Simulation Modelling Practice and Theory 58 (2015) 239–254

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

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

# Open-source simulators for Cloud computing: Comparative study and challenging issues

Wenhong Tian [a,b,*], Minxian Xu [a], Aiguo Chen [c,*], Guozhong Li [a], Xinyang Wang [a], Yu Chen [a]

[a] School of Information and Software Engineering, University of Electronic Science and Technology of China, China
[b] Big Data Research Center, University of Electronic Science and Technology of China, China
[c] School of Computer Science and Engineering, University of Electronic Science and Technology of China, China

ARTICLE INFO

*Article history:*
Available online 12 August 2015

*Keywords:*
Cloud computing
Data centers
Simulators for Cloud computing
Resource scheduling

ABSTRACT

Resource scheduling in infrastructure as a service (IaaS) is one of the keys for large-scale Cloud applications. Extensive research on all issues in real environment is extremely difficult because it requires developers to consider network infrastructure and the environment, which may be beyond the control. In addition, the network conditions cannot be controlled or predicted. Performance evaluations of workload models and Cloud provisioning algorithms in a repeatable manner under different configurations are difficult. Therefore, simulators are developed. To understand and apply better the state-of-the-art of Cloud computing simulators, and to improve them, we study four known open-source simulators. They are compared in terms of architecture, modeling elements, simulation process, performance metrics and scalability in performance. Finally, a few challenging issues as future research trends are outlined.



## 1. Introduction

Cloud computing is developed based on various recent advancements in virtualization, Grid computing, Web computing, utility computing and related technologies. Cloud computing provides both platforms and applications on demand through the Internet or intranet [16]. Some of the key benefits of Cloud computing include the hiding and abstraction of complexity, virtualized resources and efficient use of distributed resources. Some examples of emerging Cloud computing platforms are Google App Engine [13], IBM blue Cloud [14], Amazon EC2 [6], and Microsoft Azure [17]. Cloud computing allows the sharing, allocation and aggregation of software, computational and storage network resources on demand. Cloud computing is still considered in its infancy as there are many challenging issues to be resolved [16,1,19]. Youseff et al. [15] establish a detailed ontology of dissecting Cloud into five main layers from top to down: Cloud application (SaaS), Cloud software environment (PaaS), Cloud software infrastructure (IaaS), software kernel and hardware (HaaS), and illustrate their interrelations as well as their inter-dependency on preceding technologies.

Cloud data center can be a distributed network in structure, which is composed of many computing nodes (such as servers), storage nodes, and network devices. Each node is formed by a series of resources such as CPU, memory, network bandwidth and so on. Each resource has its corresponding properties. There are many different types of resources for

* Corresponding authors at: School of Information and Software Engineering, University of Electronic Science and Technology of China, China (W. Tian).
*E-mail addresses:* wenhong_tian@uestc.edu.cn (W. Tian), xmxyt900@gmail.com (M. Xu), agchen@uestc.edu.cn (A. Chen), aa416517258@hotmail.com (G. Li), uestc_wxy@qq.com (X. Wang), uestc-chenyu@foxmail.com (Y. Chen).

http://dx.doi.org/10.1016/j.simpat.2015.06.002
1569-190X/

Cloud providers. The definition and model defined by this paper are aimed to be general enough to be used by a variety of Cloud providers. In this paper, we focus on Infrastructure as a service (IaaS) in Cloud data centers.

In a traditional data center, applications are tied to specific physical servers that are often over-provisioned to deal with workload surges and unexpected failures [5]. Such configuration rigidity makes data centers expensive to maintain with wasted energy and floor space, low resource utilizations and significant management overheads. With virtualization technology, today's Cloud data centers become more flexible, secure and on-demand allocating.

One key technology plays an important role in Cloud data center is resource scheduling. One of the challenging scheduling problems in Cloud data center is to consider allocation and migration of reconfigurable virtual machines and integrated features of hosting physical machines.

It is extremely difficult to research widely for all these problems in real platforms because the application developers cannot control and process network environment. What is more, the network conditions cannot be predicted or controlled.

The research of dynamic and large-scale distributed environment can be achieved by building data center simulation system, which supports visualized modeling and simulation in large-scale applications in Cloud infrastructure. Data center simulation system can describe the application workload statement, which includes user information, data center position, the amount of users and data centers, and the amount of resources in each data center. Using this information, data center simulation system generates requests and allocates these requests to virtual machines.

By using data center simulation system, application developers can evaluate suitable strategies such as distributing reasonable data center resources, selecting data center to match special requirements, improving resource utilization and load balancing, reducing total energy-consumptions, reducing costs and so on. We will look at some closely related work firstly.

### 1.1. Related work

There is quite intensive research conducted for Cloud simulators. In this paper, we concentrate on open-source simulators which we can easier access. Dumitrescu and Foster [8] introduce GangSim tool for grid scheduling. Buyya et al. introduce GridSim [20] toolkit for modeling and simulation of distributed resource management for grid computing. Calheiros et al. [21] introduce modeling and simulations of Cloud computing environments at application level, a few simple scheduling algorithms such as time-shared and space-shared are discussed and compared. Sakellari et al. [12] complement a survey of mathematical models, simulation approaches and testbeds in Cloud computing, which aims to enable researcher to find suitable modeling approach and simulation implementation. Ikram et al. [2] introduce a novel Cloud resource management service model and its simulation-based evaluations are mainly focusing on two applications dynamic service composition. Nuu et al. [23] propose a scheme for modeling and experimenting combined smart sleep and power scaling algorithms in energy-aware data center networks. Gurout et al. [22] provide a survey on energy-aware simulation techniques with DVFS (Dynamic Voltage and Frequency Scaling). CloudAnalyst [7] aims to achieve the optimal scheduling among user groups and data centers based on the current configuration. Both CloudSim and CloudAnalyst are based on SimJava [11] and GridSim [20], which treat a Cloud data center as a large resource pool and consider application-level workloads. Kliazovich et al. [10] propose an energy-aware simulation environment named GreenCloud for Cloud datacenters at package level. Nunez et al. [4] introduce a new simulator of Cloud infrastructure named iCanCloud using C++ and compare the performance with CloudSim. Tian et al. [27] propose CloudSched, a novel lightweight simulation tool for VM scheduling with lifecycle in Cloud data centers.

### 1.2. Comparative guideline of open-source Cloud simulators

Cloud simulators can be divided into various categories according to their features. In this section, we will give a brief comparison with different categories by extending the comparison category in [9]. The open-source simulators are selected because we can study their source codes in details, develop new algorithms and improve them if necessary. The four open source simulators, namely CloudSim, iCanCloud, GreenCloud, CloudSched, are representative of many related simulators because we study the architecture design, modeling elements, simulation process, performance metrics and scalability. These simulators have common features such as in architecture, modeling elements, simulation process as well as their own characteristics such as focusing on different service layers and with different performance metrics. CloudSim is well known simulator for Cloud computing, it can be extended easily but currently it does not consider parallel experiments or lifecycles of VMs. The iCanCloud implements parallel experiments but does not consider energy consumption or VM migration. GreenCloud models detailed energy consumptions for different physical components. CloudSched can model lifecycle of requests, and provide different metrics for load-balance, energy efficiency and utilization etc. Four open source Cloud data centers simulators (CloudSim, GreenCloud, iCanCloud, CloudSched) are compared together in Table 1.

#### 1.2.1. Platform

The platform that the simulator based on makes it bind with some specific features. CloudSim and CloudSched are both implemented with Java, so they can be executed on any machine installed JVM. While built based on GridSim and SimJava, CloudSim is heavy to execute. GreenCloud is an extension of NS2 network simulator, and it's a packet level simulator. As for iCanCloud, it's based on OMNET, which can simulate in-depth physical layer entities.