

Assessing Cost and Response Time of a Web Application Hosted in a Cloud Environment

Karim Abouelmehdi, Abderrahim Beni-hssane, Zakaria Benlalia, Abdellah Ezzati and Abdelmajid Moutaouikkil

Abstract Cloud computing offers a possibility of web applications hosting more cheaply and easy compared to traditional hosting in servers that have very limited capacities. The main objective of our study is to show the importance of assessing cost, and response time of a web application which will be hosted in a cloud environment, in several geographical areas before choosing the optimum configuration as cloud solution provider.

Keywords Cloud computing · Hosting cost · Datacenter response time · Web application

1 Introduction

Cloud computing progress open up many new opportunities for developers of Internet applications. Previously, the main concern of developers was the deployment and application hosting because it required acquisition of a server which has a fixed capacity capable, in peak demand, to manage expected applications, installation and maintenance of the entire software infrastructure of the platform supporting the application.

K. Abouelmehdi (✉) · A. Beni-hssane · Z. Benlalia · A. Moutaouikkil
LAROSERI laboratory, Sciences Faculty, Computer Science Department,
Chouaib Doukkali University, El Jadida, Morocco
e-mail: karim.abouelmehdi1@gmail.com

A. Beni-hssane
e-mail: abenhssane@yahoo.fr

Z. Benlalia
e-mail: benlalia.zakaria@gmail.com

A. Ezzati
LAVETE laboratory, Sciences and Technics Faculty, Hassan 1st University,
Settat, Morocco
e-mail: abdezzati@gmail.com

In addition, the server was underused because the peak traffic is occurred only at specific times. With the advent of cloud, deployment and hosting has become cheaper and easier. Faced with the ongoing rise in costs of implementation and maintenance of IT systems, companies are increasingly outsourcing their IT services by entrusting them to specialized enterprises such as cloud providers. In this context, a number of challenges including isolation (resources, failures, performances and user spaces), optimal management of resources and security arise to allow a real development of cloud [1].

In this study, we show the importance of assessing cost, and response time of a web application, which will be hosted in a cloud environment in several geographical areas, before choosing cloud provider.

2 Related Works

Many studies using simulation techniques to study the behavior of large-scale distributed systems and tools to support such researches. Among these simulators: Micro-grids [2], Christophe Gaultier [3], SimGrid [4], GridSim [5] and CloudANALYST [6]. While the first three simulators focus on Grid computing systems, the Cloudanalyst is the only simulation Framework designed for Cloud computing systems study. However, grid simulators were used to evaluate the costs of running distributed applications in cloud infrastructures [7, 8]. In 2012, Dhaval and all [9] proposed an algorithm that selects the most cost-effective datacenter. This reduces the overall cost, but increases the response time. This algorithm is tested by cloud analyst. In 2012, Sarfaraz [10] also proposed a solution that selects the data center with a high-performance configuration. This reduces the response time, but increases the overall cost. It is also tested by Cloud analyst.

The objective of our study is to show the simplicity of using cloud technology to host a large-scale web application in several geographic areas, by assessing, with CloudAnalyst simulator, the hosting cost and response time. And then make decision of the convenient configuration.

3 Case Study: Case Simulation of a Social Network Application

In this case study that is made by the simulator CloudAnalyst we worked with data from social network application “Facebook” on which we will calculate the transfer time and cost of deploying Datacenter. For each region, we used several scenarios. We will work with the data found in 2012 with a scale of 1/1000. This application is accessed by users from all around the world. Asia (2678100 users), North America (2365200 users), South America (1444000 users), Europe (2499900 users), Oceania (146200 users) and Africa (504,300 users).

Table 1 User bases used in the experiment

User base	Region	Peak hours	Online users during peak hrs	Online users during off-peak hrs
UB1	N. America	3:00–9:00	236520	23652
UB2	S. America	3:00–9:00	144400	14440
UB3	Europe	3:00–9:00	249990	24999
UB4	Asia	3:00–9:00	267810	26781
UB5	Africa	3:00–9:00	50430	5043
UB6	Oceania	3:00–9:00	14620	1462

3.1 Configuring the Simulation

Table 1 Based on the above information, we set for the number of users each user group was approximately 10 % of all users during peak hours, the number of users accounting for nonpeak hours to 10 % of peak time users. The detailed settings are showed in Table 1. And we assume the Americas, Europe, Oceania user requests once every 5 min; Asia users request once every 3 min; Africa once the user requests every 10 min. The user requests data size is 100 b each time [4]. Cloud resource services agency policy is the closed data center policies. Data center contains 100 identically configured servers, each with the x86 architecture, Linux operating system, 2 G memory, 100 G storage spaces, and four processors. And the processor speed is 10,000 MIPS; Bandwidth is 1000 Mbps; Virtual machine resource scheduling policy is the time-sharing. Applications in the experiments virtual local storage size is 100 MB, the virtual machine has 1 GB of RAM and 10 MB of available bandwidth. Users are grouped by a factor of 1000, and requests are grouped by a factor of 100. Each user request requires 250 instructions to be executed. In terms of the cost of hosting applications in a Cloud, we assume a pricing plan which closely follows the actual pricing plan of Amazon EC2. The assumed plan is: Cost per VM per hour (1024 Mb, 100 MIPS): \$ 0.10; Cost per 1 Gb of data transfer (from/to Internet):\$0.10 [11].

3.2 Simulated Scenarios

Several scenarios are considered in our case study. The easiest one is to model the case in which only one centralized data center is used to host the application of social network “Facebook”. In this model, all requests of all users across the globe are processed by this single data center. A data center has 40 virtual machines assigned to the application is located in North America.

The second scenario consists to use two data centers, each one of them has 20 virtual machines dedicated to the application and they are located in Europe and North America. **The third scenario** is similar to the second one; the only difference

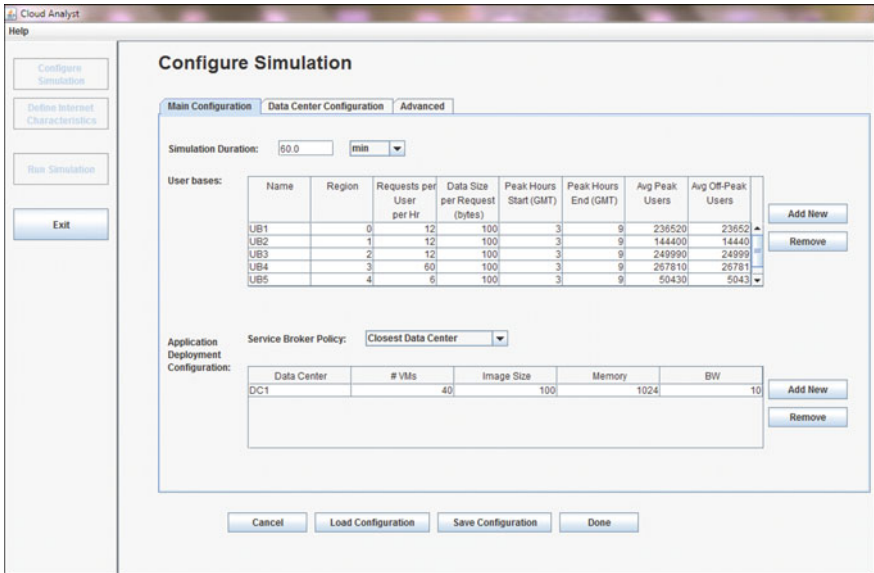


Fig. 1 Data entry

is that each data center has 40 virtual machines without load sharing between them. **The fourth and fifth scenarios** are similar to the third one. We just changed the load balancing algorithm. In **the sixth scenario** three data centers, each one has 40 virtual machines, are used. In this case, data centers are located in North America, Europe and Asia. Finally, in the last scenario, three data centers each contains the following number of virtual machines (60, 40, and 20) are also located in North America, Europe and Asia (Figs. 1 and 2).

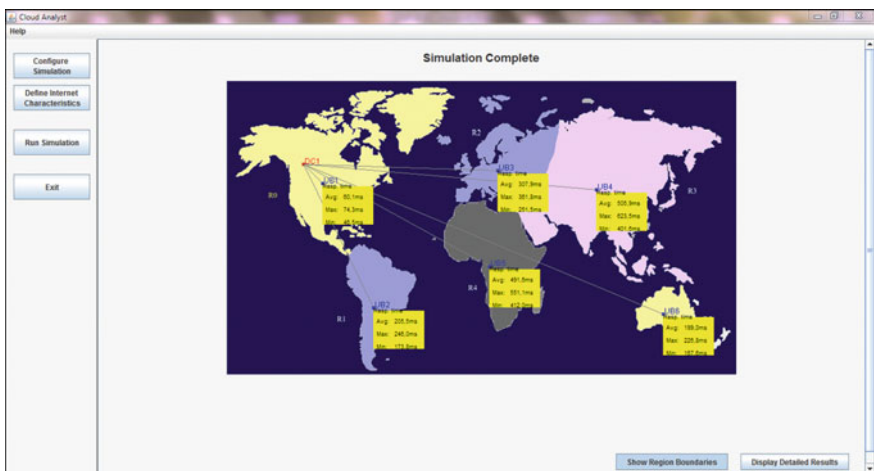


Fig. 2 Result of the first scenario

Table 2 Simulation results

	Configuration	Load balancing policy across VM	Overall average response time (ms)	Average data centre processing time (ms)	VM cost (\$/hrs)	Data transfert cost (\$/hrs)	Total cost ()
1	1 data center with 40 VMs (N. America)	Round robin	401,34	3,75	4,02	23,33	27,35
2	2 data centers with 20 VMs each (N. America, Europe)	Round robin	242,52	9,49	4,02	23,33	27,35
3	2 data centers with 40 VMs each (N. America, Europe)	Round robin	237,72	4,61	8,03	23,33	31,37
4	2 data centers with 40 VMs each (N. America, Europe)	Equally spread current execution load	237,74	4,56	8,03	23,33	31,37
5	2 data centers with 40 VMs each (N. America, Europe)	Throttled	236,82	3,69	8,03	23,33	31,37
6	3 data centers with 40 VMs each (N. America, Europe, Asia)	Throttled	73,57	5,93	12,05	23,33	35,38
7	3 data centers with 60, 40, 20 VMs (N. America, Europe, Asia)	Throttled	78,13	10,65	12,05	23,33	35,38

3.3 Results

(Table 2).

3.4 Discussion of Results

For a simple and visible debate, the results are shown in Figs. 3 and 4

- When more than Datacenter are used, the average response time is reduced
- The higher the number of virtual machines in the data center, the higher the average time of datacenter treatment is minimal even in the case of a single Datacenter. Therefore the most important is the number of VMs and not the number of Datacenter

Fig. 3 Performance comparison with recent data center services agency policy

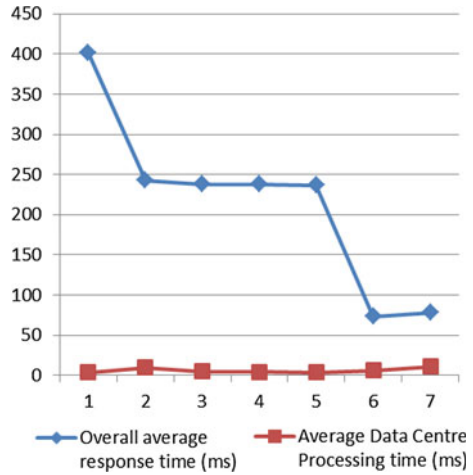
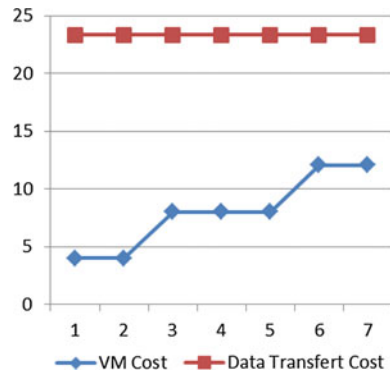


Fig. 4 The variation in costs



- The greater the number of virtual machines, the more the cost is high
- The data transfer costs remains constant because we worked with the same amount of information on the 7 scenarios.

4 Conclusion

In this work, we showed the importance of assessing cost, and response time of a web application to be hosted in a cloud environment in several geographic areas before choosing cloud provider. In fact, it can provide an idea about the optimum configuration depending on what suits us with an optimum cost before committing to a Cloud solution provider. This type of simulation experiment has a great potential to help testers to identify new features in order to model and develop new mechanisms and algorithms for managing cloud resources.

References

1. Weiss, A.: Computing in the clouds. *netWorker* **11**, 16–25 (2007)
2. Liu, X., Song, H., Jakobsen, D., Bhagwan, R., Zhang, X., Taura, K., Chien, A.: The MicroGrid: a scientific tool for modeling computational Grids. In: Proceedings of the ACM/IEEE Supercomputing Conference, IEEE Computer Society, Nov. 2001
3. Dumitrescu, C., Foster, I.: GangSim: a simulator for grid scheduling studies. In: Proceedings of the 5th International Symposium on Cluster Computing and the Grid (CCGrid 05), IEEE Computer Society, May 2005
4. Legrand, A., Marchal, L., Casanova, H.: Scheduling distributed applications: the SimGrid simulation framework. In: Proceedings of the 3rd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid 07), pp. 138–145, May 2001
5. Buyya, R., Murshed, M.: GridSim: a toolkit for the modelin and simulation of distributed resource management and scheduling for Grid computing. *Concurr. Comput. Pract. Exp.* **14** (13), 1175–1220 (2002)
6. Buyya, R., Ranjan, R., Calheiros, R.N.: Modeling and simulation of scalable cloud computing environments and the cloudsim toolkit: challenges and opportunities. In: Proceedings of the 7th High Performance Computing and Simulation Conference (HPCS 09), IEEE Computer Society, June 2009
7. Deelman, E., Singh, G., Livny, M., Berriman, B., Good, J.: The cost of doing science on the Cloud: the Montage example. In: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing, IEEE, Nov. 2008
8. Assunção, M., di Costanzo, A., Buyya, R.: Evaluating the cost benefit of using cloud computing to extend the capacity of clusters. In: Proceedings of the 18th International Symposium on High Performance Distributed Computing. ACM Press, June 2009
9. Limbani, D., Oza, B.: A proposed service broker strategy in cloudanalyst for cost-effective data center selection. *Int. J. Eng. Res. Appl. India* **2**(1), 793–797 (2012)
10. Ahmed, A.S.: Enhanced proximity-based routing policyfor service brokering in cloud computing. *Int. J. Eng. Res. Appl. India* **2**(2), 1453–1455 (2012)
11. Gustedt, J., Jeannot, E., Quinson, M.: Experimental methodologies for large-scale systems: a survey. *Parallel Process. Lett.* **19**, 399–418 (2009)