Computers and Electrical Engineering 47 (2015) 69–81

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Workload modeling for resource usage analysis and simulation in cloud computing

Deborah Magalhães [a,b,*], Rodrigo N. Calheiros [b], Rajkumar Buyya [b], Danielo G. Gomes [a]

[a] Group of Computer Networks, Software Engineering, and Systems (GREat) Departamento de Engenharia de Teleinformática, Universidade Federal do Ceará, Fortaleza-Brazil

[b] Cloud Computing and Distributed Systems (CLOUDS) Laboratory, Department of Computing and Information Systems, The University of Melbourne, Australia

ARTICLE INFO

Article history:
Received 26 December 2014
Revised 25 August 2015
Accepted 27 August 2015

Keywords:
Cloud computing
Workload modeling and characterization
Simulation
Distribution analysis
Web applications

ABSTRACT

Workload modeling enables performance analysis and simulation of cloud resource management policies, which allows cloud providers to improve their systems' Quality of Service (QoS) and researchers to evaluate new policies without deploying expensive large scale environments. However, workload modeling is challenging in the context of cloud computing due to the virtualization layer overhead, insufficient tracelogs available for analysis, and complex workloads. These factors contribute to a lack of methodologies and models to characterize applications hosted in the cloud. To tackle the above issues, we propose a web application model to capture the behavioral patterns of different user profiles and to support analysis and simulation of resources utilization in cloud environments. A model validation was performed using graphic and statistical hypothesis methods. An implementation of our model is provided as an extension of the CloudSim simulator.

1. Introduction

Clouds are being used as a platform for various types of applications with different Quality of Service (QoS) aspects, such as performance, availability and reliability. These aspects are specified in a Service Level Agreement (SLA) negotiated between cloud providers and customers. The failure to comply with QoS aspects can compromise the responsiveness and availability of service and incur SLA violations, resulting in penalties to the cloud provider. The development of resource management policies that support QoS is challenging and the evaluation of these policies is even more challenging because clouds observe varying demand, their physical infrastructure has different sizes, software stacks, and physical resources configurations, and users have different profiles and QoS requirements [1]. In addition, reproduction of conditions under which the policies are evaluated and control of evaluation conditions are difficult tasks.

In this context, workload modeling enables performance analysis and simulation, which brings benefits to cloud providers and researchers. Thereby, the evaluation and adjustment of policies can be performed without deployment of expensive large scale environments. Workload models have the advantage of allowing workload adjustment to fit particular situations, controlled

* Corresponding author at.: Cloud Computing and Distributed Systems (CLOUDS) Laboratory, Department of Computing and Information Systems, The University of Melbourne, Australia. Tel.: +61 415722721.

E-mail address: deborah.vm@gmail.com, deborah@great.ufc.br (D. Magalhães).

http://dx.doi.org/10.1016/j.compeleceng.2015.08.016

modification of parameters, repetition of evaluation conditions, inclusion of additional features, and generalization of patterns found in the application [2], providing a controlled input for researchers. For cloud providers, the evaluation and simulation of resource management policies allow the improvement of their systems' QoS. Finally, the simulation of workloads based on realistic scenarios enables the production of tracelogs, scarce in cloud environments because of business and confidentiality concerns [3,4].

Workload modeling and characterization is especially challenging when applied in a highly dynamic environment, such as cloud data centers, for different reasons:(i) heterogeneous hardware is present in a single data center and the virtualization layer incurs overhead caused by I/O processing and interactions with the Virtual Machine Monitor (VMM); and (ii) complex workloads are composed of a wide variety of applications submitted at any time, and with different characteristics and user profiles. These factors contribute to a lack of methodologies to characterize the different behavioral patterns of cloud applications.

To tackle the above issues, a web application model able to capture the behavioral patterns of different user profiles is proposed to support analysis and simulation of resources utilization in cloud environments. The proposed model supports the construction of performance models used by several research domains. Performance models improve resource management because they allow the prediction of how application patterns will change. Thus, resources can be dynamically scaled to meet the expected demand. This is critical to cloud providers that need to provision resources quickly to meet a growing resource demand by their applications.

In this context, the main contribution of this paper is a model capable of representing resource demand of Web application supported by different user profiles in a context of cloud environment. The workload patterns are modeled in the form of statistical distributions. Therefore, the patterns fluctuate based on realistic parameters in order to represent dynamic environments. A model validation is provided through graphical and analytical methods in order to show that the model effectively represents the observed patterns. A secondary contribution of this paper is the validation and implementation of the proposed model as an extension of the CloudSim simulator [1], making the model available for the cloud research community.

The rest of the paper is organized as follows: Section 2 presents the challenges and importance of workload modeling in clouds. Section 3 describes related works. Section 4 details the adopted methodology and how it was achieved. Section 5 presents and discusses the modeling and simulation results. Section 6 concludes and defines future research directions.

2. Problem statement and motivation

Workload characterization and modeling problems have been addressed over the last years, resulting in models for generation of synthetic workloads similar to those observed on real systems [2]. The main objective of such models is enabling the behavior patterns detection on the collected data.

2.1. Challenges of workload modeling in clouds

Workload modeling and characterization is especially challenging when applied in a highly dynamic environment such as a cloud, for various reasons, as discussed below:

1. Hardware platforms heterogeneity: Information Technology (IT) managers update about 20% to 25% of their platforms every year [5], resulting in the combination of different hardware in the same data center. Besides, the virtualization layer promotes an overhead caused by I/O processing and interactions with the Virtual Machine Manager (VMM). This overhead depends on the hardware platform.
2. Business and confidentiality: due to business and confidentiality reasons, there are few cloud tracelogs available for analysis. Thus, there is a lack of methodologies to characterize the different behavioral patterns of cloud applications [3,4]. Nevertheless, recent efforts in this direction, such as, Google TraceLog [6] and Yahoo!, enable data analysis and characterization for specific scenarios [7].
3. Workload complexity: the cloud hosts a wide variety of applications submitted at any time, with different characteristics and user profiles, which have heterogeneous and competing QoS requirements [1]. This leads to complex workloads depending on users' behavior and resource consumption. Thus, it is challenging to predict workload patterns over time.

2.2. Importance of workloads modeling in cloud

Workload modeling increases the understanding of typical cloud workload patterns and leads to more informed decisions to better manage resources [3,7]. From the workload characterization, performance models can be constructed to support research topics such as energy-efficiency and resource management and to answer important research questions, such as: how are overall cloud data center utilization levels affected by user behavior? How cloud data center energy efficiency can be improved while the QoS is maintained?

In addition, workload modeling in cloud computing enables performance analysis and simulation, which brings benefits to cloud providers and researchers as it allows: (i) the evaluation, through simulation, of resource management policies allowing the improvement of cloud services' QoS; (ii) the evaluation of these policies without deployment and execution of the applications in expensive large-scale environments; and (iii) the simulation of realistic cloud environments with controlled modification, adjustment, and repetition. The use of simulation models enables the production of tracelogs based on realistic scenarios, filling the gap previously identified in the cloud computing area [3].