International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT2390554

A Study on Virtual Machine Placement, its Parameters and Challenges

Sharanayya B Hiremath

Lecturer at SVS College of BBA and BCA, Ilkal, Karnataka, India

ARTICLEINFO

Article History:

Accepted: 02 Dec 2023

Published: 19 Dec 2023

Publication Issue

Volume 9, Issue 6

November-December-2023

Page Number

247-254

ABSTRACT

In real-world scenarios, cloud computing data centres house hundreds of thousands of virtual machines (VMs).Computing resources are provisioned as metered on-demand services across networks, and may be promptly allocated and released with low administration effort, thanks to the rise of cloud computing. The virtual machine is one of the most often used resource carriers in the cloud computing paradigm for encapsulating business services. In this regard, Virtual Machine Placement (VMP) is one of the most difficult problems in cloud infrastructure management, given the enormous number of alternative optimization criteria and differences in cloud infrastructure management.

Keywords : Data centre, VM placement, Markov model, Queuing theory, ARIMA

I. INTRODUCTION

Cloud computing is rapidly gaining traction as a critical technology for hosting a variety of IT services for businesses, including on-demand virtual resources based on a pay-per-use model [1, 2]. Large-scale data centres (DCs) with several servers or physical machines (PMs) are used by cloud service providers (CSPs) [3, 4]. In cloud DCs, virtualization is used to provide clients with virtual machines (VMs) that are contained by a software layer known as VMM or VM Monitor [5]. The VMM simplifies the management of PMs' shared resources and increases the security of VMs [6].However, hosting multiple VMs in a single PM is a difficult problem [7]. For example, due to over- and under-utilization issues in the PMs, application performance may suffer or high-cost resources may be squandered. As a result, resource management in cloud DCs is a difficult issue that affects both CSPs and their customers [8]. As PMs or servers consume nearly 26 percent of the power consumed by cloud DCs. Proper VM placement and dynamic management can significantly reduce DC power consumption, improve throughput, and increase CSP profit while preventing SLA Violation (SLAV) [10]. However, the VM placement process employs costly VM migration operations, and incorrect VM placement may result in a slew of VM migration processes, degrading the cloud DC's performance. Various VM placement solutions for various cloud computing environments have been

proposed in the literature, which can be divided into reactive and proactive/predictive schemes.

Reactive VM placement can be performed in response to overutilization or underutilization events, and it only considers the current status of the DC [11]. However, predictive VM placement frameworks use historical resource usage data to predict the future state of PMs and make better placement decisions [12, 13]. Predictive VM placements primarily aim to reduce the number of PMs, VM migrations, and DC network traffic while maintaining guaranteed QoS [14]. The predictive VM placement schemes use a variety of prediction algorithms and techniques [15–18].

Virtual Machine Placement refers to the process of determining which virtual machines (VMs) should be located (i.e. executed) on each physical machine (PM) in a data centre (VMP). The VMP problem has been extensively researched in the cloud computing literature, with several surveys already presented. Existing surveys concentrate on specific issues such as: (1) the use of energy-efficient techniques to solve the problem [2] [3] (2) specific architectures in which the VMP problem is used, specifically federated clouds [4], and (3) methods for

comparing the performance of placement algorithms in large on-demand clouds [5]. There are numerous parameters and considerations (for example, performance, cost, and location). Involved in the decision of where and when to place and reallocate data objects and computation resources in cloud environments. Some of the considerations are consistent with one another while others may be contradicting. At the same time, we are witnessing an increasing trend towards hosting soft real-time applications, such as airline reservation systems, virtual reality applications, Netflix video streaming and Coursera online digital learning, on the cloud.

These applications necessitate more stringent performance specifications, such as being sensitive to latency and response times. Because multiple collocated VMs caused by resource overbooking can cause significant performance interference [6][7][8][9] for applications hosted on their respective VMs, cloud providers' use of resource overbooking may have a negative impact on their performance.

Although prior work on performance isolation [9] among VMs collocated on an overbooked host machine exists, it remains a difficult task to shield the VMs from its neighbours due to the nature of resource sharing, resource overbooking practises used, and the fluctuating workload characteristics in the cloud. As a result, an application running on one VM may have an effect on the performance of another application running on a different VM on the same host machine. Network-intensive and compute-intensive applications, in particular, may be severely impacted.

Because performance interference is caused by how one VM interacts with another collocated VM, addressing performance interference issues resulting from resource overbooking and meeting the response time requirements of soft real-time applications will necessitate effective VM placement on host machines while carefully considering the actual workload characteristics of the VMs. Traditional and offline heuristics such as bin packing will not be relevant for interference-aware VM placement in cloud computing due to the changing dynamics of the workloads on the VMs and also because VMs frequently relocate from one physical machine to another for a variety of reasons.

II. CLOUD COMPUTING

Cloud computing is a paradigm shift in the way that current enterprise IT infrastructure is built, and it is a new paradigm in which computing is given as a service rather than a product, with shared resources, software, and information provided to customers as a utility across networks.

2.1. Hardware Virtualization

Virtualization is a technique that combines or divides computing resources to provide

one or more operating environments through the use of methodologies such as hardware and software partitioning, partial or total machine emulation, time-sharing, and others.Virtualization is a computing technique that decouples computational functions and implementations from physical hardware.

It is the foundation of cloud computing because it allows for separations between hardware and software, users, and processes and resources. Virtualization technologies find major applications in a variety of domains, including server consolidation, secure computing platforms, supporting multiple operating systems, kernel debugging and development, system migration, and so on, resulting in broad use. The majority of them exhibit comparable operating environments to the end user; nevertheless, they differ greatly in the layers of abstraction at
which they work and the underlying architecture.

There are three techniques to hardware virtualization: (i) full virtualization, (ii) partial virtualization, and (3) para virtualization.

Deployable services in cloud systems can be contained in virtual appliances (VAs) [30] and deployed by instantiating virtual machines with their virtual appliances [31]. We identified the following abstraction levels: instruction set level, hardware abstraction layer (HAL) level, operating system level, library level, and application level virtual machines.

By divorcing the hardware and operating system infrastructure supplier from the application stack provider, virtual appliances enable economies of scale on one side to be leveraged by economies of simplicity on the other.

2.2. XaaS Service Models

Commonly associated with cloud computing are the following service models:

i) Software as a Service (SaaS)

Software applications are offered as services that run on infrastructure managed by
the SaaS vendor in the SaaS model. Consumers can access services via a variety of clients, including web browsers and programming interfaces, and are often charged on a subscription basis. It is built on the idea of renting an application from a service provider rather than purchasing, installing, and operating software on one's own.

ii) Platform as a Service (PaaS)

Cloud providers supply a computer platform and/or solution stack in the PaaS
paradigm, which often includes an operating system, programming language execution environment, database, and web server [32].

Application developers can develop and run their software on a cloud platform without having to manage or control the underlying hardware and software layers, such as network, servers, operating systems, or storage, but they retain control over the deployed applications and possibly application-hosting environment configuration settings [33].Force.com, Microsoft Azure, and Google App Engine are a few examples.

iii) Infrastructure as a Service (IaaS)

Computing resources such as storage, network, and computation resources are
provisioned as services in the IaaS model. Consumers can deploy and run arbitrary software, such as operating systems and apps. Consumers do not manage or control the underlying cloud infrastructure, but must maintain their own virtual infrastructure, which is often made up of virtual machines hosted by the IaaS operator. Amazon EC2 and S3, Rack space, AT&T, and Verizon are a few examples.

III. Cloud computing scenarios

Two key stakeholders in a cloud provisioning scenario can be identified based on the
cloud services provided:

(i) Provider of Infrastructure (IP) (ii) Provider of Services (SP).

IP who provides infrastructure resources such as virtual machines, networks, storage, and so on, which are utilised by SP to deliver end-user services such as

SaaS to their consumers, with these services produced using PaaS technologies. As stated in [14], four major types of cloud scenarios have been identified:

i) Private cloud

And the organisation provides services through internal infrastructure, fulfilling the

functions of both SP and IP. Many of the security and privacy risks associated with hosted sensitive material in public clouds can be avoided in private clouds; the latt er is when the SP leases IaaS services with publicly visible IPs. Because the entire infrastructure can be managed inside the same domain, a private cloud provides greater assurances of control, monitoring, and performance.

ii) Cloud Bursting

Private clouds may offload capacity to other IPs during periods of excessive workload or for other reasons, such as planned server maintenance. As the providers form a hybrid architecture known as cloud bursting. Less sensitive processes are performed in the public cloud, whilst tasks needing higher degrees of security are performed on private infrastructure.

iii) Cloud Federated

Federated clouds are IPs that collaborate through collaborative load-sharing agreements to offload capacity to one another [15] in a manner similar to how power companies swap capacity. The federation occurs in a transparent manner at the IP level. In other words, if an SP delivers services to one of the IPs in a federation, the SP is not notified if the service is off-loaded to another IP in the federation. However, the SP can direct which IPs the service is supplied on, for example, by setting location constraints in the service manifest.

iv) Multi Cloud

In multi-cloud scenarios, the SP is in charge of dealing with the added complexity of coordinating the service across several external IPs, i.e., planning, initiating, and monitoring service execution.

IV. VIRTUAL MACHINE PLACEMENT

Given a set of admitted services and the availability of local and maybe remote resources, a number of placement problems must be addressed in order to identify where to keep data and execute VMs. The sections that follow discuss the challenges and state-of-the- art of VM placement and scheduling in cloud systems.

4.1. Parameters and Considerations

A plethora of characteristics and considerations go into deciding where and when to reallocate data objects and computations in cloud settings. An automated placement and scheduling method should analyse the trade-offs and allocate resources in a way that favours the stakeholder for whom it is designed (SP or IP). This frequently leads to the problem of maximising price or performance given a set of restrictions, which frequently includes the one of price and performance that is subject to optimization. Among the most important factors to consider are:

i) Performance:

To increase physical resource efficiency, data centres are increasingly using virtualization and consolidation to handle a large number of different applications operating concurrently on server platforms. The performance obtained with different virtual machine placement strategies can vary greatly [27].

ii) Cost:

In the early stages of cloud adoption, fixed prices dominated the pricing paradigm.

However, the cloud market trend indicates that the use of dynamic pricing strategies is increasing [36]. Investment reductions are feasible by dynamically shifting services among clouds or dynamically reconfiguring services (e.g., resizing VM sizes without affecting service performance). Internal costs for VM placement, such as interference and overhead caused

by one VM on other concurrently operating VMs on the same physical host, should also be considered.

iii) Locality: In general, for usability and accessibility reasons, VMs should be positioned close to users (which could be other services or VMs). However, due to legal considerations and security concerns, for example, location may constitute a constraint for optimal placement.

iv) Reliability and continuous availability:

Service reliability and availability are important goals for VM placement. To do this, virtual machines (VMs) may be placed/replicated/migrated across various (at least two) geographical zones. Factors like as the relevance of the data/service wrapped in VMs, its predicted usage frequency, and the stability of the various data centres must be considered during this operation.

4.2. Challenges

Given the range of relevant factors, the set of restrictions and objective functions of

potential interest, and the variety of deployment scenarios, there are a number of hurdles to developing broadly applicable placement methods, some of which are discussed below. For starters, there is no generic model to represent multiple scenarios of resource scheduling, particularly when customers' expectations are ambiguous and difficult to encode using modelling languages.

Second, model parameterization, or finding appropriate values for parameters in a given model, is a time-consuming operation when the problem is vast. For example, in a multi-cloud scenario with n cloud providers and m VMs, (m*n) assignments are required to describe the VM migration overheads while ignoring any VM size changes. As a result, techniques that can assist in automatically capturing certain variables are required.

Third, the VM placement problem is often described as a variation of the NP-hard class constrained multiple-knapsack problem [37]. As a result, trade-offs between solution quality and execution time must be considered. Given the size of real-world data centres, this is a critical issue. For example, Amazon EC2 [38], the largest cloud provider, has around 40,000 servers and schedules 80,000 VMs every day [39].

V. 5.0 CONCLUSION

One of the most critical concerns in cloud DCs is effective VM management, which has a direct impact on energy usage, cost, scalability, and other environmental issues such as CO2 emissions. VM placement is a critical step in VM management that attempts to place VMs in the most appropriate PMs by taking into account a variety of parameters. To make better VM placement selections and save their costs .It is preferable to forecast the future load of the VMs when migrating. The forecasting method leads in a reduction in reaction time, a number of VM migrations, and a reduction in costs. SLA infractions given the significance of this issue, numerous studies have been done in order to more efficiently manage VMs.

VI. REFERENCES

[1]. Masdari, M., Salehi, F., Jalali, M., Bidaki, M.: A survey of PSO-based scheduling algorithms in cloud computing. J. Netw. Syst. Manag. 25, 122 – 158 (2017)

[2]. Alizadeh, M., Abolfazli, S., Zamani, M., Baharun, S.,Sakurai, K.: Authentication in mobile cloud computing: a survey. J. Netw. Comput. Appl. 61, 59 –80 (2016)

[3]. Masdari, M., ValiKardan, S., Shahi, Z., Azar, S.I.: Towards workflow scheduling in cloud computing: a comprehensive analysis. J. Netw. Comput. Appl. 66, 64 –82 (2016)

[4]. Cheraghlou, M.N., Khadem-Zadeh, A., Haghparast, M.: A survey of fault tolerance architecture in cloud computing. J.Netw. Comput. Appl. 61, 81 –92 (2016)

[5]. Masdari, M., Nabavi, S.S., Ahmadi, V.: An overview of virtual machine placement schemes in cloud computing. J.Netw. Comput. Appl. 66, 106 –127 (2016)

[6]. Masdari, M., Jalali, M.: A survey and taxonomy of DoS attacks in cloud computing. Security and Communication Networks. 9, 3724 –3751 (2016)

[7]. Ahmad, R.W., Gani, A., Hamid, S.H.A., Shiraz, M., Yousafzai, A., Xia, F.: A survey on virtual machine migration and server consolidation frameworks for cloud data centers. J. Netw. Comput. Appl. 52, 11 –25 (2015)

[8]. Song, F., Huang, D., Zhou, H., Zhang, H., You, I.: Anoptimization-based scheme for efficient virtual machine placement. Int. J. Parallel Prog. 42, 853 –872 (2014)

[9]. Rong, H., Zhang, H., Xiao, S., Li, C., Hu, C.: Optimizing energy consumption for data centers. Renew. Sust. Energ. Rev. 58, 674 –691 (2016)

[10]. J. Xu and J. Fortes, "A multi-objective approach to virtual machine management in datacenters," in Proceedings of the 8th ACM international conference on Autonomic computing, 2011, pp. 225–234

[11]. Ding, Y., Qin, X., Liu, L., Wang, T.: Energy efficient scheduling of virtual machines in cloud with deadline constraint. Futur. Gener. Comput. Syst. 50, 62 –74 (2015)

[12]. S. Chaisiri, B.-S. Lee, and D. Niyato, "Optimal virtual machine placement across multiple cloud providers," in Services Computing Conference, 2009. APSCC 2009.IEEE Asia-Pacific, 2009, pp. 103–110

[13]. Weingärtner, R., Bräscher, G.B., Westphall, C.B.: Cloud resource management: a survey on forecasting and profiling models. J. Netw. Comput. Appl. 47, 99 –106 (2015)

[14]. Roh, H., Jung, C., Kim, K., Pack, S., Lee, W.: Joint flow and virtual machine placement in hybrid cloud data centers. J. Netw. Comput. Appl. 85, 4 –13 (2017)

[15]. Lin, W., Xu, S., Li, J., Xu, L., Peng, Z.: Design and theoretical analysis of virtual machine placement algorithm based on peak workload characteristics. Soft. Comput. 21, 1301–1314 (2017)

[16]. Addya, S.K., Turuk, A.K., Sahoo, B., Satpathy, A., Sarkar, M.: A game theoretic approach to estimate fair cost of VM placement in cloud data center. IEEE Syst. J. 1–10 (2017)

[17]. A. p. Xiong and C.-x. Xu, "Energy efficient multiresource allocation of virtual machine based on PSO in cloud data center," Mathematical Problems in Engineering, vol. 2014, 2014

[18]. J. L. L. Simarro, R. Moreno-Vozmediano, R. S. Montero, and I. M. Llorente, "Dynamic placement of virtual ma- chines for cost optimization in multi-cloud environments, " in High Performance Computing and Simulation (HPCS), 2011 International Conference on, 2011, pp. 1–7

[19]. S, K. (2016). Study of Virtual Machine Placement, its Parameters, . International Journal of Advanced Computer Science and Technology.

[20]. Zangakani, M. M. (2019). Green Cloud Computing Using Proactive Virtual Machine. Spring Nature.

[21]. E. Elmroth, J. Tordsson, F. Hern´andez, A. Ali-Eldin, P. Sv¨ard, M. Sedaghat, and W. Li. Self-management challenges for multi-cloud architectures. In W. Abramowicz, I. Llorente, M. Surridge, A. Zisman, and J. Vayssi`ere, editors, Towards a Service-Based Internet, volume 6994 of Lecture Notes in Computer Science, pages 38–49. Springer Berlin/Heidelberg, 2011.

[22]. A. Beloglazov, J. Abawajy, and R. Buyya, “Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing,” Future Generation Computer Systems, vol. 28, no.5, pp. 755–768, 2012

[23]. L. Salimian and F. Safi, "Survey of energy efficient data centers in cloud computing," in Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing. IEEE Computer Society, 2013, pp. 369–374.

[24]. M. Gahlawat and P. Sharma, "Survey of virtual machine placement in federated clouds," in Advance Computing Conference (IACC), 2014 IEEE International. IEEE, 2014, pp. 735–738.

[25]. K. Mills, J. Filliben, and C. Dabrowski, "Comparing vm-placement algorithms for on-demand clouds," in Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on. IEEE, 2011, pp. 91–98.

[26]. R. Nathuji, A. Kansal, and A. Ghaffarkhah, "Q-clouds: managing performance interference effects for qos-aware clouds," in Proceedings of the 5th European conference on Computer systems. ACM, 2010, pp. 237–250.

[27]. O. Tickoo, R. Iyer, R. Illikkal, and D. Newell, "Modeling virtual machine performance: challenges and approaches," ACM SIGMETRICS Performance Evaluation Review, vol. 37, no. 3, pp. 55–60, 2010.

[28]. X.Pu, L. Liu, Y. Mei, S. Sivathanu, Y. Koh, and C. Pu, "Understanding performance interference of i/o workload in virtualized cloud environments," in Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on. IEEE, 2010, pp. 51–58.

[29]. X. Zhang, E. Tune, R. Hagmann, R. Jnagal, V. Gokhale, and J. Wilkes, "Cpi2: Cpu performance isolation for shared compute clusters," in Proceedings of the 8th ACM European Conference on Computer Systems, ser. EuroSys '13. New York, NY, USA: ACM, 2013, pp. 379–391.

[30]. G. Kecskemeti, G. Terstyanszky, P. Kacsuk, and Z. Nem´eth. An Approach for Virtual Appliance Distribution for Service Deployment. Future Gener. Comput. Syst., 27(3):280–289, March 2011.

[31]. G. Kecskemeti, P. Kacsuk, T. Delaitre, and G. Terstyanszky. Virtual Appliances: A Way to Provide Automatic Service Deployment. In F.

[32]. Davoli, N. Meyer, R. Pugliese, and S. Zappatore, editors, Remote Instrumentation and Virtual

[33]. Laboratories, pages 67–77. Springer US, 2010.Platform as a Service. http://en.wikipedia.org/wiki/Platform_as_a_service, visited May, 2012.

[34]. P. Mell and T. Grance. The NIST definition of cloud computing. National Institute of Standards and Technology (NIST), 2011.

[35]. M. Ahronovitz et al. Cloud computing use cases white paper, v4.0. www.cloudusecases.org, visited May 2012.

[36]. B. Rochwerger, D. Breitgand, E. Levy, A. Galis, K. Nagin, I. Llorente, R. Montero, Y. Wolfsthal E. Elmroth, J. Caceres, M. Ben-Yehuda, W. Emmerich, and F. Galan. The RESERVOIR model and architecture for open federated cloud computing. IBM Journal of Research and Development, 53(4):1–11, 2009.

[37]. J. Lucas Simarro, R. Moreno-Vozmediano, R. Montero, and I. Llorente. Dynamic Placement of Virtual Machines for Cost Optimization in MultiCloud Environments. In Proceedings of the 2011

[38]. International Conference on High Performance Computing and Simulation (HPCS), pages 1 –7, july 2011.

[39]. T. Chunqiang, S. Malgorzata, S. Michael, and P. Giovanni. A scalable application placement controller for enterprise data centers. In Proceedings of the 16th international conference on World Wide Web, WWW'07, pages 331–340. ACM, 2007.

[40]. Amazon Elastic Compute Cloud. http://aws.amazon.com/ec2/, visited May, 2012.

[41]. D. Erickson, B. Heller, S. Yang, J. Chu, J. D. Ellithorpe, S. Whyte, S. Stuart, N. McKeown, G.

M. Parulkar, and M. Rosenblum. Optimizing a virtualized data center. In Proceedings of the 2011 ACM SIGCOMM Conference (SIGCOMM'11), pages 478–479, 2011.

[42]. Gurobi Optimization, http://www.gurobi.com, visited October 2011.

Cite this article as :

Sharanayya B Hiremath , "A Study on Virtual Machine Placement, its Parameters and Challenges", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 6, pp.247-254, November-December-2023.
Available at doi :
https://doi.org/10.32628/CSEIT2390554
Journal URL : https://ijsrcseit.com/CSEIT2390554