Green Cloud Communication System for Big Data Management

Christos L. Stergiou, Kostas E. Psannis
Department of Applied Informatics
University of Macedonia
Thessaloniki, Greece
c.stergiou@uom.edu.gr
k.psannis@uom.edu.gr

Yutaka Ishibashi
Graduate School of Engineering
Nagoya Institute of Technology
Nagoya, Japan
ishibasi@nitech.ac.jp

***Abstract*—This paper makes an effort to survey and study the open challenges in the field of energy-efficient and green Cloud infrastructures. The useful software that offers the possibility to implement and evaluate Cloud environments is CloudSim, which we also use in order to demonstrate and propose our idea. Moreover, we consider the CloudSim's simulator architecture in order to achieve our proposal. Consequently, we investigate and propose a system framework for better use of Big Data management, based on Cloud federated network. Additionally, we propose an algorithm for achieving an energy-efficient resource allocation technique for Big Data management on Green Cloud environment. The experimental results demonstrate that our proposed model has immense potential as it offers significant performance gains regarding the cost saving and the better data management under large workload scenarios.**

Keywords-Cloud Computing; Green; Big Data; Management; Energy Efficient; CloudSim.

I. INTRODUCTION

This paper investigates the field of energy-efficient and green Cloud infrastructures. The challenge of consuming less energy by achieving more processing power is the new field arising, regarding the management and transfer of large data-sets need to be analyzed.

Cloud Computing is a technology that counts on huge data centers, and as a result counts on enormous amounts of electricity. Conventional Cloud applications offered from social networking, content delivery, web hosting, data processing and others. Such applications have multiple and differerent requirements. Also, volume of the system, and power performance are two open challenges of these applications on Cloud Service Providers. Cloud Service Providers (CSPs) use their infrastructures to provide all the available Cloud models (IaaS, PaaS, and SaaS) [1]. Energy amounts of such CSPs data centers are rising every year due to users demands. Cloud federation is a novel architecture that might offer possibilities and opportunities to address the open challenges.

Moreover, in order to achieve a better Cloud usage a lot of scenarios are proposed by the researchers. Cloud simulators are the most efficient method to facilitate and to operate a Cloud infrastructure without the additional costs of the hardware required. There are many Cloud Computing simulators, which are simulating many aspects of Cloud environments. The most widely known are: CloudSim, CloudAnalyst, GreenCloud, NetworkCloud, EMUSIM and MDCSim [2] [3]. In this paper we will use the CloudSim simulator in order to setup and evaluate our proposed scenario which relying on Green Cloud infrastructure.

Consequently, our basic contributions are listed bellow:

- Surveys related CloudSim energy-efficient scenarios.
- Considers CloudSim's simulator stable architecture.
- Investigate and propose a system framework for better use of Big Data management, based on Cloud federated network.
- Propose an algorithm for achieving an energy-efficient resource allocation technique for Big Data management.

The rest of this paper is organized as follows. Section 2 introduces the related work researches which have been made in the field of CloudSim simulation regarding energy-efficient scenarios. In addition, Section 3 provides details of our theoretical approach, count on CloudSim Architecture and the challenge of model of energy consumption. Section 4 offers the information about the experimental setup of our system and its evaluation process. Furthermore, Section 5, provides the experimental results and analysis of our system, along with the outcomes contacted from them. Finally, Section 6 concludes this paper and cites some future directions of our experiments.

II. RELATED WORK

For the purpose of our work we study and analyze previous literature which has been contributed with Cloud simulators, and more specifically with CloudSim.

There are a few studies that are more relevant to our work. Initially, Louis et al. [4] propose CloudSimDisk, which is a scalable module for modeling and simulation of energy-aware storage in a cloud system. Authors work's contribution is in the field of Cloud Computing, aiming to extend the widely used CloudSim simulator. Also, Dr. R. Malhotra & P. Jain [5] initially define the use of CloudSim. Thereafter, they explore all the variants that are available in Cloud Simulators, such as CloudAnalyst, GreenCloud, Network CloudSim, EMUSIM and MDCSim. Consequently, Dr. R. Malhotra & P. Jain compare the use of all CloudSim Variant with respect to networking, platform and language. Moreover, Buyya et al. [6] propose three aspects: (a) Architectural principles for energy-efficient management of Clouds, (b) Energy-efficient resource allocation policies and scheduling algorithms with consider to the quality-of-service expectations and devices power usage characteristics, and (c) A novel software technology for energy-efficient management of Clouds. Furthermore, A. V. Sajitha & Dr. A. C. Subhajini [3] present a comprehensive information about default energy conscious Virtual Machine placement algorithms in the CloudSim. Through their experiments, they can conclude that CloudSim simulator is better than the others regarding the energy oriented Data Center evaluation, especially in dynamic environments. They reach this conclusion because CloudSim is an extensible open source software. Finally, Long et al. [7] introduce CloudSim simulator as a framework which provides simulation scenarios for power to manage services and modeling of cloud infrastructure. In addition to this, they analyze CloudSim's architecture, and how it could be used it in order to model a Cloud environment.

TABLE I. RELATED WORK COMPARISON.

Papers / Features to explore	Louis et al. [4]	Dr. R. Malhotra & P. Jain [5]	Buyya et al. [6]	A. V. Sajitha & Dr. A. C. Subhajini [3]	Long et al. [7]
Fr	X				X
Pl		X	X	X	
Ar			X	X	X
BDP			X	X	
BDA		X			X
BDM	X	X	X	X	X
IS		X	X		X
CSM	CloudSimDisk	Multiple Simulators	CloudSim: Energy-Aware Scenario	CloudSim: Energy Efficient Green Cloud Data Center	CloudSim: Energy Efficient Cloud Environment

TABLE II. FEATURES TO EXPLORE ABBREVIATIONS.

Fr: Framework	**BDA:** Big Data Analytics
Pl: Platform	**BDM:** Big Data Management
Ar: Architecture	**IS:** Integration Scenario
DBP: Big Data Processing	**CSM:** CloudSim Model

Particularly, Table I illustrates the major features contributed more to our working scenario, as mentioned in previous works. Table II presents the meanings of the abbreviations listed in Table I. Specifically, we can observe that each work demonstrates a different CloudSim scenario regarding the simulation expectations. Additionally, three of the related works perform a platform scenario, three perform an architecture scenario and only two of them perform a framework scenario. In addition to this, two of the related works perform a combine scenario of platform and architecture and one of them perform a combine scenario of framework and architecture. Moreover, all of them contribute Big Data Management scenario, in contrast of Big Data Analytics and Big Data Processing that contributed by only two of them. Finally, three of the related works illustrates an integrated technologies scenario.

III. THEORETICAL APPROACH

A. *CloudSim Simulator Architecture*

CloudSim simulator provides a plenty of layers that works on them. Its layers support the modeling and the simulation of a Cloud environment. It also offers the ability of setting requirements about memory, storage and bandwidth parameters of both VMs and Cloud servers. The whole layer setup of CloudSim is shown in figure 1. Due to figure 1 and the architecture supported from CloudSim, a Cloud Service Provider (CSP) has the ability to evaluate a customized method based on these layers in order to verify and competitive the various policies of VM provisioning [8].

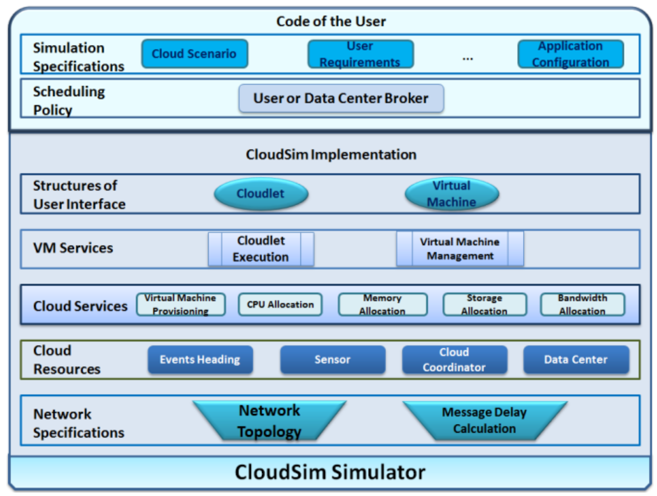

Figure 1. CloudSim Architecture with layers.

B. *Data center energy efficiency - Green Grid Consortium*

Finally, there is another scenario, due to "Green Grid Consortium", in order to calculate the Power Usage Effectiveness (PUE). This could be delivered by dividing the IT Equipment Power (ITEP) which demonstrates the energy facilities consumed from the equipment which is used in order to manage, process, transfer, store, operate, and route data through the data, with Total Facility Power (TFP) which demonstrates the data centers' entirely power that is delivered. This is an inverse version of PUE calculation, described as Data Center's Infrastructure Efficiency (DCIE), and illustrated by Equation (1).

$$DCIE = \frac{ITEP}{TFP} \qquad (1)$$

Count on the complexity of each Cloud environment and the forceful nature each resource demanding, Reinforcement Learning (RL) scenario could be utilized in order to suggest optimal allocation rules.

C. *Modeling Power Consumption*

Several works which have been made in this field [9] [10] show that applying Dynamic Voltage and Frequency Scaling (DVFS) on a CPU could result in nearly linear power to frequency relationship. This fact based both on the restricted number of the states that could place the value of frequency and voltage of a CPU, and also on the reason that the DVFS applied only on CPU and not to any other hardware component. Furthermore, related works exhibit that the consumption of an idle server is approximately in average of 75% power consumption of a server operating at maximum CPU speed. Regarding this, the switching of idle servers off aiming to reduce the overall Power Consumption is justified. This justification concludes in the definition of Equation (2) bellow:

$$P_s = pf * PC_m + (1 - pf) * PC_m * CPU_u \qquad (2)$$

In Equation (2), PC_m represents the maximum value of Power Consumption, in the moment where the server is fully utilized. Then, pf illustrates the value of the Power Consumed Fraction of the idle server, and CPU_u represents the utilization of the CPU. CPU's utilization might be changed during the time duration count on the variability of each process workload.

Consequently, the utilization of the CPU could also be demonstrated as a function of time. This can be represented as $CPU_u(t)$. Hence, the overall energy (OE) consumed by a settled node could be defined from Equation (3) bellow, as an integral part of the power consumption function for a period of time:

$$OE = \int_{time} P_s(CPU_u(t)) \qquad (3)$$

D. Cloud Federation approach

CSPs run various data centers, in many places simultaneously, through the Internet aiming to achieve the needs of multiple customers all over the world. Therefore, their systems which are currently established do not have the ability to support policies and mechanisms for dynamically coordinating load-shredding between various data centers with the aim to outline an optimal location in order to host application services and achieve rational QoS levels.

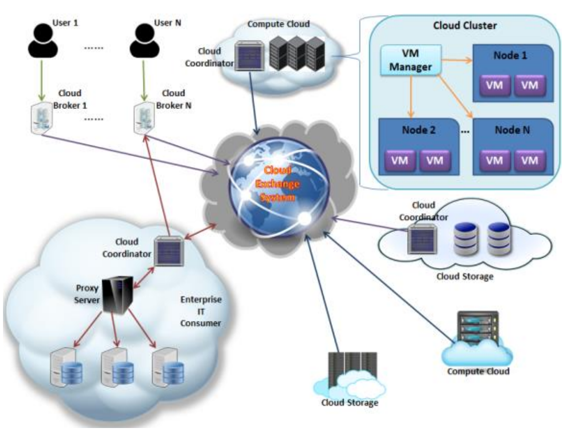

Figure 2. Federated Cloud Network.

Figure 2 illustrates a Cloud Computing system architecture consisting of services for different users' needs. These needs count on Cloud SaaS model, brokering, and CSPs coordinator services, which could be able to boost utility-driven Cloud services, such as application provisioning and workload migration.

A federated Internet-based system of supervisory distributed Cloud, which could provide particular benefits both in finance and performance. Such benefits could be the improvement of SaaS model's ability by achieving the QoS levels, and also the enhance of the peak-load of data that could be handled and the dynamic expansion of storage capacity to every user that could have access to the Cloud system. Thus, the Cloud federated system could enlarge the reliability of the CSPs participation except of ensuring the business continuity.

Cloud Coordinator (CCo) is one of the major components of Federated Cloud architecture. CCo is established in every Cloud system and it has the main responsibilities of system's operation. CCo could export Cloud services in whole federated infrastructure, and also has the ability to track the data load through the Cloud resources and then undertaking negotiation with the other CSPs established in the federation aiming to handle the sudden peak in the resource demand at each local Cloud system. In addition to this, CCo could monitor every application's execution and their lifecycle.

IV. Experimental Setup & Evaluation

In this section we try to clarify the way of thinking that led to the parameters set for the proposed scenario.

A. Data Center Configuration Setup

To our scenario, the data center needs to be created with at most 250 heterogeneous host machines, and at least 50 heterogeneous host machines. For each of them two types of server configurations are used as shown in Table III.

TABLE III. Server Configurations.

Host Type 1	Host Type 2
HP ProLiant ML110 G4	HP ProLiant ML110 G5
Intel Xeon 3040	Intel Xeon 3075
Cores: 2 (1800MHz/core)	Cores: 2 (2200MHz/core)
4GB RAM memory	4GB RAM memory
1TB storage memory	1TB storage memory

B. Experimental Configuration Setup

The whole simulation analysis contacted on the two aforementioned types of hosts, randomly selected, for respectively 50, 100, 150, 200, and 250 heterogeneous VMs each time. For each VM the configuration setup illustrated in Table 4.

TABLE IV. VMs Configurations.

Virtual Machine
1 x CPU (with 1 core)
1000 or 2000 MIPS capacity (randomly)
8GB RAM memory
2TB storage memory

The value of power consumed by the VMs is defined according our proposed model. Model's host consumption is set from 150Watt on 0% CPU utilization up to 300Watt on 100% CPU utilization. Each user who contacted the system is considered as a different independent user, who submit its requests to the host for provisioning of 50, 100, 150, 200, or 250 VMs that fills the full capacity of data center. The VM simulated to our scenario runs a web application with variable workload value, which affects to CPU's utilization. The web application runs to each VM operates for 150000 MIPS. The whole experiment runs for 7 days.

V. Experimental Analysis

In this section, the experimental results of the simulation scenarios implemented on CloudSim simulator along with the other findings are demonstrated and analyzed.

A. Algorithm Approach

Through this work we propose an algorithm scenario that embeds novel techniques for energy-efficient Big Data management in CloudSim toolkit. Reinforcement and Federated approaches are combined in our model in order to save computation power and distribute it to multiple CSPs.

As we already know from the literature, CloudSim could be used as a virtualization software aiming to simulate and operate multiple scenarios for the function of Cloud Data Center [10]. Also, CloudSim could be used as a framework for scalable simulation process which facilitates support scenarios and experimentations of virtual data centers in Cloud environments. As regards the data management, it offers scenarios of services for VMs, based on memory, storage and bandwidth, with multiple configurations. These scenarios could be established without limitation on data volume, and as a result could be operated on large-scale data.

Algorithm 1

CloudSim impute values:

- ➢ ***nH***: Total number of the various Hosts of data centers.
- ➢ ***nVM***: Total number of the Virtual Machines used in the operation process.
- ➢ ***WC***: The value of Workload Consumption of CPU's power.
- ➢ ***PF***: The value of Power Consumed Fraction of idle server.
- ➢ ***cu***: The value of CPU's utilization.
- ➢ ***N***: The maximum value of process operation.

CloudSim outcomes:

- ✓ OA: Overall Allocation of the VMs use
- ✓ OE: Overall Energy consumed by entire process

Proposed Scenario:

initialize nH, nVM, WC // create multiple Hosts, VMs and Workload Consumption of CPU's power
initialize Cloud Environment() // initialize state: ***S***, effect: ***E***, and ***Y*** values and ***X*** counter values
for (VM ϵ nVM) and (Host ϵ nH)
S = {WC, h, vm}
for i=0, i<=N, i++
E_i = E ϵ $\max_E * Y_i(S_i, E)$
with E_i count E_{i+1}
recompense X_{i+1}
$Y_{i+1}(S_i, E_i) \leftarrow X + [WC \cdot \max_E \cdot Y_i(S_{i+1}, E_{i-1})]$
$S_i = S_{i+1}$
end
$OE_N = (PF * WC + (S_i * Y_{i-1})) * cu_N$
return h
distribute(Host, VM)
end
return OA and OE

Regarding the proposed scenario, through CloudSim toolkit we can perform a better use and performance of the associate obstacles in order to achieve better services and policies based on data management techniques. Count on the existing CloudSim architecture, we could support Cloud data centers infrastructures with the Data Center's Infrastructure Efficiency (DCIE) applied on it. Therefore, our proposed CloudSim scenario illustrates a framework which encompasses all the factors needed and integrated in order to achieve an efficient and Green Cloud environment, based on the existing CloudSim architecture.

B. DCIE resource allocation results

The DCIE emphasis metric and its affect are produced over the 50% of overall scenario level. The proposed scenario contacted for seven days duration, for four dimension of the number of each VM.

Figure 3 demonstrates the DCIE metric comparison of our proposed scenario. Due to figure 3, the DCIE value of resource allocation shows that the Day 1 has the highest value of DCIE, in contrast with Day 7 which has the lowest value of DCIE. Moreover, the lowest value for each category of VMs are: a) for 50 VMs is approximately 47%, b) for 100 VMs is approximately 50%, c) for 150 VMs is approximately 52%, d) for 200 VMs is approximately 54%, and d) for 250 VMs is approximately 62%. Count on the aforementioned results, the value of DCIE increases by approximately 2% per day. As regards the grow of the value of DCIE, the impact of this value to the infrastructure efficiency of the data center, summarized from the proposed algorithm, additionally guarantees and provides such a good and efficient infrastructure environment.

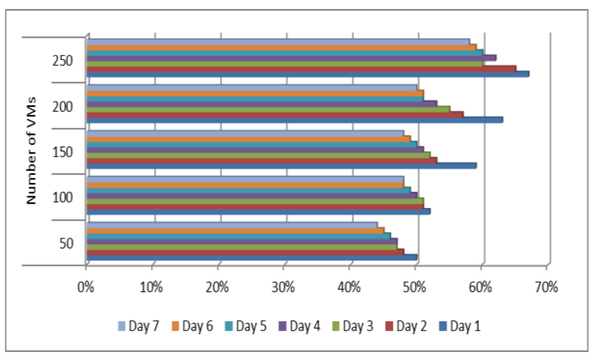

Figure 3. Metric of DCIE count on proposed scenario.

C. *Performance analysis*

The whole power consumed of the system could be delivered by multiple figures analyzed in the current subsection. The factors that we take into account are Power Consumed Fraction, CPU's utilization, Overall Energy consumed, and Data Center's Infrastructure Efficiency.

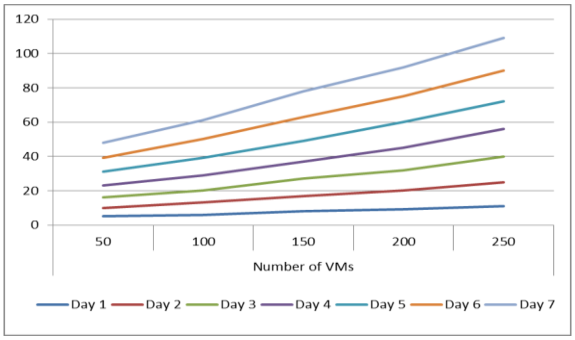

Figure 4. Power Consumed Fraction by energy-efficient algorithm.

Figure 4 presents the effects of Power Consumed Fraction of the simulation regarding the different need of VMs, over the seven days of the procedure. In particular, we can define that as the days past the need of power consumed rises. Additionally, as many VMs we contribute such many power resources needed.

VI. CONCLUSION & FUTURE WORK

After we ended up using the CloudSim simulator as the ideal Cloud simulator for our scenario, we investigated the related works on the field that try to offer energy-efficient and green Cloud environment for data management and processing through the CloudSim simulator. CloudSim could offer aspects such VMs and CCo that established and used equally in real-time scenarios and infrastructures. Based on our study, we proposed a system framework for better use of Big Data management, based on Cloud federated network. Furthermore, we tried to offer an algorithm approach of our scenario, by proposing a novel model for achieving an energy-efficient resource allocation technique for Big Data management on Green Cloud environment. Finally, we demonstrated experimental results which show that our proposed model has immense potential as it offers significant performance gains regarding the cost saving and the better data management under large workload scenarios.

For the future we are oriented to involve to the proposed framework the major issue of security, with the aim to provide a Green and Secure Cloud environment, through the scenarios of federated learning and reinforcement learning.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and feedback which was extremely helpful in improving the quality of the paper. Also, the whole work is a part of the PhD project contacted from C. L. Stergiou.

REFERENCES

[1] C. Stergiou, K. E. Psannis, "Recent advances delivered in Mobile Cloud Computing's Security and Management challenges", IGI Global, Modern Principles, Practices, and Algorithms for Cloud Security, 2019.

[2] C. Stergiou, K. E. Psannis, "Efficient and secure BIG data delivery in Cloud Computing", Springer, Multimedia Tools and Applications, vol. 76, issue: 21, pp. 22803–22822, November 2017.

[3] A. V. Sajitha, Dr. A. C. Subhajini, "Analysis of Cloud Sim Toolkit for Implementing Energy Efficient Green Cloud Data Centers", International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 6, issue IV, pp. 4613-4624, April 2018.

[4] B. Louis, K. Mitra, S. Saguna, C. Ahlund, "CloudSimDisk: Energy-Aware Storage Simulation in CloudSim", in Proceedings of 2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC), Limassol, Cyprus, 7-10 December 2015, pp. 11-15. [DOI: 10.1109/UCC.2015.15]

[5] Dr. R. Malhotra, P. Jain, "Study and Comparison of CloudSim Simulators in the Cloud Computing", The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), vol. 1, no. 4, pp. 1-5, October 2013. [DOI: 10.9756/sijcsea/v1i4/0104510201]

[6] R. Buyya, A. Beloglazov, J. Abawajy, "Energy-Efficient Management of Data Center Resources for Cloud Computing: A Vision, Architectural Elements, and Open Challenges", in Proceedings of the 2010 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2010), 12-15 July 2010, Las Vegas, USA. [arXiv:1006.0308]

[7] W. Long, L. Yuqing, X. Qingxin, "Using CloudSim to Model and Simulate Cloud Computing Environment", in Proceedings of 2013 Ninth International Conference on Computational Intelligence and Security, 14-15 December 2013, Leshan, China.

[8] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, R. Buyya, "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms", Wiley, Journal of Softwaer: Practice and Experience, vol. 41, issue: 1, pp. 23-50, January 2011. [DOI: 10.1002/spe.995]

[9] A. Gandhi, M. Harchol-Batler, R. Das, C. Lefurgy, "Optimal power allocation in server farms", in Proceedings of ACM 11th international joint conference on Measurement and modeling of computer systems, pp. 157-168, June 2009, Seattle, WA, USA.

[10] R. Raghavendra. P. Ranganathan, V. Talwar, Z. Wang, X. Zhu, "No "power" struggles: coordinated multi-level power management for the data center", ACM SIGARCH Computer Architecture News, vol. 36, no. 1, March 2008. [DOI: 10.1145/1353534.1346289]