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 Centres, 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-deman 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. This research aims to enhance resource allocation in the IaaS model; this concept is fundamental as it deals with the balancing of resources provided to clients and the workload/user requests on servers.

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.