITHM FOR THE DATA CENTRES TO OPTIMIZE CLOUD COMPUTING "** *Journal of Science and Technology, Vol. 08, Issue 04,-April 2023, pp72-79*
Article Info
Received: 22-02-2023 Revised: 12-03-2023 Accepted: 25-03-2023 Published: 19-04-2023

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 back end 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 under loaded 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 span.

I. Introduction

As we shift more towards online storage and services, Cloud Computing technology becomes an essential part of the business. This technology provides services through various kinds such as in software via web browsers, in Platforms such as designing and developing cloud-based applications. In the Infrastructure, the backend is managed by Cloud Service Providers (CSPs) such as maintaining Data Centers, servers, etc. Although there exist many other service delivery models in this technology, however, in this research, the focus is on the Infrastructure as a Service (IaaS) model. It deals with the server-side of this technology for resource allocation.

Virtualization is the backbone and essential feature of cloud-based applications. This technique can significantly affect the performance of the scalable and on-demand services provided to clients if the migration process and allocation of virtual machine resources are handled inefficiently. According to, cloud performance is proved to be in the top three Cloud Computing challenges.

The cloud users access services by sending requests; these are represented in Virtual Machines (VMs) in the cloud environment. CSPs should deliver services that are beneficial to businesses and increase user satisfaction. Thus, the proposed Load Balancing algorithm is developed mainly focusing on the IaaS model out of the three service models in the cloud where authors deal with the Cloud Computing technology's backend, such as server workload.

There are two components in a typical cloud environment: the frontend is the user side, and it is accessible by connecting to the Internet. The backend side handles the cloud service models where the Data Center store multiple physical machines (known as servers). Incoming user requests are received from the application are dynamically scheduled, and through virtualization, the necessary resources are allocated to clients. The virtualization technique is also responsible for balancing the load in the entire system, scheduling, and efficient allocation of resources. CSPs and cloud users can leverage the advantage of virtualization as well as dynamic task scheduling techniques. Thus, efficient scheduling can highly reduce execution time and increase the ratio of resource utilization in cloud-based applications.

II. LITERATURE SURVEY

''Cloud computing virtualization of resources allocation for distributed systems,'

Cloud computing is a new technology which managed by a third party "cloud provider" to provide the clients with services anywhere, at any time, and under various circumstances. In order to provide clients with cloud resources and satisfy their needs, cloud computing employs virtualization and resource provisioning techniques. The process of providing clients with shared virtualized resources (hardware, software, and platform) is a big challenge for the cloud provider because of over-provision and under-provision problems. Therefore, this paper highlighted some proposed approaches and scheduling algorithms applied for resource allocation within cloud computing through virtualization in the datacenter. The paper also aims to explore the role of virtualization in providing resources effectively based on clients' requirements. The results of these approaches showed that each proposed approach and scheduling algorithm has an obvious role in utilizing the shared resources of the cloud data center. The paper also explored that virtualization technique has a significant impact on enhancing the network performance, save the cost by reducing the number of Physical Machines (PM) in the datacenter, balance the load, conserve the server's energy, and allocate resources actively thus satisfying the clients' requirements. Based on our review, the availability of Virtual Machine (VM) resource and execution time of requests are the key factors to be considered in any optimal resource allocation algorithm.

''Cloud computing: A paradigm shift in the way of computing,''

Introduction Cloud computing is a new trend of computing where resources like storage, computation power, network, applications etc. are delivered as services. These services are available to the customers as subscription-based model i.e. pay-as-you go. In this model, customers can get these services on their demands regardless of where these services are hosted and customers have to pay depending on their usage of services. In cloud computing, resources are made virtual and unlimited. Also, the resources can be provisioned from anywhere i.e. always available at any location. So, cloud computing is a new paradigm where we can provision resources dynamically, deploy applications, and can access platform-independent services. Cloud computing, successor of internet computing, is a technology, where the concept of utility, scalability, on-demand services are incorporated. Figure 1 illustrates " Internet Computing " vs. " Cloud Computing " . Defining Cloud in IT According to the U.S. National Institute of Standards and Technology (NIST) , Cloud is a classical model which enable omnipresent, convenient, on-demand network access to a publicly accessible pool of configurable resources like servers, storage, network components, applications; that can be accessed, manipulated and released with minimal management effort, less cost and minimal service provider interaction. Cloud computing can be defined by the following important properties. Service on demand: Cloud users can use services on their demands, whenever they need from any place and at any time without making any direct communication with cloud service provider. Wide network access: Services can be accessed over

the network using different devices (like laptops, mobile phones, PDA, tablets, office computer etc.). Services can be provisioned in any platform, which means cloud services are platform independent. Pooled Resources: In cloud computing, resources are pooled together so that cloud providers can offer multi-tenant services. Multi-tenant supports multiple users to be served at a time with physical and virtual resources. These resources can be dynamically assigned and released according to the user's choice. Increased elasticity: There is no limit for provisioning resources via cloud. So services can be easily and quickly scale in and scale out. For example, an online shopping site uses the resources from the cloud in terms of users.

''Toward cloud computing: Security and performance,''

Security and performance are basic requirements for any system. They are considered the criteria for the measurement of any progress in a security system. Security is an indicator that affects the level of performance through the threats that influence the performance of parts of the cloud during the rendering of services. Both security and performance demonstrate the efficiency of cloud computing which indicates that the performance and security are measurements for the extent of the development of the cloud. In this paper, the relationship between performance and security will be examined to know the extent of their impact on the progress of cloud computing.

III.SYSTEM ANALYSIS

3.1. EXISTING SYSTEM

This subsection provides a review of previous existing algorithms in the field of Load Balancing and Task Scheduling. Many recent algorithms aimed to improve Task Scheduling and Load Balancing. Yet, few limitations still exist due to the underlying basic algorithms used, such as Round Robin or First Come First Serve. These algorithms can increase the waiting time or Make span in scheduling tasks.

Authors in proposed a dynamic Load Balancing algorithm to minimize the Make span time and utilize resources efficiently. It sorts tasks using length and processing speed by using the bubble sort algorithm. Then, tasks are allocated to Virtual Machines in a First-Come-First-Serve order. After allocation is complete, balancing the load is done considering and calculating the load of Virtual Machines.

This approach can easily optimize the resources and reduce Make span; however, it does not consider priority or any QoS parameters such as Deadline.

3.2. PROPOSED SYSTEM

In this subsection, we describe this research's objective in an illustrative diagram to explain the problem in Load Balancing and the role of the proposed LB algorithm, as seen in figure 3 below. This proposed model's main goal is to provide efficient resource allocation in a cloud environment whereby it avoids unbalanced workload in Cloud Computing applications. This model resolves issues related to workload migration and task rejection in the cloud. The proposed framework consists of two layers:

• Top Layer: deals with requests from multiple different clients (application's users) of both mobile and desktop. Clients can access the Internet using different devices to send requests to the cloud. In this layer, the model uses the Cloudlet Scheduler Time Shared algorithm to submit tasks in a random order (Arrival Time) and schedule them to Virtual Machines by considering two main parameters: Deadline and Completion Time. In Cloud Computing, Data Center (DC)

can be described as big storage for cloud servers and data. DC receives requests and sends them to the active load balancer. In this layer of the model, the proposed algorithm is implemented as a Load Balancer, which acts as the primary balancer in the cloud environment to perform migration in the case of violation, which has not been addressed in the previous literature up to the author's knowledge.

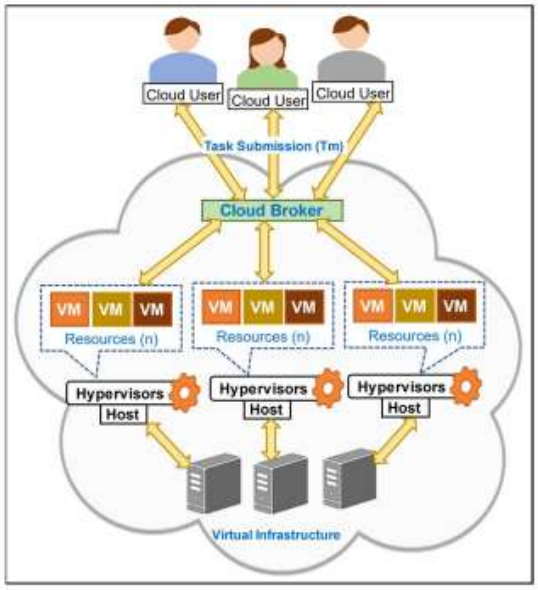
• Bottom Layer: deals with allocation of user requests to Virtual Machines (VMs). As the figure illustrates, we have a primary batch of VMs; VM2's status is set to high priority since it violates the SLA requirement, which means its Completion Time is higher than the Deadline. Thus, the proposed LBA should apply a migration technique to transfer the workload to another available Virtual Machine by reconfiguring the MIPS of both VMs before and after allocating the resources to them. The allocation table is then updated whenever a Virtual Machine becomes violated or not, along with the number of requests it's been allocated. There is a case where there is no SLA violation. Suppose the Time to Complete (TTC) is less than SLA (Deadline) given for tasks to run on VMs. Then, no SLA violation occurs.

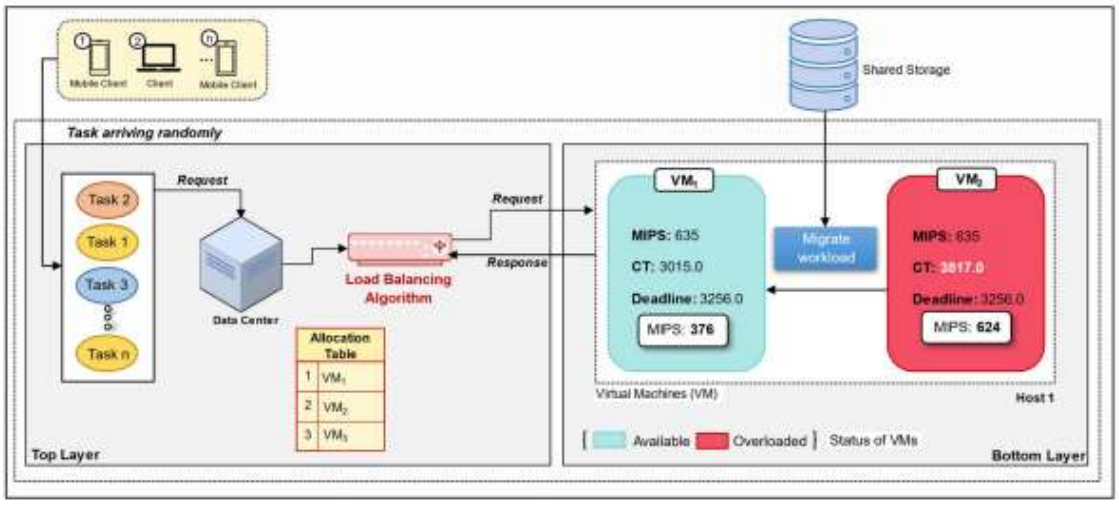
Overall, the proposed framework supports dynamic scheduling and load balancing to fully utilize the CPU and fully the cloud resources.

IV.SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture of Comparative Evaluation for Traditional Machine Learning and Deep Learning Classification Techniques for Sentiment Analysis.





4.1. System Architecture

V. SYSTEM IMPLEMENTATION

5.1. MODULES

There are 3 modules:

- Task Scheduler
- User
- Cloud
-

Task Scheduler: -

- Login
- View Task
- Assign Task
- VM Resources
- Logout

User: -

- Register
- Login
- File Upload
- My Files
- My Profile
- Logout

Cloud: -

- Login
- Add Virtual Machine
- Manage Virtual Machine
- Task Scheduling Details
- Load Balancer
- Make span

VI. RESULTS

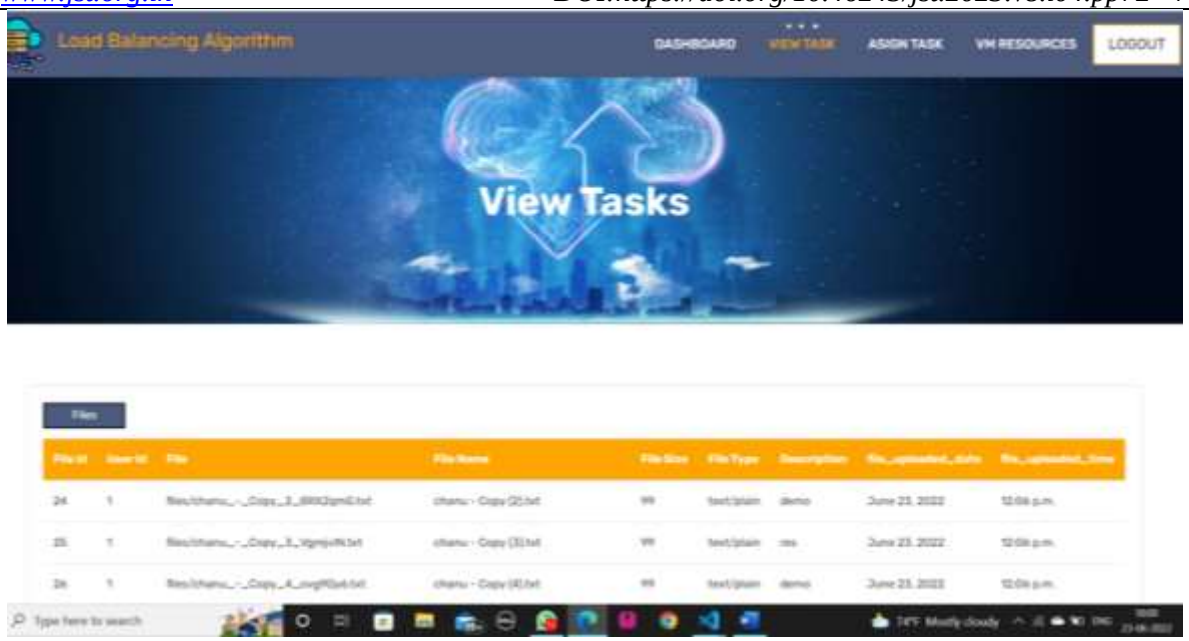


Fig. 6. Uploaded Records Page

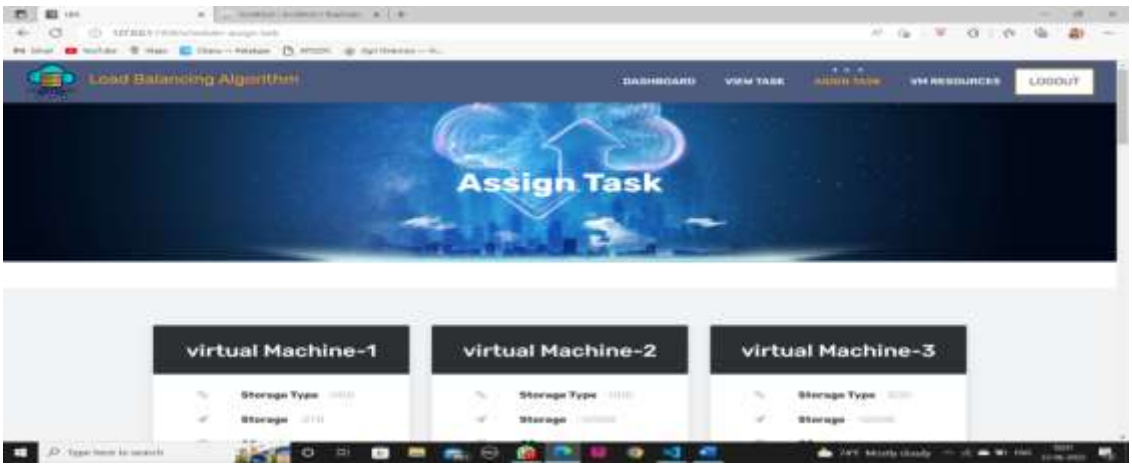


Fig 6.1 load handling

Journal of Science and Technology
ISSN: 2456-5660 Volume 8, Issue 04 (April -2023)
www.jst.org.in *DOI:https://doi.org/10.46243/jst.2023.v8.i04.pp72 - 79*

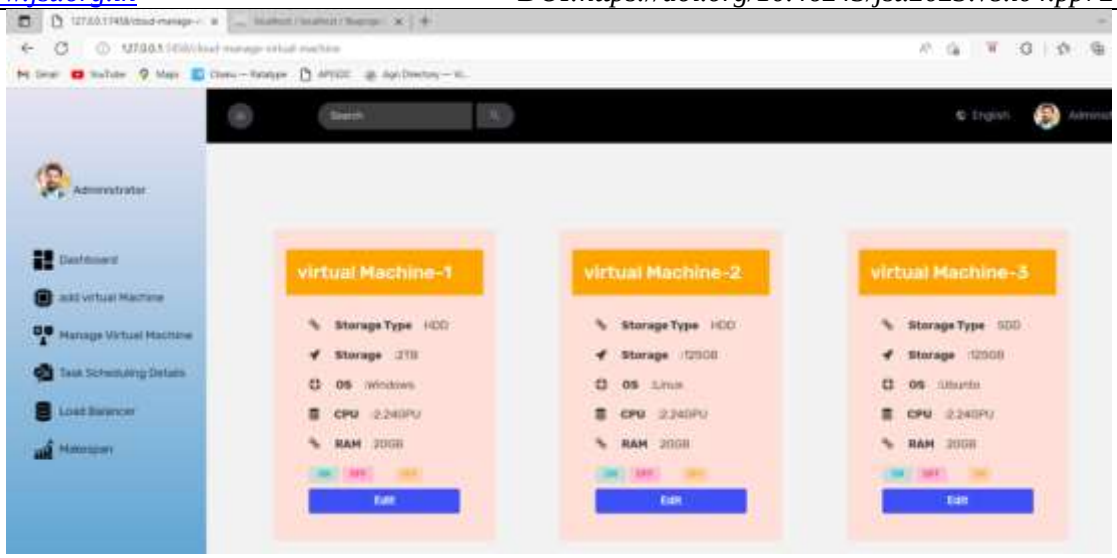


Fig. 6.2 Load balancing

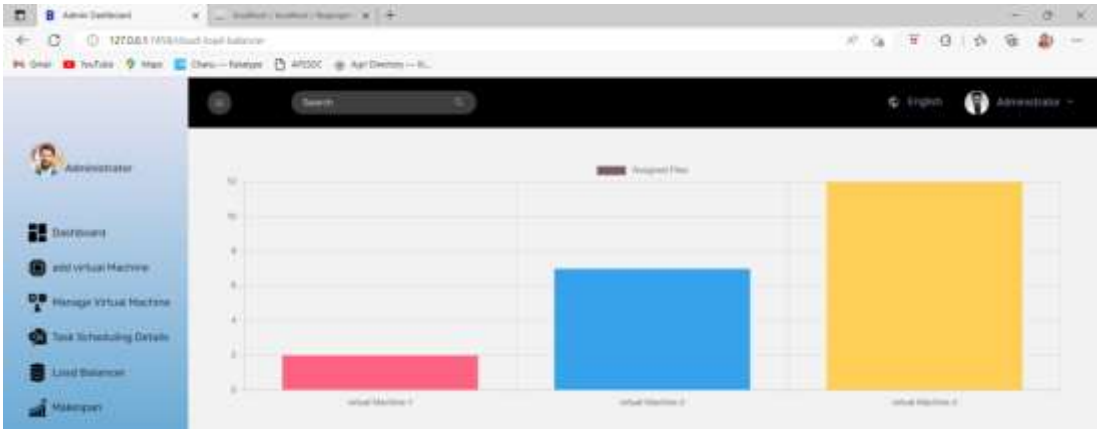


Fig. 6.2 Load balancing analysis

VII.CONCLUSION AND FUTURE WORK

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 MAKE SPAN 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] H. Shukur, S. Zeebaree, R. Zebari, D. Zeebaree, O. Ahmed, and A. Salih, ‘‘Cloud computing virtualization of resources allocation for distributed systems,’’ J. Appl. Sci. Technol. Trends, vol. 1, no. 3, pp. 98–105, Jun. 2020, doi: 10.38094/jastt1331.

[2] M. Agarwal and G. M. Saran Srivastava, ‘‘Cloud computing: A paradigm shift in the way of computing,’’ Int. J. Mod. Educ. Comput. Sci., vol. 9, no. 12, pp. 38–48, Dec. 2017, doi: 10.5815/ijmecs.2017.12.05.

[3] N. Zanoon, ‘‘Toward cloud computing: Security and performance,’’ Int. J. Cloud Comput.: Services Archit., vol. 5, no. vol. 5, nos. 5–6, pp. 17–26, Dec. 2015, doi: 10.5121/ijccsa.2015.5602.

[4] C. T. S. Xue and F. T. W. Xin, ‘‘Benefits and challenges of the adoption of cloud computing in business,’’ Int. J. Cloud Comput.: Services Archit., vol. 6, no. 6, pp. 1–15, Dec. 2016, doi: 10.5121/ijccsa.2016.6601.

[5] D. A. Shafiq, N. Jhanjhi, and A. Abdullah, ‘‘Proposing a load balancing algorithm for the optimization of cloud computing applications,’’ in Proc. 13th Int. Conf. Math., Actuarial Sci., Comput. Sci. Statist. (MACS), Dec. 2019, pp. 1–6, doi: 10.1109/MACS48846.2019.9024785.

[6] S. K. Mishra, B. Sahoo, and P. P. Parida, ‘‘Load balancing in cloud computing: A big picture,’’ J. King Saud Univ.–Comput. Inf. Sci., vol. 32, no. 2, pp. 149–158, 2020, doi: 10.1016/j.jksuci.2018.01.003.

[7] I. Odun-Ayo, M. Ananya, F. Agono, and R. Goddy-Worlu, ‘‘Cloud computing architecture: A critical analysis,’’ in Proc. 18th Int. Conf. Comput. Sci. Appl. (ICCSA), Jul. 2018, pp. 1–7, doi: 10.1109/ICCSA.2018.8439638.

[8] A. Jyoti, M. Shrimali, and R. Mishra, ‘‘Cloud computing and load balancing in cloud computing -survey,’’ in Proc. 9th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence), Jan. 2019, pp. 51–55, doi: 10.1109/confluence.2019.8776948.

[9] S. H. H. Madni, M. S. Abd Latiff, M. Abdullahi, S. M. Abdulhamid, and M. J. Usman, ‘‘Performance comparison of heuristic algorithms for task scheduling in IaaS cloud computing environment,’’ PLoS ONE, vol. 12, no. 5, May 2017, Art. no. e0176321, doi: 10.1371/journal.pone.0176321.

[10] M. Adhikari and T. Amgoth, ‘‘Heuristic-based load-balancing algorithm for IaaS cloud,’’ Future Gener. Comput. Syst., vol. 81, pp. 156–165, Apr. 2018, doi: 10.1016/j.future.2017.10.035.

[11] B. Singh and G. Singh, ‘‘A study on virtualization and hypervisor in cloud computing,’’ Int. J. Comput. Sci. Mobile Appl., vol. 6, no. 1, pp. 17–22, 2018.

[12] M. Kumar, S. C. Sharma, A. Goel, and S. P. Singh, ‘‘A comprehensive survey for scheduling techniques in cloud computing,’’ J. Netw. Comput. Appl., vol. 143, pp. 1–33, Oct. 2019, doi: 10.1016/j.jnca.2019.06.006..