A Quantitative Study on Live Virtual Machines Migration in Virtualized Computing Environment

Marcela Tassyany Galdino Santos, Edlane de Oliveira Gusmão Alves, Anderson F. B. F. da Costa
Federal Institute of Paraíba (IFPB)
Campina Grande, Brazil
e-mail: marcelatassyany@ieee.org, edlaneoliveira@ieee.org, anderson@ifpb.edu.br

***Abstract*—The live virtual machines migration is a widely used technique in cloud computing environments because it is not necessary to stop services hosted on the migrated virtual machines. This paper aims to conduct an experimental and quantitative study on the migration of virtual machines through the evaluation of different scenarios. For this, a fully virtualized computing environment was implemented with the purpose of performing experiments that allow to analyze the influence of the live migration process in the performance of the services offered by the virtual machines migrated. The following metrics were evaluated: number of requests per second, server response time, throughput, latency and total migration time. Benchmarks ab- Apache Benchmark and YCSB-Yahoo! Cloud Serving Benchmark were used to generate the workload for the Web services and Database (Cassandra), respectively. The results obtained revealed that during the migration period, the services considered presented a reduction in their performance, but although there is a decrease, the service is not interrupted, thus complying with the principles of live migration.**

Keywords - virtualization; virtual machines; live migration; xen.

I. INTRODUCTION

The current scenario of globalization has generated the need for organizations, in general, to seek a flexible Information Technology (IT) infrastructure [1]. In this scenario, virtualization appears as a central element in datacenters environments, as it allows the optimization of idle resources, thus reducing costs with equipment and energy, optimizing the use of physical space, as well as enabling rapid alteration of the computing infrastructure and the portability of computational systems.

It is possible to define virtualization as a layer of software that allows partitioning a computational system from a physical machine into independent virtual machines that simulate different systems [2]. One of the great advantages of virtualization is to enable servers consolidation by allowing a host to host more than one independent virtual server, following the "one server per service" philosophy while reducing the waste of computing resources, as well as the costs of implementation and maintenance of the infrastructure [3].

There are situations where virtual machines need to be reallocated, such as in hardware maintenance cases. For this, a technique called virtual machines migration is used. Two types of migration can be performed and are commonly referenced in the literature, namely, Stop-and-Copy and Live Migration. This last approach allows the Virtual Machines (VM) to be migrated from the source host to the destination without interruption of the service execution, from the user perspective [4].

The live migration can be further subdivided into: pre-copy and post-copy. In the pre-copy approach, while the virtual machine is kept running at the source, all the memory pages that are at the source are transferred to the destination host. If there is a modification of some page during the transfer process, then it needs to be re-copied [5]. Hypervisor Xen works using this approach. In post-copy, initially the virtual machine is suspended at the source. Subsequently, information regarding its minimum state is transferred to the destination host. Then, the process of transferring the pages of memories is started. If a page that has not been transferred is requested from the destination host, a network fault will be generated, and this fault will be forwarded to the source that responds by transferring the requested page [7].

Therefore, realizing that the correct management and planning of the entire physical and virtual infrastructure of the computational environments influence the performance of the different running applications, and in view of the insufficient work that takes into account the occurrence of simultaneous migrations. This article aims to conduct an experimental study on migration of virtual machines in real time. Will be analyzed the influence of this in the performance of the Web and Database (DB) services from the perspective of the client of the application. The following scenarios were considered: Without Migration, With Migration and Simultaneous Migrations.

The paper is organized as follows: Section 2 presents the related works. In Section 3, we expose the methodology used to perform the experiments by describing the configuration of the virtualized environment implemented, in addition to the materials, scenarios and metrics used. Data analysis is in Section 4. Then, in Section 5, we conclude this paper and discuss futures works.

II. RELATED WORKS

Several works seek to optimize the location of Virtual Machines (VMs) among the physical resources available in datacenters [8][9]. These studies address the optimization of the available physical resources, aiming at saving resources, updating hardware, saving energy, among others.

 ISBN: 978-1-61208-598-2

In [10] and [11], it is possible to find the comparison of performances between different hypervisors. The most commonly used benchmarking strategy applies to a series of benchmark software that tests the most diverse system devices such as input / output, memory, network, processor, and so on. None of the cited works realizes performance evaluation by observing a specific application running in VM.

Alkmim *et al.* [12] considered the performance analysis of VM resources (memory, processing and file system - share VM image) before and during the migration process. Their results showed that the file system transfer rate was reduced by 55% during the migration period and its latency increased considerably.

In [4], a comparative evaluation was performed between stop and copy and pre-copy migrations using metrics such as total migration time, downtime, response time, and demand flow. Their results showed that stop-and-copy had five times more downtime than the pre-copy approach. However, the latter presented higher values in total migration time and response time. However, it did not show unavailability of the service, as in stop-and-copy.

Ye et al. in [13], carried a performance analysis using metrics such as downtime and total migration time, considering simultaneous migrations. It also evaluated resource reservation techniques for migration. However, no metrics were used to assess the migration impact from the user's perspective of the application.

Elsaid et al. in [14] developed an empirical model of single and multiple migration performance analyses to be used in estimating migration overhead. However, only factors related to CPU (Central Processing Unit) processing, transmission rate and dirty pages, for example, are considered and not the influence of the type of workload and the impact that the migration process can cause on the service offered in the migrated VM.

Bezerra et al. [15] conducted a preliminary statistical study evaluating the performance under the perspective of the user of the application, considering the scenarios With Migration and Without Migration to real and virtual environments. The results showed that there is a reduction of performance in the occurrence of migration, and that for virtual environments this reduction is more accentuated. However, scenarios with simultaneous migrations were not considered in this work.

III. METHODOLOGY

A. Xen

Xen is an open source virtual machine monitor (or hypervisor), which uses the para-virtualization concept by default [5]. In addition, it performs all the management, control and sharing of the resources of the hosts where the VMs will be dynamically allocated [3]. In addition to providing live migration support, it is modular, allows scalability, robustness and adequate security even for large and critical environments.

According to [16] and [17], Citrix, Microsoft and VMware, are the companies that have the best solutions in this market. Considering live migration, Xen is the most used solution [18]. Of the companies contributing to the Xen Project, we can mention: Alibaba / Aliyun [19], AWS [20], AMD [21], Citrix [22], Google [23], Intel [24], Oracle [25], Rackspace [26] and Verizon [27]. In this sense, it is possible to verify that Xen is commonly adopted, so it was the solution chosen for the experimentation environment implemented in this work.

B. Fully virtualized experimental environment

Figure 1 illustrates the fully virtualized environment configured on an actual machine of 16GB with RAM (Random Access Memory), Intel Dual Core 1066 MHz and 500GB of disk. Four VMs have been configured using the VMware Workstation Player: Xen1, Xen2, Client, and NFS (Network File System).

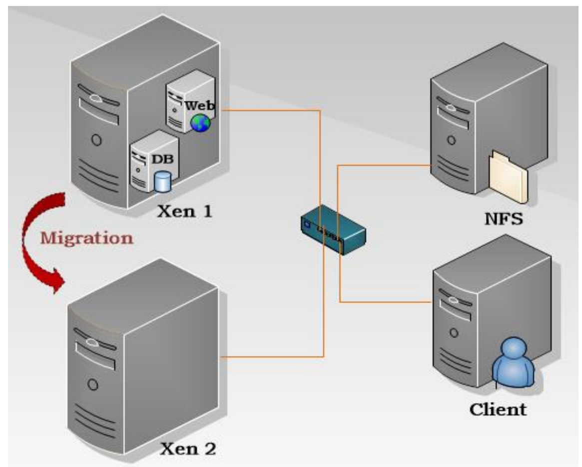

Figure 1. Fully virtualized experimental environment

Xen1 and Xen2 are the source and destination machines in the migration process. They are both virtualized with the Xen hypervisor. That way, Xen1 and Xen2 are responsible for hosting the Web Server and DB-Server machines, as well as their respective applications.

The Web-Server and DB-Server VMs are the instances that will be migrated and host the Apache Web HTTP services, as well as the Apache Cassandra™ Database server, respectively.

The VM Client is the element responsible for generating the workload for the VMs Web Server and DB-Server, using the benchmarks ab- Apache Benchmark [28] and YCSB - Yahoo! Cloud Serving Benchmark [29], respectively.

The Network File System (NFS) is the VM responsible for sharing the virtual machines' images and virtual disks involved in the migration process. The existence of this element is one of the requirements of the Xen hypervisor.

Table I shows the hardware and operating system configurations of all machines that make up the fully virtualized experimentation environment. One caveat to be made is that all network traffic is established through a LAN (Local Area Network) Fast Ethernet network.

ISBN: 978-1-61208-598-2

TABLE I. MEMORY, DISK, AND SOFTWARE CONFIGURATIONS OF THE VMS CREATED IN THE DEPLOYED ENVIRONMENT

Virtual Machines	Memory	Disk	Operating System
Client	1 GB	25 GB	Ubuntu 14.04
Xen 1	6 GB	40 GB	Ubuntu 14.04
Xen 2	6 GB	40 GB	Ubuntu 14.04
DB-Server	1 GB	10 GB	Ubuntu Server 14.04
Web-Server	1 GB	10 GB	Ubuntu Server 14.04
Storage NFS	1 GB	50 GB	Ubuntu 14.04

Some contexts and metrics were considered and collected during the experiments in the environment implemented in this work.

C. *Scenarios and metrics*

Experiments were carried out in order to individually evaluate the Web and Database Services. For both cases, three scenarios were considered and thirty samples were collected for analysis in each of the experiments.

As specified in Section III B, the workload generated for both the Web service and the database service was performed through the Benchmarks ab and YCSB, respectively. For both, we simulated the existence of 10 Clients making requests simultaneously. For YCSB, the number of operations and records were set to 200000. For Apache, the number of requests was set to 1000000.

For both cases, three scenarios were considered:

- Without Migration: This scenario was implemented with the purpose of observing the influence of the workload generated by the Client to the VM-Server, without considering the migration occurrence.
- With Migration - No Load: In this scenario, the influence of the workload generated by the Client is observed during the VM-Server migration. The total migration time was analyzed in this scenario.
- With Migration: The VM-Server receives the Client's workload while it is being migrated.
- Simultaneous Migration: For this scenario, two VMs with the same service are migrated concurrently while receiving the Client workload.

Some metrics have been evaluated for the Web Service, considering all scenarios described above. These metrics are the following:

- Number of requests per second: Corresponds to the number of requests served by the server in one second.
- Transfer rate: It used to measure the ratio of the amount of data that is transferred in a second between the Client and the Web server
- Server Response Time (SRT): Corresponds to the time in milliseconds that the server takes to respond to a request.

For the database service, some other metrics were also considered. These metrics are the following:

- Throughput (operations per second): Matches the number of operations that are performed on the database in one second.
- Read Latency (nanoseconds): This metric is the average time between the request and response of a read operation.
- Update Latency (nanoseconds): Similar to Read Latency, this metric refers to the average time between the request and response of an update operation.

For both services, the Total Migration Time (TMT) was also evaluated. The TMT is the time between the start of the migration, the transfer of all memory pages (registers states, CPUs, network interfaces, etc.) until the moment the VM is executed at the destination. For each experiment, thirty repetitions were performed.

IV. RESULTS AND ANALYSIS

To analyze the influence of the migration on the Web service and the Database service, the scenarios: No Migration, With migration, and Simultaneous Migration were considered for all metrics. For the evaluation of the total migration time the scenario With Migration - Without Load, was also considered, as described in section III B.

Thirty samples were collected during each run of the experiments. From this, the measurements used to better represent the dataset were: the mean, the standard deviation and the median.

A. *Web service*

Tables II and III, respectively, show the Number of Requests per Second and the Transfer Rate, considering the scenarios Without Migration, With Migration and Simultaneous Migration.

TABLE II. NUMBER OF REQUESTS PER SECOND FOR SCENARIOS: WITHOUT MIGRATION, WITH MIGRATION AND SIMULTANEOUS MIGRATION

Number of requests per second	Without Migration	With Migration	Simultaneous Migration
Average	676,35	555,58	524,97
Medium	677,29	556,84	518,58
Standard deviation	9,66	6,21	33,5

Considering the Web service, as shown in Tables II and III, in relation to the Number of Requests per Second and the Transfer Rate, a performance reduction of 17.85% was observed in the scenario With Migration and 22.38% for Simultaneous Migrations, comparing both with the Without Migration scenario.

TABLE III. SCENARIO TRANSFER RATE: WITHOUT MIGRATION, WITH MIGRATION AND SIMULTANEOUS MIGRATION

Transfer Rate (Bytes / second)	Without Migration	With Migration	Simultaneous Migration
Average	539,63	443,28	418,85
Medium	540,38	444,28	413,75
Standard deviation	7,71	4,95	26,74

Comparing the scenarios With Migration and Simultaneous Migration with regard still to the Number of Requests per Second and the Transfer Rate, a small reduction was observed which indicates that equivalent performances may occur. Although the decrease shown is not so evident, it is possible to observe that the TMT for the simultaneous migration was considerably higher, as shown in Table IV. This suggests that although the values for the referred metrics did not show such a significant reduction, the migration time increased significantly between these two scenarios.

Table IV also displays the TTM for the case where the migrated VM receives no load from the Client. With this it is possible to observe that the presence of workload implies an increase in the total time of migration. This is because under these circumstances, the memory pages are constantly being modified and it is necessary to resend them to the destination (as is typical of the pre-copy migration), thus increasing the time for the migration process to end.

TABLE IV. TOTAL MIGRATION TIME FOR THE SCENARIOS: WITH MIGRATION-NO LOAD, WITH MIGRATION, AND SIMULTANEOUS MIGRATION

TMT	With Migration - No Load	With Migration	Simultaneous Migration
Average	104,43	267,88	652,09
Medium	106,78	267,62	649,63
Standard deviation	4,26	9,53	0,57

Table V shows the Server Response Time for each of the scenarios considered.

TABLE V. SERVER RESPONSE TIME FOR SCENARIOS: WITHOUT MIGRATION, WITH MIGRATION AND SIMULTANEOUS MIGRATION

Server Response Time	Without Migration	With Migration	Simultaneous Migration
Average	8,46	9	9,93
Medium	8	9	10
Standard deviation	0,51	0	0,36

It is possible to observe that the presence of migration resulted in an increase in the SRT, which increases in the case of of Simultaneous Migrations.

B. Database service

For the Database service, Throughput presented a reduction of 39.16% for the scenario With Migration and 45.42% when considering Simultaneous Migrations, as shown in Table VI.

TABLE VI. THROUGHPUT FOR THE SCENARIOS:WITHOUT MIGRATION, WITH MIGRATION AND SIMULTANEOUS MIGRATION

Throughput (operations / second)	Without Migration	With Migration	Simultaneous Migration
Average	277,12	168,59	151,27
Medium	275,98	167,62	159,56
Standard deviation	51,47	13,41	35,64

As shown in Tables VII and VIII, it was observed that there is an increase of Latency in the presence of migration.

TABLE VII. AVERAGE READ LATENCY FOR SCENARIOS: WITHOUT MIGRATION, WITH MIGRATION AND SIMULTANEOUS MIGRATION

Average Latency - READ (ns)	Without Migration	With Migration	Simultaneous Migration
Average	46,08	73,16	100,40
Medium	44,28	69,77	84,20
Standard deviation	13,38	11,69	43,77

Compared to the Without Migration scenario, the latency for read operations showed an increase of 37% for the scenario With Migration and 54.1% for the scenario with Simultaneous Migration. Latency in Update operations showed an increase of 33.04% and 28.41%, for the same scenarios, respectively.

TABLE VIII. LATENCY AVERAGE UPDATE FOR THE SCENARIOS: WITHOUT MIGRATION, WITH MIGRATION AND SIMULTANEOUS MIGRATION

Average latency - UPDATE (ns)	Without Migration	With Migration	Simultaneous Migration
Average	27,97	41,76	39,07
Medium	28,11	44,72	39,72
Standard deviation	2,22	9,98	6,47

We note that there was a reduction in the performance related to the metrics evaluated for Simultaneous Migrations, when compared to the scenario With Migration. However, the results indicate that equivalent performances can occur.

TABLE IX. TOTAL MIGRATION TIME FOR SCENARIOS: WITHOUT MIGRATION, WITH MIGRATION AND SIMULTANEOUS MIGRATION

TMT	Without Migration - No Load	With Migration	Simultaneous Migration
Average	104,43	352,18	699,06
Medium	106,78	7,18	44,14
Standard deviation	4,26	350,19	706,918

It was also observed that the total migration time increased significantly, growing approximately twice in the occurrence of simultaneous migrations, as shown in Table IX.

V. CONCLUSION

Migration is an important technique commonly used in cloud computing environments. This article presented a study carried out with the objective of evaluating the influence of the migration in the services offered by the migrated machines.

Through the experiments, it was possible to verify that the migration of virtual machines generates an impact on the performance of the services offered to the users. It was also observed that in scenarios with multiple migrations the impact generated was small in relation to the metrics analyzed compared to the scenario with a single migration.

In future works, we intended to evaluate other metrics, implement other configurations, as well as to consider and evaluate the results obtained in a concrete context of a real environment.

Finally, it is worth mentioning that in the experiments network failures or requests were not observed, which allows to conclude that, although there is a downtime corresponding to the period in which the VM is interrupted at the origin and put into execution at the destination, this time is so minimal that it does not entail a failure in the services offered, thus achieving the goal of online migration that does not interrupt, from the user's point of view, the services during the process.

REFERENCES

[1] M. Veras, Virtualização: Central Component of the Datacenter. Preface Marco Américo D. Antonio. Editora Brasport: Rio de Janeiro, Brasil, 2011.

[2] A. Cassimiri, Virtualization: Basic Principles and Applications. Em: Escola Regional de Alto Desempenho - ERAD, Caxias do Sul, 2009.

[3] S. Citrix, Inc. Xenserver Open Source Virtualization. XenServer. Available in: <http://xenserver.org/>. Access on 15 de Sep 2016.

[4] V. M. Deborah, M. S. José, G. G. Danielo, “Analysis of the Impact of Migration of Virtual Machines on Virtualized Computational Environment”. XXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, Campo Grande, MS, 2011.

[5] C. Clark, et al, “Live Migration of Virtual Machines”. in NSDI’05: 2nd Symposium on Networked Systems Design Implementation, 2005.

[6] Xen, Xen Project. Available in: <https://www.xenproject.org/>. Access on Oct 31 2017.

[7] M. G. Bezerra, Automated Resource Management System for Virtualized Environments. Federal University of Rio de Janeiro. Polytechnic School: Department of Electronics and Computation, Rio de Janeiro, 2013.

[8] S Chaisiri, B. S. Lee, D. Niyato, “Optimal virtual machine placement across multiple cloud providers”. Services Computing Conference, APSCC. IEEE Asia-Pacific, 2009.

[9] M. Tsugawa, P. Riteau, A. Matsunaga, J. Fortes, “User-level virtual networking mechanisms to support virtual machine migration over multiple clouds”. The 2nd IEEE International Workshop on Management of Emerging Networks and Services (IEEE MENS), 2010.

[10] L. Jack, et al, “Performance overhead among three hypervisors: An experimental study using hadoop benchmarks”. IEEE International Congress on BigData Congress, 2013.

[11] P. V. V. Reddy, “Performance Evaluation of Hypervisors in the Private Cloud based on System Information using SIGAR Framework and for System Workloads using Passmark”. International Journal of Advanced Science and Technology, Vol. 70, pp. 17-32, 2014.

[12] G. P. E. U. Alkmim, “A low cost solution for Virtual Machine Migration”. XXIX Congress of the Brazilian Computer Society. Bento Gonçalves, 2009.

[13] K. Ye, X. Jiang, D. Huang, “Live Migration of Multiple Virtual Machines with Resource Reservation in Cloud Computing Environments”. IEEE 4th International Conference on Cloud Computing, 2011.

[14] M. E. Elsaid, “Multiple Virtual Machines Live Migration Performance Modelling VMware vMotion based Study”. IEEE International Conference on Cloud Engineering, 2016.

[15] P. Bezerra, G. Martins, C. Rocha, R. Gomes, F. Cavalcante, “Evaluating VM Live Migration Overhead With Xen Hypervisor”. 15th International Conference WWW/Internet. Alemanha, 2016.

[16] P. L. Silva, Use of VMware ESX virtualization technology to improve the utilization of hardware resources and increase performance in datacenters. Monograph (Technologist in computer science to business management) - the East Zone Technology College, São Paulo, 2009.

[17] N. Ruest, Virtualization, a begginer´s guide. New York: McGraw Hill Publications. New York, 2009.

[18] P. Padalla, X. Zhu, Z. Wang, S. Singhal, K.G. Shin, Performance evaluation of virtualization technologies for server consolidation. HP Labs Tec. Report, 2007.

[19] Alibaba, Alibaba Cloud. Available in: <https://www.alibabacloud.com>. Access on Oct 31, 2017.

[20] AWS, Amazon Web Services Cloud. Available in: <aws.amazon.com/>. Access on Oct 31, 2017.

[21] AMD, Advanced Micro Devices. Available in: <www.amd.com/pt/home>. Access on Oct 31, 2017.

[22] Citrix, Citrix Systems. Available in: <https://www.citrix.com.br/>. Access on Oct 31, 2017.

[23] Google. Available in: <https://www.google.com>. Access on Oct 31, 2017.

[24] Intel Corporation. Available in:<https://www.intel.com>. Access on Oct 31, 2017.

[25] Oracle. Available in:<https://www.oracle.com/>. Access on Oct 31, 2017.

[26] Rackspace. Available in:< https://www.rackspace.com/>. Access on Oct 31, 2017.

[27] Verizon. Available in:< https://www.verizon.com/>. Access on Oct 31, 2017.

[28] Apache, The Apache Software Foundation. Available in: <https://httpd.apache.org/docs/2.4/programs/ab.html>. Access on Sep 15, 2016.

[29] YCSB, Yahoo! Cloud Serving Benchmark. Available in: <https://github.com/brianfrankcooper/YCSB/wiki>. Access on Sep 15, 2016.

 ISBN: 978-1-61208-598-2