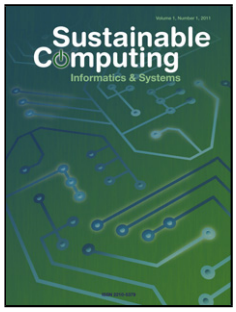

Title: Energy efficient dynamic resource management in cloud computing based on logistic regression model and median absolute deviation

Authors: Yaser Jararweh, Manar Bani Issa, Mustafa Daraghmeh, Mahmoud Al-Ayyoub, Mohammad A. Alsmirat

PII: S2210-5379(18)30011-8
DOI: https://doi.org/10.1016/j.suscom.2018.07.005
Reference: SUSCOM 261

To appear in:

Received date: 15-1-2018
Revised date: 2-7-2018
Accepted date: 16-7-2018

Please cite this article as: Jararweh Y, Issa MB, Daraghmeh M, Al-Ayyoub M, Alsmirat MA, Energy efficient dynamic resource management in cloud computing based on logistic regression model and median absolute deviation, *Sustainable Computing: Informatics and Systems* (2018), https://doi.org/10.1016/j.suscom.2018.07.005

Yaser Jararweh*, Manar Bani Issa*, Mustafa Daraghmeh*, Mahmoud Al-Ayyoub*, Mohammad A. Alsmirat*
*Jordan University of Science and Technology, Irbid, Jordan
{yijararweh, maalshbool, masmirat}@just.edu.jo

Paper Highlights

- Optimize resource utilization and energy efficiency in cloud data centers.
- proposing host overloading detection algorithm based on logistic regression model and median absolute derivation.
- Promote the VM consolidation technique.

Abstract

The unprecedented trend of using public cloud computing services by increasing number of customers motivates cloud services providers to optimize their resources usage and management to the limit. This is including managing cloud user's virtual machines (VM) running on one or more of the thousands of hosting servers or physical machines (PMs) of the cloud datacenters. The cloud service providers are mainly concerned on answering the two main questions that dramatically impact their infrastructure usage and utilization; Where to initially place the VMs and where to move them in case we need to move them. Along with the VM consolidation technique, VMs migration will help in protecting the physical servers from being overloaded or reduce the number of active physical servers for better resources utilization and energy saving. Efficiently detecting overloaded servers will help in improving the cloud system performance and reduce the total operational costs which will provide competitiveness for the cloud provider in the market. In this work, we are proposing a general host overloading detection algorithm based on logistic regression model and median absolute derivation. The proposed algorithm is scalable and can be used with any VM placement and migration algorithms. An extensive evaluation procedure is used with dynamic workload to proof the efficiency of the proposed algorithm. The archived results show that the proposed algorithm outperforms all other known host status prediction techniques.

***Index Terms* — Cloud computing, resource management, energy efficiency, dynamic consolidation, host detection algorithm, logistic regression, median absolute deviation**

I. INTRODUCTION

With the huge development of information and communication technologies and the with the ever-increasing volume of data, the need for huge computing services and hardware resources (such as memory, permanent storage, network bandwidth, etc.) has become very urgent [1]. These huge computing services and hardware resources are currently offered, under the name of Cloud Computing, by many companies as a paid service. Customers usually use these services and resources to complete their resource demanding tasks. Cloud computing customers usually do not concern themselves with how and where the work is actually carried out. What is more interesting for the customers is the cost arising from the use of these services and resources. The customers are also concerned with the quality of service provided to them. There are some significant challenges that face cloud computing service providers, such as the optimal use of the resources and how they can manage and control these resources (Hardware and Services). Also, providing services to the customers with the highest quality at the lowest cost and with a commitment to Service Level Agreement (SLA) [2].

SLA is a contract between the cloud service provider and the cloud customer [3], in which the requirements of the cloud customer request are determined [4]. In addition, it contains comprehensive and detailed descriptions of the services provided by the cloud service provider, such as performance and availability, and the terms and conditions of the cloud customer, such as the cost [5]. SLA contains also the procedures that must be taken (such as penalties) in case of any agreement violation.

Cloud service providers usually offer huge amount of resources in the shape of different physical machines (aka data center) that are connected to each other using a fast network. To perform its tasks, the customer is provided with a subset of these resource, that may range from a subset of a single physical machine resources to the resources of a set of physical machines, according to some SLA. However, these resources might be underutilized. The problem of underutilized resources may force the cloud service providers to provide more resources and upgrade their existing infrastructure. Moreover, energy consumption increases because