XXI Escola Regional de Redes de Computadores (ERRC). DOI: 10.5753/errc.2024.4590

ARTIGO COMPLETO/FULL PAPER

Evaluating SimGrid and CloudSim Plus for Hybrid Cloud Scientific Workflows

João Antonio Soares • jansoares@inf.ufpel.edu.br
Universidade Federal de Pelotas (UFPel)

Rafael Burlamaqui Amaral • rafael.amaral@inf.ufpel.edu.br
Universidade Federal de Pelotas (UFPel)

Gerson Cavalheiro • gerson.cavalheiro@inf.ufpel.edu.br
Universidade Federal de Pelotas (UFPel)

ABSTRACT. Hybrid cloud, integrating public and private clouds, presents a promising environment for scientific applications by combining scalability with cost efficiency. However, the complexity of these environments requires tools to support infrastructure planning and optimization prior to actual deployment. This study evaluates two simulators: SimGrid and CloudSim Plus. The focus is on assessing their suitability for simulating the execution of scientific applications within hybrid cloud environments; particularly regarding scalability. Scientific workflows modeled as directed acyclic graph, were used to evaluate a hybrid cloud infrastructure, and both simulators were assessed on workload representation, cloud infrastructure, and scheduling strategies. The results suggest that SimGrid offers greater flexibility in network modeling, while CloudSim Plus excels in resource allocation policy simulation. This comparative analysis aims to help researchers select the most appropriate tool for simulating and optimizing scientific applications and cloud scheduling strategies. Future research should explore the integration of emerging technologies, such as containers and microservices, within these simulators.

KEYWORDS: cloud computing • scientific workflow • hybrid cloud

1 Introduction

Scientific and technological innovation has accelerated the adoption of flexible, intelligent computational infrastructures, prioritizing scalability, efficiency, and privacy [1]. Through service models such as IaaS, PaaS, and SaaS, cloud computing offers broad access to computational resources via virtualization [2]. On-demand services enable the scalable provisioning of resources, particularly for scientific applications.

Cloud infrastructures can be deployed in various configurations, including hybrid clouds, which combine private and public clouds to optimize resource usage [3]. Hybrid clouds leverage the scalability of public clouds to meet peak demands [4]. In academic institutions, local infrastructure provides scalable resources, with the option to expand into the public cloud during periods of high demand. However, this flexibility brings challenges, including configuring applications for distribution and managing performance bottlenecks [5]. Additionally, the complexity of integrating and managing resources across private and public clouds can complicate efficient usage [6].

In this context, simulation tools emerge as alternatives for modeling and evaluating different strategies before real-world deployment [7]. Given the variety of cloud simulators, each with its strengths and limitations, a detailed understanding of their capabilities is important for informed decision-making, both in academic and industrial contexts. While several simulators exist [8, 9], this paper focuses on SimGrid [10] and CloudSim Plus [11] due to their popularity. Thus, this paper proposes a comparison of two tools for simulating scientific applications in hybrid cloud, aiming to elucidate their strengths and limitations and assist in selecting the most suitable tool for different research and development needs.

The organization of this study is as follows: Section 2 compares the cloud simulators; Section 3 outlines the hybrid cloud scenario used; Section 4 presents experimental results; Section 5 reviews related work; Section 6 concludes and suggests future research directions.

2 Cloud Simulators

The evolution of cloud computing has brought challenges and opportunities in simulating scenarios involving complex infrastructures, such as hybrid clouds. Simulation have emerged as popular solutions among researchers and developers due to their ability to represent distributed computing scenarios and enable the analysis of different techniques and policies. This work addresses the comparison of simulation tools in terms of their capabilities and limitations for hybrid cloud scenarios.

One of the main differences lies in their structural approach to modeling distributed infrastructures. CloudSim Plus emphasizes modularity and extensibility, adhering to software engineering principles like SOLID and KISS, which makes its architecture flexible. This flexibility allows developers to develop new classes and customize existing modules. Its class hierarchy – including Datacenter, Host, Vm, and Cloudlet – is designed to represent specific elements of a cloud environment, enabling the configuration of diverse scenarios and the experimentation with various allocation policies.

SimGrid employs a realistic simulation model focused on accurately modeling resources and network interactions, featuring an architecture with various modules for complex distributed systems. Its S4U (SimGrid for You) interface helps define cloud components and manage interactions like network connections and process synchronization, excelling in network topology modeling for detailed control over communication infrastructure.

2.1 Workload Representation in Cloud Simulators

Workload representation is an aspect to consider when comparing simulators, as the tools differ in how they model and manage these loads.

CloudSim Plus uses the concept of a `cloudlet` to represent a workload. Each `cloudlet` is configured with parameters such as the number of instructions and input and output size, allowing for detailed customization of simulated applications. The `networkCloudlet` is a specialization of this class, which allows for the representation of communication between different tasks, although modeling dependencies between tasks depends on custom events and programmatic control.

SimGrid adopts a different approach to representing scientific applications and workflows. With the `Exec` and `Comm` classes, it is possible to describe computing and communication within the simulated infrastructure. Moreover, native support for modeling workflows using input files, such as `.dot` and `.json`, allows SimGrid to be used effectively to simulate complex applications, including those with data dependencies and synchronization.

While CloudSim Plus is more suitable for simulating individual and independent workloads, SimGrid offers considerable advantages in modeling workflows, being more appropriate for applications that require accuracy in modeling dependencies and communications.

2.2 Modeling Cloud Infrastructure with Simulation Support

When comparing the simulation tools, the capability to represent various cloud infrastructures, including public, private, and hybrid clouds, can be an important consideration. For example, CloudSim Plus allows for distinguishing between public and private infrastructures through the `DatacenterCharacteristics` class, which defines the characteristics of the datacenter, such as the type of infrastructure and the costs associated with resource use. In addition, the use of [12] allows scenarios to be described via YAML files, facilitating the reproduction of experiments, although the use of more advanced features still requires a programmatic approach.

SimGrid does not have a mechanism to differentiate public and private infrastructures but offers the ability to define different network zones, allowing the simulation of hybrid platforms. Network modeling is one area where SimGrid surpasses CloudSim Plus, offering detail and support for complex network topologies. Models like [13] are used to estimate network link latencies. Thus, while CloudSim Plus facilitates distinguishing between different types of infrastructure, SimGrid provides greater flexibility in modeling network topologies and characterizing distributed platforms.

2.3 Discussion

The comparison underscores their differing modeling capabilities. CloudSim Plus's modularity and extensibility make it ideal for researchers creating custom simulations, with Maven integration and examples enhancing usability. Conversely, SimGrid is tailored for simulating distributed systems with detailed communication models and network topologies, which are suitable for scenarios demanding realistic representations of distributed systems.

While both offer specific advantages in terms of usability, CloudSim Plus is more user-friendly for those seeking a smoother learning curve, while SimGrid requires a more significant initial investment in terms of understanding the modules and interface.

Therefore, the choice between CloudSim Plus and SimGrid should be guided by the specific requirements of each project. CloudSim Plus is more suitable for scenarios requiring modular implementation and ease of use, while SimGrid excels in environments that demand precise and detailed modeling, especially in relation to communication and resource distribution.

CloudSim Plus and SimGrid offer different advantages and challenges, depending on the requirements of hybrid cloud simulation. While CloudSim Plus is highly flexible and developer-friendly, SimGrid offers precision and direct features for modeling networks and distributed systems. The choice between them should be based on the specific objectives of each simulation, considering factors such as flexibility, detailing, precision, and usability. Thus, this work provides a critical analysis intended to assist in decision-making about which tool to use for projects involving the simulation of hybrid cloud environments.

3 The proposed Hybrid Cloud Scenario Architecture

This section introduces the architecture of the proposed hybrid cloud scenario, designed to evaluate the applicability of different cloud simulation tools in modeling complex scientific workflow applications. The hybrid cloud scenario is structured to replicate a real-world environment, allowing the comparison of simulator performance under varying conditions of resource availability and workload complexity.

The three capacities – low, medium, and high – represent different degrees of computational power and storage capabilities, intended to simulate both federated cloud environments and hybrid cloud solutions with public cloud extensions as shown in Table 1. For the low-capacity scenario, only one datacenter was employed, with two racks comprising a total of six computational nodes. The medium-capacity scenario added another datacenter, thereby increasing the available nodes to 18 distributed across five racks. Finally, the high-capacity scenario included a public cloud datacenter, bringing the total host count to 34 and incorporating enhanced processing capabilities through additional processing resources.

Table 1. Size Configurations

Size	*Hosts*	*PEs*	Network	Storage
small	6	72	40	1
medium	18	216	100	3
high	34	816	100	5

The public cloud extension was modeled to demonstrate cloud bursting – a mechanism where additional workloads are migrated to public infrastructure when private resources are exhausted. This hybrid setup provides insight into the advantages and costs associated with leveraging public cloud services, particularly concerning scalability and resource optimization.

4 Experimental Results

The experiments involved simulating multiple users deploying a scientific application on the infrastructure provided over a 24-hours. The date of user entry into the infrastructure was modeled using an exponential statistical distribution. Each user can allocate up to 4 VMs at startup. It is assumed that all users utilize their maximum quota of 4 VMs, and the VMs cannot be shared among users. The simulated workload is the synthetic workflow Montage https://github.com/wfcommons/WfInstances/tree/main/pegasus/montage, with 391 tasks, where each user submits a workflow for execution.

The values in Table 2 represent the average of 15 runs on each simulator, totaling 30 experiments for each infrastructure size. The results indicate that infrastructure capacity has a considerable impact on application performance. Low-capacity infrastructure frequently experiences overload, with resource utilization rates exceeding 100%. In contrast, medium-capacity infrastructure reduces execution time to 1203.78 seconds, which suggest performance improvements due to greater resource availability.

Table 2. Example of Montage workflow [14] simulated makespan

Size	Workflow Duration(s)	Average Task Duration (s)
small	2921.89	267.09
medium	1203.78	162.36
high	786.23	120.6

The high-capacity infrastructure achieves the best performance, with an average execution time of 786.23 seconds and a more balanced resource utilization rate over time. However, the performance improvement compared to the medium-capacity infrastructure is less significant. This reduced gain is likely due to the medium-capacity infrastructure already meeting the processing requirements of the workflow, with the additional resources of the high-capacity infrastructure providing only marginal benefits.

The analysis of the results reveals that the efficiency of Montage execution improves significantly with an increase in infrastructure capacity, especially when moving from low to medium capacity. The high-capacity infrastructure, while offering the best performance, shows marginal gains compared to the medium-capacity infrastructure. The results suggest that, for a small Montage workflow, choosing a medium-capacity infrastructure may be the most cost-effective in terms of cost-benefit, meeting the processing needs without incurring excessive costs.

4.1 Simulation Performance Comparison

The experiment demonstrates the versatility of cloud simulators in modeling complex scientific workflows and evaluating their execution in different cloud environments. SimGrid showed greater efficiency in terms of simulation runtime compared to CloudSim Plus, as seen in Table 3. This difference in performance is attributed to the underlying programming language (C++ for SimGrid versus Java for CloudSim Plus), as well as the native handling of task concurrency and resource allocation strategies.

Table 3. Simulation Execution Time

Simulator	Total Simulation Time (s)	Standard Deviation (s)
CloudSim Plus	181.2	36.82
SimGrid	44.4	5.70

The experiments also highlight the importance of selecting the appropriate infrastructure capacity for scientific workflows. Highly parallel workflows like Montage benefit significantly from higher capacity setups, emphasizing the trade-off between cost and performance.

In conclusion, the use of cloud simulation tools such as CloudSim Plus and SimGrid effectively facilitates the evaluation of scalability in hybrid cloud environments. Such tools allow researchers to conduct extensive experiments under controlled settings, avoiding the prohibitive costs and complexity of real-world cloud deployments while providing insight into workload behavior and infrastructure requirements.

5 Related Work

Cloud computing has been studied in terms of innovations in distributed resource management. However, due to the complexity and cost of managing real platforms, simulation tools have emerged as a practical alternative for studying cloud environments. Consequently, the literature broadly evaluates these tools to support research.

In [9], several cloud simulation tools are used to help researchers choose the appropriate tool for specific projects. Their study highlighted that while tools such as CloudSim provide general flexibility, they often need to catch up when it comes to handling emerging needs like mobile cloud computing, where performance and adaptability are significant.

[15] conducted an extensive survey on cloud computing simulators, covering the architectures and features of 33 different tools. They found that a significant portion of these tools are based on the CloudSim framework, yet many need help with scalability and comprehensive modeling. Their findings underscore the need for improved models to meet the growing complexity of cloud environments, especially in hybrid settings.

[16] reviewed cloud simulation tools, focusing on their capabilities for dynamic workload management and energy modeling. They pointed out that most tools, including those derived from CloudSim, need more robust support for dynamic changes and energy efficiency, which limits their effectiveness in modern cloud scenarios where adaptability is critical.

Unlike the reviewed works, which focus on generalized cloud modeling or public clouds, this study takes a practical approach to evaluate these tools in a hybrid cloud infrastructure scenario modeled as Directed Acyclic Graphs (DAGs). It shows how these simulators address the specific needs of scientific applications, considering the diversity in workflow structures, dependencies, and resource requirements.

6 Conclusion and Future Work

This study demonstrated the applicability of cloud simulation tools for modeling hybrid cloud infrastructures to support scientific applications. By comparing SimGrid and CloudSim Plus, we identified distinct strengths and weaknesses. SimGrid offers flexibility, albeit with more significant programming effort and fewer cloud-specific features. CloudSim Plus provides a comprehensive environment for modeling cloud scenarios, but is limited in network aspects. The results emphasize the value of these tools for developing effective cloud resource management strategies for scientific workflows.

For future work, we aim to explore network considerations for workload scheduling and execution in greater detail and extend these tools to support emerging cloud technologies, such as lightweight virtualization through containers and microservices-based workload models. These improvements will help ensure that these tools remain relevant and scalable options for simulating diverse and complex hybrid cloud scenarios.

Declarations

Availability of data and additional materials

The implementations in the respective simulators are publicly available in the online code repository at: https://github.com/JoaoNevesSoares/simcloud.git.

References

1 Carrillo, O. *et al.* Scalable Architectures to Support Sustainable Advanced Information Technologies. *In*: 2022 IEEE INTERNATIONAL CONFERENCE ON CLUSTER COMPUTING (CLUSTER). *2022 IEEE International Conference on Cluster Computing (CLUSTER)*. Heidelberg, Germany: IEEE, 2022. P. 512–515. ISBN 978-1-66549-856-2. DOI: 10.1109/CLUSTER51413.2022.00065. Available from: https://ieeexplore.ieee.org/document/9912677/. Visited on: 31 May 2024.

2 Mell, P. M.; Grance, T. *SP 800-145. The NIST Definition of Cloud Computing*. Gaithersburg, MD, USA, 2011.

3 Sotomayor, B. *et al.* Virtual Infrastructure Management in Private and Hybrid Clouds. *IEEE Internet Computing*, v. 13, n. 5, p. 14–22, 2009. ISSN 1089-7801. DOI: 10.1109/MIC.2009.119. Available from: http://ieeexplore.ieee.org/document/5233608/. Visited on: 18 Mar. 2024.

4 Mansouri, Y.; Prokhorenko, V.; Babar, M. A. An automated implementation of hybrid cloud for performance evaluation of distributed databases. *Journal of Network and Computer Applications*, v. 167, p. 102740, 2020. ISSN 1084-8045. DOI: https://doi.org/10.1016/j.jnca.2020.102740. Available from: https://www.sciencedirect.com/science/article/pii/S1084804520302149.

5 Maarouf, A.; Marzouk, A.; Haqiq, A. Comparative Study of Simulators for Cloud Computing. *In*: 2015 INTERNATIONAL CONFERENCE ON CLOUD TECHNOLOGIES AND APPLICATIONS (CLOUDTECH). *2015 International Conference on Cloud Technologies and Applications (CloudTech)*. Marrakech, Morocco: IEEE, 2015. P. 1–8. ISBN 978-1-4673-8149-9. DOI: 10.1109/CloudTech.2015.7336989. Available from: http://ieeexplore.ieee.org/document/7336989/. Visited on: 22 Feb. 2024.

6 Khan, S. U.; Ullah, N. Challenges in the Adoption of Hybrid Cloud: An Exploratory Study Using Systematic Literature Review. *The Journal of Engineering*, v. 2016, n. 5, p. 107–118, 2016. ISSN 2051-3305, 2051-3305. DOI: 10.1049/joe.2016.0089. Available from: https://onlinelibrary.wiley.com/doi/10.1049/joe.2016.0089. Visited on: 24 May 2024.

7 Buyya, R. *et al.* Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility. *Future Generation Computer Systems*, v. 25, n. 6, p. 599–616, 2009. ISSN 0167739X. DOI: 10.1016/j.future.2008.12.001. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0167739X08001957. Visited on: 31 Dec. 2023.

8 Lata, S.; Singh, D. Cloud simulation tools: A survey. *In*: AIP PUBLISHING, 1. AIP Conference Proceedings. 2022. v. 2555.

9 Shahid, M. A. *et al.* A Systematic Survey of Simulation Tools for Cloud and Mobile Cloud Computing Paradigm. *Journal of Independent Studies and Research Computing*, 2022. Available from: https://api.semanticscholar.org/CorpusID:251195288.

10 Casanova, H.; Legrand, A.; Quinson, M. SimGrid: A Generic Framework for Large-Scale Distributed Experiments. *In*: TENTH International Conference on Computer Modeling and Simulation (uksim 2008). 2008. P. 126–131. DOI: 10.1109/UKSIM.2008.28.

11 Silva Filho, M. C. *et al.* CloudSim Plus: A Cloud Computing Simulation Framework Pursuing Software Engineering Principles for Improved Modularity, Extensibility and Correctness. *In*: 2017 IFIP/IEEE SYMPOSIUM ON INTEGRATED NETWORK AND SERVICE MANAGEMENT (IM). *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. Lisbon: IEEE, 2017. P. 400–406. ISBN 978-3-901882-89-0. DOI: 10.23919/INM.2017.7987304. Available from: https://ieeexplore.ieee.org/document/7987304/. Visited on: 31 Dec. 2023.

12 Silva Filho, M. C.; Rodrigues, J. J. P. C. Human Readable Scenario Specification for Automated Creation of Simulations on CloudSim. *In*: Hsu, R. C.-H.; Wang, S. (Eds.). *Internet of Vehicles – Technologies and Services*. Cham: Springer International Publishing, 2014. P. 345–356. ISBN 978-3-319-11167-4.

13 Dabek, F. *et al.* Vivaldi: A decentralized network coordinate system. *ACM SIGCOMM Computer Communication Review*, ACM New York, NY, USA, v. 34, n. 4, p. 15–26, 2004.

14 Jacob, J. C. *et al.* Montage: An astronomical image mosaicking toolkit. *Astrophysics Source Code Library*, ascl–1010, 2010.

15 Mansouri, N.; Ghafari, R.; Zade, B. M. H. Cloud Computing Simulators: A Comprehensive Review. *Simulation Modelling Practice and Theory*, v. 104, p. 102144, 2020. ISSN 1569190X. DOI: 10.1016/j.simpat.2020.102144. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1569190X20300836. Visited on: 31 Dec. 2023.

16 Fakhfakh, F.; Kacem, H. H.; Kacem, A. H. Simulation Tools for Cloud Computing: A Survey and Comparative Study. *In*: 2017 IEEE/ACIS 16TH INTERNATIONAL CONFERENCE ON COMPUTER AND INFORMATION SCIENCE (ICIS). *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. Wuhan, China: IEEE, 2017. P. 221–226. ISBN 978-1-5090-5507-4. DOI: 10.1109/ICIS.2017.7959997. Available from: http://ieeexplore.ieee.org/document/7959997/. Visited on: 5 Feb. 2024.