A Survey of Cloud Computing Simulations and Cloud Testing

Azn Oujani, azinoujani@wustl.edu (A project report written under the guidance of Prof. Raj Jain)

Abstract:
Cloud computing is a hot topic all over the world nowadays, through which customers can access information and computer power via a web browser. Hence, it eliminates the need for maintaining expensive computing facilities. The characteristics of a typical cloud are: on-demand access, scalability, elasticity, cost reduction, minimum management effort, and device/location independence. As the adoption and deployment of cloud computing increase, it is critical to evaluate the performance of cloud environments. Modeling and simulation technologies are suitable for evaluating performance and security issues. Testing cloud-based software systems needs techniques and tools to deal with infrastructure-based quality concerns of clouds. These tools can be built on the cloud platform to take advantage of virtualized platforms and services as well as substantial resources and parallelized execution.

This paper explores the concept of cloud computing, surveys various modeling and simulation techniques, and introduces cloud testing concepts along with the recently developed benchmarks that are used in cloud testing.

Keywords:

Table of Contents:
• 1. Introduction
  • 1.1. Introduction to Cloud Computing
  • 1.2. What Is A Cloud?
  • 1.3. Why Is The Cloud Used?
  • 1.4. How to Utilize Cloud Computing?
    ◦ 1.4.1. SaaS
    ◦ 1.4.2. IaaS
    ◦ 1.4.3. PaaS
  • 1.5. Issues with Performance Analysis in Cloud Computing
• 2. Cloud Computing Modeling and Simulations
  • 2.1. CloudSim
    ◦ 2.1.1. GreenSim
    ◦ 2.1.2. NetworkCloudSim
    ◦ 2.1.3. CloudAnalyst
    ◦ 2.1.4. EMUSIM
  • 2.2. SPECI
  • 2.3. GROUDSIM
  • 2.4. DCSim
• 3. Cloud Testing
  • 3.1. What Is Cloud Testing?
  • 3.2. Why Cloud Testing Is Used?
  • 3.3. New Requirements and Features in Cloud Testing
  • 3.4. Issues and Challenges in Cloud Testing
• 4. Cloud Testing Benchmarks
  • 4.1. YCSB
  • 4.2. Enhanced TPC-W
  • 4.3. CloudStone
  • 4.4. MalStone
• 5. Conclusion
• 6. List of Acronyms
• 7. References

1. Introduction
Cloud computing can be viewed from two different perspectives: cloud application, and cloud infrastructure as the building block for the up layer cloud application. It has achieved two important goals for distributed computing: high scalability and high availability. Scalability means that the cloud infrastructure can be expanded to very large scale while availability means that the services are available even with the failure of quite a number of nodes.

This paper first defines the concept and main categories of cloud computing along with different issues regarding performance analysis in cloud computing, and then explores recent solutions in modeling and simulation for cloud environments. Then it introduces cloud testing, and finally investigates common benchmarks to support cloud testing.

1.1. Introduction to Cloud Computing
Cloud computing can be viewed from two different perspectives: cloud application, and cloud infrastructure as the building block for the up layer cloud application. It has achieved two important goals for distributed computing: high scalability and high availability. Scalability means that the cloud infrastructure can be expanded to very large scale while availability means that the services are available even with the failure of quite a number of nodes.

This paper first defines the concept and main categories of cloud computing along with different issues regarding performance analysis in cloud computing, and then explores recent solutions in modeling and simulation for cloud environments. Then it introduces cloud testing, and finally investigates common benchmarks to support cloud testing.
1.2. What Is A Cloud?

A cloud is a combination of infrastructure, software, and services that are not local to a user. Data is neither stored on the local hard drive of your computer, nor on servers that are down in the basement of your company. Instead it is out in the cloud; the infrastructure is outside of your organization and you access the applications, the infrastructure, and all those services typically through the web-based interface.

1.3. Why Is The Cloud Used?

It is easy to use and efficient. Users do not have to put in software and also the on-demand, pay-as-you-go model creates a flexible and cost-effective means to access compute resources.

1.4. How is cloud computing utilized?

There are three general categories of cloud computing which are briefly discussed below.

1.4.1. SaaS(Software as a Service)

This type is the most popular way of using cloud computing. It utilizes a multitenant architecture, in which the system is built in a way that allows several customers to share infrastructure in an isolated way, without compromising the privacy and security of each customer's data.

Using multi-tenancy, cloud computing delivers an application through thousands of customers' web browsers. There is no upfront investment in servers or software licensing for customers, and maintaining just one application makes the cost low for providers compared to conventional hosting.

Google Doc, Gmail, and Zoho are some examples in this category.

1.4.2. IaaS(Infrastructure as a Service)

This type of cloud computing used to be called as HaaS (Hardware as a Service). IT infrastructure was rented out with pre-determined hardware configuration for a specific period of time, and the client had to pay for the configuration and time regardless of the actual use. With IaaS customer accesses the resources, IT infrastructure such as storage, processing, networking, over the Internet. The client can dynamically scale the configuration, and pay only for the services actually used.

IaaS is offered in three different models: Private, Public, and Hybrid.

Private cloud implies that the infrastructure resides at the customer premises, and the internal cloud is based on a private network behind a firewall. That is to say, the cloud services are implemented using resources of a single organization.

Public cloud is located on cloud computing platform vendor's data center and provides public accessible services over the Internet, i.e. a service provider makes resources available to the public over the Internet.

Hybrid cloud is a combination of the aforementioned models with customer choosing the best of each worlds. It allows an organization to provide and manage few of its resources in-house and has others externally.

Amazon S3, Simple Storage Service, is an example in this category.

1.4.3. PaaS(Platform as a Service)

This type of cloud computing not only deals with operation systems, but also provides a platform that you can run existing applications or develop and test new ones without hurting your internal system by allowing the customer to rent virtualized servers and associated services. Using PaaS, enables development teams that are geographically distributed to work on a single software development project.

AppEngine, and Bungee Connect are two examples in this category.

1.5. Issues with Performance Analysis in Cloud Computing

Massive scalability, and component interactions complexity, i.e. dynamic configuration, are major challenges that performance analysts face in large-scale cloud computing systems to determine the system performance characteristics since one of the key requirements in performance maintenance is to make sure that system performance is SLA-driven [1]. The System is managed dynamically by SLA, Service Level Agreement, which is a negotiated contract between a customer and a service provider that clarifies all service features that are to be provided, and consequently the policies that are to be taken in this way. An SLA generally uses response time, how quickly responses to requests need to be delivered, as a performance metric [27]. For instance, if the system encountered peaks in load, in order to abide by the committed service levels, it would create additional instances of the application on more servers.

In most cloud computing categories, service components are located on different hosts. Passing user requests through these components generates many different types of execution paths that make it challenging for performance analysts to determine system behavior. For instance, finding out which service component might be the main source of the issue when the system performance does not satisfy the expectation, or specifying the critical paths among execution paths could be challenging [1].

2. Cloud Computing Modeling and Simulations

Cloud simulators are required for cloud system testing to decrease the complexity and separate quality concerns. They enable performance analysts to analyze system behavior by focusing on quality issues of specific component under different scenarios [2].

Some of the published cloud computing simulators for evaluating cloud computing systems performance are described briefly in this section.

2.1. CloudSim

CloudSim is a simulation application which enables seamless modeling, simulation, and experimentation of cloud computing and the application services, proposed by [7,8,9] due to the problem that existing distributed system simulators were not applicable to the cloud computing environment.

The authors have mentioned that users could analyze specific system problems through CloudSim, without considering the low level details related to Cloud-based infrastructures and services. The layered CloudSim architecture is depicted in Figure1 [5].
Several works have been done from then on to improve CloudSim as are described briefly below.

2.1.1. CloudAnalyst

CloudAnalyst was derived from CloudSim and extends some of its capabilities proposed in [12, 13]. This simulator can be applied to examining the behavior of large scaled Internet application in a cloud environment and separating the simulation experimentation exercise from a programming exercise. It also enables a modeler to repeatedly perform simulations and to conduct a series of simulation experiments with slight parameters variations in a quick and easy manner. The CloudAnalyst architecture is shown in Figure 2 [5].

2.1.2. GreenCloud

GreenCloud is an improvement in CloudSim to prove the green cloud computing approach. The lack of detailed simulators, and having no provisioning system for analyzing energy efficiency of the clouds was the motivation behind GreenCloud development to interact and measure cloud performance. It's an advanced packet-level simulator with concentration on cloud communication proposed in [16]. It provides a detailed fine-grained modeling of the energy consumed by the data center IT equipment such as computing servers, network switches, and communication links. GreenCloud is considered as an extension of the network simulator Ns2 [17] [5].
2.1.3. NetworkCloudSim

NetworkCloudSim is an extension of CloudSim as a simulation framework that supports generalized applications such as high performance computing applications, workflows, and e-commerce besides real cloud data centers modeling proposed in [9]. The architecture of CloudSim-based NetworkCloudSim is depicted in Figure 3 [5].

![Architecture of CloudSim-based NetworkCloudSim](https://www.cse.wustl.edu/~jain/cse567-13/ftp/cloud/index.html)

2.1.4. EMUSIM

EMUSIM is an integrated architecture proposed in [18] to anticipate service's behavior on cloud platforms to a higher standard, which is built on Automated Emulation Framework (AEF) [25] for emulation and CloudSim for simulation [7] [5].

2.2. SPECI

The size of data centers that provide cloud computing services is increasing, and some middleware properties that manage these data centers will not scale linearly with the number of components.

Simulation Program for Elastic Cloud Infrastructures (SPECI) proposed in [15] based on [14] is a simulation tool, which allows analyzing large data center behavior under the size and design policy of the middleware as inputs. SPECI is composed of two packages: data center layout and topology, and the components for experiment execution and measuring [5].

2.3. GroudSim

GroudSim, proposed in [28], is an event-based simulator that needs just one simulation thread for scientific applications on grid and cloud environments. It is mainly concentrated on the IaaS, but it is easily extendable to support additional models like PaaS, or cloud storage [5].

More investigation was carried out in [19] in order to allow the user to simulate their experiments from the same environment used for real applications, by integrating GroudSim into the ASKALON environment [20] [5].

2.4. DCSim

DataCenter Simulator, proposed in [30], is concentrated on virtualized data centers, offering IaaS to Multiple tenants, in order to achieve a simulator to evaluate and develop data center management techniques. Figure 4 depicts the architecture of DCSim [5].
Table 1 [5], depicts the analysis and comparison of the cloud computing simulators based on underlying platform, developing language, software or hardware. Most of these simulators are software based and are developed using Java.

<table>
<thead>
<tr>
<th>Simulator</th>
<th>Underlying Platform</th>
<th>Programming Language</th>
<th>Software/Hardware</th>
</tr>
</thead>
<tbody>
<tr>
<td>CloudSim</td>
<td>GridSim</td>
<td>Java</td>
<td>Software</td>
</tr>
<tr>
<td>CloudAnalyst</td>
<td>CloudSim</td>
<td>Java</td>
<td>Software</td>
</tr>
<tr>
<td>GreenCloud</td>
<td>Ns2</td>
<td>C++, OTcl</td>
<td>Software</td>
</tr>
<tr>
<td>NetworkCloudSim</td>
<td>CloudSim</td>
<td>Java</td>
<td>Software</td>
</tr>
<tr>
<td>EMUSIM</td>
<td>AEF, CloudSim</td>
<td>Java</td>
<td>Software</td>
</tr>
<tr>
<td>SPECI</td>
<td>SimKit</td>
<td>Java</td>
<td>Software</td>
</tr>
<tr>
<td>GroudSim</td>
<td>-</td>
<td>Java</td>
<td>Software</td>
</tr>
<tr>
<td>DCSim</td>
<td>-</td>
<td>Java</td>
<td>Software</td>
</tr>
</tbody>
</table>

**3. Could Testing**

Cloud testing is defined as TaaS, Testing as a Service. This testing includes both functional testing, including redundancy and performance scalability, as well as non-functional testing, including security, stress, load, performance, and interoperability, of numerous applications and products.

**3.1. What Is Cloud Testing?**

Cloud testing is not testing the cloud. It is a subset of software testing in which cloud-based web applications are tested by simulated real-world web traffic. Migrating IT organizations to cloud solutions makes cloud testing essential to validate functional system and business requirements.

**3.2. Why Is Cloud Testing Used?**

There are two types of cloud testing services: On-Premise, On-Demand.

*On-Premise*: Cloud testing can be used for validating and verifying different products owned by individuals or organizations.

*On-Demand*: On-demand testing is getting increasingly popular nowadays, and it's used to test on-Demand software.

Cloud testing is used to take unique advantages, based on [3], such as:

- Using scalable cloud system infrastructure to test and evaluate system performance and scalability.
- Leveraging On-demand testing to perform extensive and effective real-time online validation.
- Reducing costs by taking advantage of using computing resources in clouds

The other key benefits are flexibility, simplicity, geographic transparency and traceability.
3.3. New Requirements and Features in Cloud Testing

New requirements and features in cloud testing, according to a survey in [3], are:

- **Cloud-based testing environment:** Using a selected IaaS or PaaS as a base to form a prepared test bed in which both virtual and physical computing resources can be included and deployed inside.

- **SLAs:** Service Level Agreements such as system reliability, availability, security, and performance agreements could be part of testing and quality assurance requirements.

- **Price models and service billing:** Price models and utility billing are basic parts and service for TaaS.

- **Large-scale cloud-based data and traffic simulation:** In performance testing and system-level function validation, simulating large-scale online user accesses and traffic data at interface connections is essential in cloud testing.

3.4. Issues and Challenges in Cloud Testing

Some of the issues and challenges in cloud testing based on the published papers are discussed below [3]:

- **Constructing on-demand test environments:** Providing on-demand testing environments is necessary for customers who want to test their applications on the cloud. There is no supporting solution to serve engineers in a cost-effective way to establish their required test environment in a cloud since most of the existing tools for testing cloud-based applications are not cloud-enabled. To overcome this shortage, TaaS providers are to provide a systematic solution that enables users to setup their required test environment based on their selection.

- **Scalability and performance testing:** Most of the published papers that have paid attention to scalability and performance testing so far such as [23][24][25][26], emphasize the scalability metrics and frameworks for parallel and distributed systems that have preconfigured resources and infrastructures. Therefore, metrics, frameworks, and solutions for these static systems can not consider scalable and dynamic testing environments, SLA-based requirements, and cost-models.

- **Testing security and measurement in clouds:** Providing secured services inside clouds is a crucial concern in modern SaaS and cloud technology. Assuring user privacy in a cloud infrastructure, guaranteeing the security of cloud-based applications inside a third-party cloud infrastructure, finding techniques, tools, and models for testing security of end-to-end applications in clouds, and determining the QoS standards for security oriented quality assurance for end-to-end applications in clouds, are some of the challenges in security validation and quality assurance.

- **Integration testing in clouds:** In a cloud infrastructure, engineers must deal with the integration of different applications in the cloud in a black-box view according to their APIs and protocols, and there is a lack of well-defined validation methods and quality assurance standards to address the connectivity protocols, interaction interfaces, and service APIs provided by applications and clouds APIs.

- **Regression testing issues and challenges:** Software changes and bug fixing would cause regression-testing challenges. We lack dynamic software validation methods and solutions in order to address these regression-testing issues, especially for on-demand software, and the dynamic features of SaaS and clouds

4. Cloud Testing Benchmarks

There is no single or ideal approach for cloud testing. This is basically due to the fact that there exists different factors such as cloud architecture design, non-functional and compliance requirements, etc., which need to be taken into account to ensure successful and complete testing when an organization starts cloud testing.

Some common benchmarks developed to support cloud testing are briefly introduced below.

4.1. YCSB

A key design goal of Yahoo! Cloud Serving Benchmark's tool [6] is extensibility; it is designed to be extensible and portable to mixed clouds to provide a comparison between cloud storage systems. Since this benchmark is under an open source license, others are able to use and extend the tool, and contribute new workloads and database interfaces. YCSB can be used to measure the performance of several cloud systems, and it is intended to deal with various quality concerns such as performance, scalability, availability and replication.

Performance examines the response time with increasing throughput until database saturation. Scaling tests how increasing in the number of machines affects system performance. Scale-up and speed-up are used as scaling metrics. A workload generator, which defines YCSB tool, utilizes user-defined workload descriptions, and a standard workload package, which is a collection of programs representing typical cloud operations [2].

4.2. Enhanced TPC-W

The architecture of TPC-benchmarks and its metrics are designed for transactional database systems therefore they are not suitable for cloud systems. Hence [22] suggested a new benchmark system specifically for cloud scalability, pay-as-you-go and fault-tolerance testing and evaluation. The benchmark defines web interactions as benchmark drivers with the usage of e-commerce scenarios. Scalability, fault tolerance, cost, and peaks are defined as new metrics for cloud storage system evaluation [2].

- **Scalability:** Cloud services are expected to scale linearly with a constant cost per web interaction. It's been proposed in [22] that the deviations of response time to the perfect linear scale can be measured by using correlation coefficient, R^2.

- **Fault tolerance:** Since hardware failures are common in IaaS, the metric is determined to analyze the potential of cloud self-healing. The recoverability of failures in a period of time is defined as the ratio between WIPS, Web Interactions Per Second, in RT, real-time, and Issued WIPS.

- **Cost:** Cloud Performance economy is measured by $/WIPS used by conventional TPC-W benchmark.

- **Peaks:** This metric is to measure how well a cloud can adapt to peak loads, scale-up and scale-down. The adaptability is defined by the ratio between WIPS in RT and Issued WIPS.

4.3. CloudStone

TCloudstone is an open source project [21] to deal with the new system architectures' performance characteristics, and to compare across different development stacks using social applications and automated load generators. The metric dollars-per-user-per-month regarding to realistic usage and cost is used to analyze cloud performance.

"Cloudstone is built upon open source tools including Olio, a social-event calendar web application, and Faban, a Markov-based workload generator, and automation tools to execute Olio following different workload patterns on various platforms." [2]
4.4. MalStone

MalStone is mainly designed for cloud middleware performance testing for data intensive computing by Open Data Group [23]. MalGen is developed to generate synthetic log-entity files that are used for testing inputs. It can generate tens of billions of events on cloud with over 100 nodes.

MalGen, an open source software package, generates the data for MalStone, synthetic log-entity files to test inputs. MalGen generates large, distributed data sets suitable for testing and benchmarking software designed for data intensive computing [2][14].

5. Conclusion

Cloud computing has been one of the fastest growing parts in IT industry. The characteristics of a typical cloud are: multi-tenancy, elasticity and scalability, service billing, connectivity interfaces and technologies, self-managed function capabilities, offering on-demand application services, providing virtual and/or physical appliances for customers [3].

It is necessary to evaluate performance and security risks that cloud computing is faced with, since users are concerned about security problems that exist with the prevalent implementation of cloud computing. Simulation-based approaches become popular in industry and academia to evaluate cloud computing systems, application behaviors and their security.

Several simulators have been specifically developed for performance analysis of cloud computing environments, including CloudSim, GreenSim, NetworkCloudSim, CloudAnalyst, EMUSIM, SPECi, GROUDSIM, and DCSim.

Cloud testing is becoming a hot research topic in cloud computing and software engineering community. This paper provides a review on cloud testing by discussing new requirements, issues, and challenges as well as conducting a survey of new benchmarks uniquely created for cloud testing, including YCSB, Enhanced TPC-W, CloudStone, and MalStone.

Research efforts continue, and researchers are working to further develop and expand upon measurement tools, models, and simulations especially for cloud computing environments.

6. List of Acronyms

SaaS Software as a Service
IaaS Infrastructure as a Service
HaaS Hardware as a Service
PaaS Platform as a Service
SI Simple Storage Interface
SLA Service Level Agreement
TaaS Testing as a Service
YCSB Yahoo! Cloud Serving Benchmark
TPC_W Transactional Processing Performance Council E-Commerce Benchmark
WIPS Web Interactions Per Second
RT Real Time

7. References

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6340205


http://seij.dce.edu/Paper%201.pdf


http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6486505

http://dl.acm.org/citation.cfm?id=1807152


http://arxiv.org/pdf/0907.4878v5

http://dl.acm.org/citation.cfm?id=1695506

http://link.springer.com/chapter/10.1007%2F978-3-642-21878-1_38?LI=true

http://www.cloudbus.org/students/MEDC_Project_Report_Bhathiya_318282.pdf

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5474733

http://www.opencloudconsortium.org


http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5683561

[17] The Network Simulator Ns2
http://www.isi.edu/nsnam/ns/

http://www.cloudbus.org/papers/EMUSIM-SPE.pdf


http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1542733


http://www.eecs.berkeley.edu/~kraska/pub/dbtest09-cloudbench.pdf

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=862209

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=242438

http://eprints.ucl.ac.uk/4990/1/4990.pdf

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4100388

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5610422

Last Modified: April 24, 2013
This and other papers on latest advances in Performance Analysis and Modeling are available on line at http://www.cse.wustl.edu/~jain/cse567-13/index.html
Back to Raj Jain's Home Page