Journal of Network and Computer Applications 88 (2017) 50–71

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

Load-balancing algorithms in cloud computing: A survey

Einollah Jafarnejad Ghomi[a], Amir Masoud Rahmani[a,*], Nooruldeen Nasih Qader[b]

[a] Science and Research Branch, Islamic Azad University, Tehran, Iran
[b] Computer Science, University of Human Development, Sulaimanyah, Iraq

ARTICLE INFO

Keywords:
Cloud computing
Load balancing
Task scheduling
Hadoop MapReduce

ABSTRACT

Cloud computing is a modern paradigm to provide services through the Internet. Load balancing is a key aspect of cloud computing and avoids the situation in which some nodes become overloaded while the others are idle or have little work to do. Load balancing can improve the Quality of Service (QoS) metrics, including response time, cost, throughput, performance and resource utilization.

In this paper, we study the literature on the task scheduling and load-balancing algorithms and present a new classification of such algorithms, for example, Hadoop MapReduce load balancing category, Natural Phenomena-based load balancing category, Agent-based load balancing category, General load balancing category, application-oriented category, network-aware category, and workflow specific category. Furthermore, we provide a review in each of these seven categories. Also. We provide insights into the identification of open issues and guidelines for future research.

1. Introduction

Cloud computing is a modern technology in the computer field to provide services to clients at any time. In a cloud computing system, resources are distributed all around the world for faster servicing to clients (Dasgupta et al., 2013; Apostu et al., 2013). The clients can easily access information via various devices such as laptops, cell phones, PDAs, and tablets. Cloud computing has faced many challenges, including security, efficient load balancing, resource scheduling, scaling, QoS management, data center energy consumption, data lock-in and service availability, and performance monitoring (Kaur et al., 2014; Malladi et al., 2015). Load balancing is one of the main challenges and concerns in cloud environments;(Jadeja and Modi, 2012) it is the process of assigning and reassigning the load among available resources in order to maximize throughput, while minimizing the cost and response time, improving performance and resource utilization as well as energy saving (Singh et al., 2016; Goyal et al., 2016). Service Level Agreement (SLA) and user satisfaction could be provided by excellent load balancing techniques. Therefore, providing the efficient load-balancing algorithms and mechanisms is a key to the success of cloud computing environments. Several researches have been done in the field of load balancing and task scheduling in cloud environments. However, our studies showed that despite the key role of load-balancing algorithms in cloud computing, especially in the advent of big data, there are a few comprehensive reviews of these algorithms. First, we mention a few recent papers that have reviewed the load-balancing algorithms and mechanisms in cloud environments:

- Milani and Navimipour (2016) have presented a systematic review of the existing load balancing techniques. They classified the existing techniques based on different parameters. The authors compared some popular load-balancing algorithms and presented their main properties, including their advantages and disadvantages. They also addressed the challenges of these algorithms and mentioned the open issues. However, their work lacks a discussion regarding the load balancing and task scheduling techniques in Hadoop MapReduce that is an issue nowadays.
- Mesbahi and Rahmani (2016) have studied state of the art load balancing techniques and the necessary requirements and considerations for designing and implementing suitable load-balancing algorithms for cloud environments. They presented a new classification of load balancing techniques, evaluated them based on suitable metrics and discussed their pros and cons. They also found that the recent load balancing techniques are focusing on energy saving. However, their work suffers from the lack of simulating the load balancing techniques by simulator tools; in addition, a discussion of open issues and future topics that researchers should focus on is also missing.
- Kanakala et al. (2015a, 2015b) have analyzed the performance of load balancing techniques in cloud computing environments. They studied several popular load-balancing algorithms and compared

* Corresponding author.
E-mail addresses: e.jafarnejad@srbiau.ac.ir (E. Jafarnejad Ghomi), rahmani@srbiau.ac.ir (A. Masoud Rahmani), nooruldeen.qader@uhd.edu.iq (N. Nasih Qader).

http://dx.doi.org/10.1016/j.jnca.2017.04.007
Received 31 December 2016; Received in revised form 6 March 2017; Accepted 7 April 2017
Available online 08 April 2017