A Cloud Cost Optimization Synopsis: Moving Virtual Machines to Optimize Profit in Federated Clouds

Ms. Ayesha Tariq

MBA Business Analytics, United States University, Class of 2021,Cloud FinOps Systems Lead at Applied Materials, Inc

Santa Clara, CA, United States

ABSTRACT:

In a growing world of AI and automation processes where everything is sought out to be measured in either time, distance, effort, cost or any kind of saving in general, Cloud Cost optimization has become an epitome of the modern BI world. It has become absolutely crucial to utilize the workings of BI and ensure companies are achieving the maximum output using the bare minimum in effort to maximize profits. This not only comes from the need of winning the race for higher profits but also to showcase how effective their mechanisms and infrastructure is compared to others in the industry. There are multiple vendors at play when it comes to SAAS, PAAS and IAAS companies, mainly AWS, Azure, and Google Cloud. In this article we look at VMs and one of the ways to optimize cost, which is one of the highest or main purchases of the company. Cutting these costs means more money for the company - lower the cost, higher the profit, but with the same level of service. Of course, customers come first after all.

Cloud providers must figure out how to efficiently handle the hardware assets and VMs, or virtual machines, in a data center to guarantee a high level of service quality while also lowering costs when consumers flood the facility. For most cloud administration operations, virtual machine migration is the foundational technology. When the underlying hardware of a virtual machine is released, both cloud providers and consumers profit greatly from this functionality. Its cutting edge is a major area of emphasis for many academics. First, we provide an overview of virtual machine migration and go over its advantages and disadvantages in this article. Three factors are used to categorize virtual machine migration

schemes: Method, Distance, and Granularity. The research on non-live migration is only summarized, whereas the studies on live migration are thoroughly examined in light of its three primary obstacles: network connection continuity, data storage migration, and memory data migration. Further work is done on the quantitative examination of the performance of VM migrations. In some contexts (such as fog computing), user mobility has emerged as a significant driving force behind live virtual machine migration as cloud computing has developed and evolved. This also includes a summary of the research on the relationship between VM migration and user mobility. Finally, we provide a list of outstanding problems that still need work or more live VM migration improvements.

INTRODUCTION

Other than the usual BI tools like SQL, Tableau tables, Power BI, etc., there is a new way to optimize VMs. By adding a layer (such as a virtual machine manager, or hypervisor) on top of the operating system or hardware resources, virtualization technology allows a physical server to be divided into many separate execution environments (OS). Without causing any disruption to one another, each execution environment, or virtual machine (VM), operates separately with an operating system and apps. Because of a number of factors, virtualization technology were not often used at first. The CPU and memory, for instance, would be partially used by it. It was also difficult for suppliers to rent out its inactive physical servers to customers due to inadequate network bandwidth. Built on the basis of virtualization technology, cloud computing is a new service model that arises when linked technologies progress, including the use of Fibre Channel (FC), hardware performance improvements, security technology development, etc.

Large corporations may rent out their excess hardware resources piecemeal to consumers in the cloud computing space, paying for what they use. This allows users to begin working on virtual machines (VMs) immediately without having to pay for and maintain physical hardware. The ability to effectively manage virtual machines (VMs) in a data center is becoming more important as more customers choose to host their applications in cloud data centers. For instance, a server failure will affect all of the virtual machines running on it; some servers could be idle while others are overloaded; etc. With the arrival of a crucial technology VM

migration, all these issues (such as how to divide work across servers equitably and safeguard VMs against hardware failure) are resolved. The source of VM migration is process migration. Nevertheless, the issue of residential reliance with process movement makes it unsuitable for cloud administration. Once a virtual machine migrates, it becomes unfixed on the host where it was originally established. A virtual machine (VM) may be relocated between a single server to a different one, even across data centers. Servers consolidation, zero-downtime hardware repair, energy management, and traffic control are among the cloud management tasks that are facilitated by virtual machine migration. Moving virtual machines is not only a beneficial process - it additionally adds overhead to each of the roles involved: the co-located virtual machines on these two hosts, the migrated virtual machine, the source host, and the destination host. As a result, cloud management requires careful use of these VM migrations. Over the last several years, much research has been presented to improve its effectiveness and reduce its negative effects. A few earlier publications have previously attempted to compile the advancements made in the field of virtual machine migration, like Xu et al. do not concentrate on migration methodologies, but rather on research concerning VM performance overhead. They consider virtual machine migration to be one of the factors affecting virtual machine performance since it affects not only the moved virtual machine but additionally the others on the source and destination hosts.

Medina and Garcria study project-based VM migration procedures, however they do not identify shared technology. Furthermore, a thorough examination and comparison of several processes is still lacking. After creating a thorough taxonomy for virtual machine migration plans, Ahmad et al. examine migration methods from the perspectives of pre-, post-, hybrid-, and non-live migration. They did, however, only highlight a few findings on the current hot themes of multiple migration and VM migration between data centers. The technologies of catastrophe recovery and live migration for long-distance networks are summarized by Kokkinos et al. They both have to send large volumes of data across sluggish networks. Recovery from a disaster is a continuous process, while live migration involves one-time replication. However, they only examine works that address moving virtual machines across servers from the standpoint of optimizing network

performance. A few further surveys on moving virtual machines are either a straightforward elaboration or addressed by the works already listed. The technologies related to virtual machine migration are reviewed in detail in this paper, covering both non-live and live migration, LAN and WAN migration, single and multiple migration, user mobility-induced migration, and Network Function Virtualization, or NFV, Instance migration, among other scenarios. In addition to covering the entirety of the surveys that have already been listed, our paper also includes new developments in the field of live virtual machine migration (such as relocation technologies in the context of wireless edge computing), evaluations of various migration processes, and a discussion of research topics that are still pending.

Here is a summary of the contributions made by this paper:

1) The three issues that LIVE virtual machine migration in traditional cloud computing faces—memory migration of data, storage data migration, and network connection continuity—as well as the ways in which various technologies might work together to solve them, are covered.

2) A fresh taxonomy on virtual machine migration is created. Migrating distance (LAN and WAN), migration resolution (single and many), and migration mode (non-live and live) are the three categories used to classify migration strategies.

3) Based on their intended difficulties, the technologies for enhancing the performance of VM migrations are categorized and examined.

4) A number of variables affect the performance of migrations, and VM migrations are associated with overhead. Also presented are the studies that examine the VM migration process in order to determine the correlation between the migration performance and the influencing elements.

5) The development of virtualization technologies (like NFV) and cloud computing (like fog computing) has created additional obstacles for live virtual machine migration. An overview is provided of the unique migratory processes in these places.

6) The metrics taken from the literature are compared and summarized for each migration technology that is studied in this article.

7) This section concludes with a discussion of the unresolved problems with the VM migration that still need optimizations or fixes.

RELATED WORK

"Federation cloud resource management: Analysis and conversation" ,

Hamid, S. H. A. M. Tauseef, A. Gani, V. Chang, U. Shoaib, M. Liaqat, et al., 2017.

The Internet of Things, big data applications, cloud computing, and mobile cloud computing are all headed toward a federation of clouds. Increases in service quality, cost-benefits, and dependability are anticipated with the use of federated resources. A critical problem in cloud federations is resource management because of insufficient administrative rules, security, trust, and cross-domain expertise. This research divides these resource management tasks into six categories for the federated cloud environment: disaster management, resource pricing, resource allocation, monitoring, discovery, and selection. Following a discussion of each federated management of resources function, updated study findings are presented. Additionally, these resource management tasks are contrasted according to performance measures appropriate for each function. Lastly, for every resource management function that is categorized, we list the unresolved research difficulties.

"A study on virtual machines migration: Innovative techniques as well as open issues",

F. Zhang, G. Liu, X. Fu, and R. Yahya pour, 2018.

In order to maintain good service quality and save costs, cloud providers must figure out how to efficiently handle the hardware assets and virtual machines, or VMs, in a data centre as users pour in. For most cloud administration activities, virtual machine migration serves as a fundamental technological component. It releases a virtual machine from its supporting hardware. Both consumers and cloud providers gain a great deal from this functionality. On extending its cutting edge, a lot of researchers are concentrating. Initially, we provide a summary of virtual machine migration in this article and talk about its advantages and disadvantages. There are three ways to categorize VM migration plans: Method; 2) Range; and 3) Preciseness. After a brief assessment of the non-live migration research, the live migration studies are thoroughly assessed according to the three primary issues they face: migration of data from memory; migration of data from storage; and continuity of the network connection. Furthermore expanded upon are the studies on the quantitative examination of

virtual machine migration performance. User mobility is becoming a major driving force behind live virtual machine migration in various situations (such as fog computing) as cloud computing develops and evolves. Accordingly, the research on the relationship between virtual machine migration and user mobility is also summed up. We conclude by listing the unresolved problems that still need work or more live virtual machine migration improvements.

"Controlling performance excess of machine virtualization in cloud computing environments: an evaluation of state of current practice and upcoming directions" ,

Xu, X, Liu, H, Jinn, and Vasilios, A. V. (2014).

Tenants may construct virtual machines (VMs) on demand and pay for the leased computing resources solely based on the amount of time they use thanks to Infrastructure-as-a-Service (IaaS) cloud computing, which is scalable and cost-effective. Unpredictable performance of virtual machines (VMs) is a prominent drawback of the current computing paradigm for tenants, since shared computing power in datacentres are contested by VMs. This is one of the main problems with IaaS cloud computing. Thus, ensuring tenant satisfaction with virtual machine performance has been the focus of increased efforts lately. We examine and summarize the most recent studies on controlling the performance costs of virtual machines (VMs) in a variety of IaaS cloud situations, including single-server virtualization, single mega datacentres, and several geographically dispersed datacentres. To be more precise, we compare the efficacy and implementation complexity of the various overhead mitigation strategies, address performance modelling techniques with special attention to their cost and accuracy, and expose the reasons behind virtual machine performance overhead through representative scenario illustrations. With the knowledge of the benefits and drawbacks of every available option, we also outline research questions for the future related to modelling approaches and ways to reduce virtual machine performance overhead in IaaS cloud computing.

"Collaboration agents for decentralized load management within data centres centres using real-time migration from virtual machines" ,

A. Ramirez-Navarrete and J. O. Gutierrez-Garcia, 2015.

Virtual machines (VMs) have variable resource use profiles, load patterns fluctuate, energy consumption is a factor, and host hardware variety and different user needs are factors that must be considered in cloud data centre load management. For load administration in data centres with VM live migration support, this paper suggests distributed issue resolution approaches. To reduce energy consumption expenses and balance and consolidate disparate loads in a distributed way, collaborative agents are equipped with both an environment-aware consolidation protocol and a load balancing protocol. A collection of heuristics for choosing which virtual machines (VMs) to migrate, an assortment of host selection criteria for choosing where to relocate the VMs, and rules for turning on and off hosts are all given to agents. Additionally, a unique load balancing strategy is proposed in this study that moves the virtual machines (VMs) that are most resource imbalance-causing from overburdened hosts to underused hosts, where the VMs' hosting most significantly reduces resource imbalances. Data centre balancing, heterogeneous load consolidation, and energy-aware server consolidation are all accomplished efficiently and effectively by agents using distributed problem-solving approaches, according to empirical findings.

METHODOLOGY

1. Cloud Server 1: A thread-based server that takes user input and stores it in virtual machine memory... Using VM, this server may move from VM 1 to VM 2 and authenticate users.
2. Cloud Server 2: A thread-based server that stores user information in virtual machine memory, it will receive user details. To authenticate a user and switch between VM 2 to VM1, this server may call VM.
3. Migrating APP: this is an online application that allows users to register, submit information to the virtual machine of their choice, log in, and then transfer their registration information from one virtual machine to another.

RESULT AND DISCUSSION
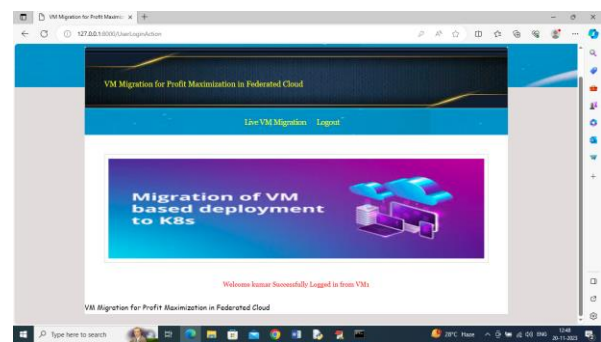
The user has successfully signed in from VM 1 in the screen above. From there, they may click on "Live VM Migration" to move between VM 1 to VM 2, as seen in the screen below.
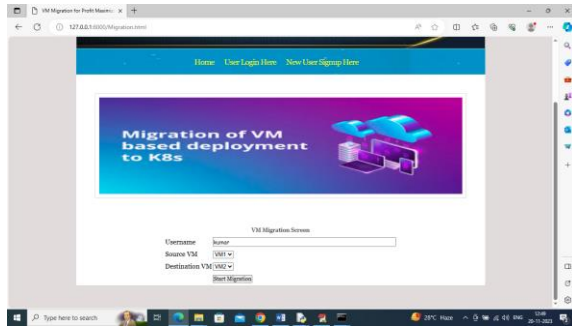
After choosing the desired virtual machine (VM) on the above screen where the user is migrating to, click the "Start Migration" button to see the page below. on the above screen, the username associated with the source virtual machine will automatically be chosen.

CONCLUSION

Users and cloud providers alike may gain a great deal from virtual machine migration. It provides the foundation for most cloud administration activities, including hardware upkeep and load balancing. The vendor lock-in problem is resolved and virtual machine mobility is further enhanced by the data centre transfer. Since this notion was put forward, several academics have worked to improve the speed of VM migration. The first section of this article provides a thorough introduction to virtual machine migration, including its benefits for cloud development and maintenance, as well as the primary difficulties and performance standards that should be considered when assessing a

migration plan. We discuss the structural theory of this study based on our proposed categorization of migration strategies. We first summarize the research on migration of non-live virtual machines. After that, we go into the tactics and optimization techniques for live virtual machine migration from the standpoint of the three difficulties it encounters: network connection continuity, data from storage migration, and memory data migration. Additionally, we summarize the research that aims to comprehend the connection between impact components and migration performances. These publications provide as useful guidelines for developing the best migration plans and optimization technologies. Additionally examined are the studies on VM migration brought on by user mobility. In this research, every evaluated technology is compared using metrics that were taken from the literature. The open research questions around virtual machine migration are finally covered.

Furthermore as we summarize yet another way to optimize cloud cost among many others, some common ways to make the process easier to do this is to use BI tools like Apptio Cloudability, VMware Cloudhealth, and others that capture back-end tech information and are able to deep dive into

areas which can be neglected by human error and incapability, but is caught by machine learning that results in countless cost reductions. For example, by performing the right analysis and making use of dashboards and reports in these tools (many that are built automatically or come built-in), a reduction in the number of Kafka brokers has shown to massively cut down cost in AWS Kafka servers. For Azure VMs even something as simple as moving the infra to a different location can save cost tremendously. Similarly , for more finance related people, software tools like Datarails, Vena, Abacum have proven to capture great insights in the cost environment of a company where managers and their analysts are able to identify areas of wastage and cost control. The use of RIs, SPs, and multiple other techniques has been used traditionally as well, but deeper insights that go unnoticed are often ignored or left out due to the simple fact of not being able to capture them when necessary. Cloud Cost Management remains one of the most important functionalities in these companies, and this proves FinOps continues to remain a hot topic as the world moves further into AI and BI.

REFERENCES

Federated cloud resource management: Review and debate, Journal of Network and Computer Applications, vol. 77, no. Supplement C, pp. 87-105, 2017. M. Liaqat, V. Chang, A. Gain, S. H. A. Hamid, M. Tauseef, U. Shoaib, et al. The paper "A survey on virtual machine migration: Challenges techniques and open issues" was published in IEEE Communications Surveys Tutorials in 2018. It was authored by F. Zhang, G. Liu, X. Fu, and R. Yahya pour.

[3] "Managing performance overhead of virtual machines in cloud computing: A survey state of the art and future directions," by F. Xu, F. Liu, H. Jinn, and A. V. Vasilios, Proceedings of the IEEE, vol. 102, no. 1, pp. 11–31, Jan 2015.

[4] J. O. Gutierrez-Garcia and A. Ramirez-Navarrete, "Live migration of virtual machines as a collaborative agent for distributed load management in cloud data canters," IEEE Transactions on Services Computing, vol. 8, no. 6, pp. 916-929, 2015.

[5] "Energy-aware virtual machine scheduling on data centres with heterogeneous bandwidths," by D. G. Lago, E. R. M. Madeira, and D. Mehdi, IEEE Transactions on Parallel and Distributed Systems, vol. 29, no. 1, pp. 83-98, Jan 2018.

[6]M. Abu-Tari, M. I. Biswas, P. Morrow, S.

McClean, B. Scoter, and G. Parr, "Quality of service scheme for intra/inter-data centre communications," Proceedings of the IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), pp. 850-856, 2017.

[7]M. Jakobe, A. Celeste, M. Fazio, M. Villani, and A. Polianite, "A method to lower energy expenses via virtual machine migrations in cloud federation," 2015 IEEE Symposium on Computers and Communication (ISCC), pp. 782-787, July 2015.

In Proc. CLOSER, pp. 563-573, 2011, N. Huber, M. von Quast, M. Hauck, and S. Konev, "Evaluating and modelling virtualization performance overhead for cloud environments."

[9]C. Devender and colleagues, "Optical networks for grid and cloud computing applications," Proc. IEEE, vol. 100, no. 5, May 2012, pp. 1149–1167.

[10]L. Wang et al., "Early understanding and experience with scientific cloud computing," Proc. IEEE 10th Int. Conf. High Perform. Compute. Common. (HPCC), pp. 825-830, 2008.

[11] A. Weiss, "Computing in the clouds," Computing, vol. 16, 2007.

[12] Cheng, L., Tamavidins, I., Ketolase, S., and Antoniou, G., "Design and evaluation of small–large outer joins in cloud computing environments," Journal of Parallel and Distributed Computing, 2017.

[13] J. Wu, S. Guo, J. Li, and D. Zeng, "Greening big data: Big data meets green challenges," IEEE Systems Journal, vol. 10, no. 3, pp. 873–887, 2016.

[14] S. Osman, D. Sobriety, G. Us, and J. Nigh, "The design and implementation of zap: A system for migrating computing environments," ACM SIGOPS Operating Systems Review, vol. 36, no. SI, pp. 361–376, 2002.

[15] J. G. Hansen and A. K. Henriksen, "Nomadic operating systems," doctoral dissertation, Cite seer, 2002.

[16] Padilla, P., X. Zhu, Z. Wang, S. Singhal, K. G. Shin, et al., "Performance evaluation of virtualization technologies for server consolidation," HP Labs Technical Report, 2007.

[17] L. Cheng and T. Li, "Efficient redistribution of data to accelerate big data analytics in large systems," in High Performance Computing (Hip), IEEE 23rd International Conference on, IEEE, 2016, pp. 91–100.

[18] R. Balanchine and R. Ramon, "Power and energy management for server systems," Computer, vol. 37, no. 11, 2004, pp. 68–76.

[19] J. Wu, "Green wireless communications:

from concept to reality [industry perspectives]," IEEE Wireless Communications, vol. 19, no. 4, 2012.

[20] Joint virtual machine place mint and routing for data centre traffic engineering, J. W. Jiang, T. Lan, S. Ha, M. Chen, and M. Chiang, in INFOCOM, 2012 Proceedings IEEE. IEEE, 2012, pp. 2876–2880.

[21] F. Xu, F. Liu, H. Jinn, and A. V. Vasilios, "Managing performance overhead of virtual machines in cloud computing: A survey, state of the art, and future directions," IEEE Proceedings, vol. 102, no. 1, pp. 11–31, 2014.