



A New Method for Load Balancing in Cloud Data Centers Using Clustering, Requests, and Virtual Machines (VMs)

Iman Momeni¹, Kamran Zamanifar^{2,*}

¹Department of Computer Engineering, Islamic Azad University, Najafabad Branch, Isfahan, Iran
Email address: Momeniiman@yahoo.com, telephone number: 00989132145973

² Department of Computer Engineering, Isfahan University, Isfahan, Iran
*corresponding author : zamanifar@eng.ui.ac.ir

ABSTRACT

There are some nodes in cloud systems to meet users' needs. The volume of the input and processing loads in some of these nodes may sometimes be different from those in other nodes of a cloud, and there may be some imbalance between them as well. Many methods have been developed and presented to balance the loads. But the main weaknesses of these methods have been the load imbalance at the moment of the entry of tasks, and lack of attention to the various computational volumes of tasks entering the cloud platform, which will reduce the performance in the cloud, and which will increase the response time, thus ultimately reducing the system performance. Using genetic algorithm techniques for producing a generalized genetic algorithm (GA), and using the K-means clustering algorithm for homogeneous clustering by means of virtual machines (VMs), we reduced response time, and increased the utilization of processing resources in the cloud platform. First, we created VMs with various computational volumes in the cloud platform, and then clustered the tasks entering the cloud and the VMs homogeneously with task clusters. Afterwards, the generalized genetic algorithm (GA) performed the most precise assignment of task clusters to VM clusters, and then performed the most precise intra cluster assignment at the next level. The method proposed in this paper, on average, has an improvement of at least 152% higher than other algorithms. Moreover, the input and experimental data in this paper are derived from the Large Hadron Data Center, which is created to reveal the emergence of the universe and the Big Bang, and which has genuine records.

Keywords: Cloud computing, Load balance, Clustering, Genetic algorithm (GA)

1. INTRODUCTION

Cloud computing is used in conjunction with the Internet. The entire Internet can be viewed as a cloud. This is an important aspect of the method for developing a business without a large investment. Because it can provide almost all kinds of services, such as: software, platforms, and storage and computing infrastructures on the Internet at a very reasonable cost. Operational costs can be reduced through cloud computing [1].

Distributed computations were developed with the idea of increasing the processing power of computers with considerable advantages, and turned into a network and/or a cluster of computations. Presentation of resource computing as a tool was a very welcomed idea around the world, which led to the success of the cloud [2].

Cloud computing embraces the virtualization concept. Through virtualization, the end user can use various services in the cloud. In cloud computing, data centers provide services in virtual forms. A number of VMs are produced in data centers, and there is a need to efficiently assign input requests to these VMs, so that their capacity is better utilized, and this is a requirement [3, 4, 5].

In the environment of a distributed system, maintaining load balances is one of the most important tasks. This is an important aspect of resource distribution that states that under different circumstances, no overload should occur in devices, and that resources should be used optimally [6].

This makes the system more efficient, and reduces the response time. The load balancer should avoid overloading in each given node [7].

The load balancer is used to assign loads to different VMs, in a way that none of the nodes are loaded severely or slowly. Load balancing must be done correctly, because failure in any of the nodes can lead to lack of data access [8].