A SURVEY ON ENERGY EFFICIENT VM CONSOLIDATION TECHNIQUES IN A CLOUD DATA CENTER

Dr S Sathyanarayana,
Guest professor
Srikantha First Grade Womens College Mysore

Abstract

A large number of services are run on dedicated physical servers in traditional data centers. The majority of the time, these data centers are not fully utilized in terms of resources. Virtualization allows for the migration of virtual machines from one host to another, allowing these data centers to consolidate their services onto fewer physical servers than were originally required. The VM should be installed on the appropriate host to maximize the use of physical resources. Various researchers have developed a number of virtual machine consolidation strategies that work in a cloud computing environment. The majority of VM consolidation methods strive to achieve a certain aim.This purpose could be to save energy by turning off some servers or to maximize resource use. In this research, we looked at various VM consolidation techniques and their flaws.

Keyword: virtual machine, data centers, virtualization, migration, power saving

I. Introduction

Many businesses have quickly adopted the extremely flexible IaaS (Infrastructure as a Service) cloud computing paradigm as a way to reduce their Total Cost of Ownership (TCO) while maximizing resource and money use. Today, many firms, such as Amazon's EC2 (Elastic Compute Cloud), are focusing on greening their data centers, which means reducing the amount of active servers. This necessitates the use of appropriate solutions such as virtualization, which serves as the backbone of cloud computing technologies. It allows computer hardware to be shared by splitting computing resources. The Hypervisor, often known as the Virtual Machine Monitor (VMM), is a small software programme that maintains and controls all VM-related actions. Load balancing and consolidation are made easier with live virtual machine migration.Many services in a datacenter only use a small part of the total resources available. This can result in a situation where numerous virtualized servers function and consume more space and resources than intended, despite the fact that their workload does not justify it. Server sprawl is the term for this issue. To avoid resource waste, numerous virtual machines are crammed onto the fewest available physical servers, while the remainder of the servers are put into sleep mode (low-power state). Server consolidation is a technique for preventing server sprawl.Data centres gain from the reduction in server count by enhancing system availability, lowering infrastructure complexity, and, of course, saving energy and money. Although VM consolidation can fulfil its purpose of increasing the number of suspended servers, there may be some issues in an IAAS environment that affect cloud energy efficiency.

- The first issue is the trade-off between running VMs' performance, energy consumption, and resource use. Because they must compete for and share the resources offered by their corresponding PM. Server suspensions and performance degradations can increase execution time, reducing energy savings. The resources being PMs—CPU, main memory, and I/O capabilities, might cause server suspensions and performance degradations.

- Another issue to consider is the overhead created by continuous live movement of VMs during consolidation, which is a result of constantly changing resource demands. When a VM is moved, the condition of its CPU, main memory, storage, and network connections are all taken into account.
- The prediction of the server's energy consumption is the next point of concern. This is a difficult task that necessitates energy-performance profiling approaches, which are unacceptably expensive due to their associated overheads.
- Cloud data centre resources (e.g. CPU, memory, network bandwidth, and storage) in particular must be assigned with equal emphasis on reducing energy consumption as on meeting the Quality of Service (QoS) standards stated by users via Service Level Agreements (SLAs).
- The final point to consider is that achieving effective server consolidation is a difficult NP-Hard challenge.

Many efficient ways to server consolidation have been presented to date, keeping these hard features in mind. These cutting-edge methods handle difficulties such as physical resource heterogeneity and the dynamic nature of virtual machines and workloads. We cover VM scheduling techniques that try to consolidate servers in this work, with an emphasis on the VM placement algorithms utilised by various scheduling approaches. The virtual machine scheduling technique helps to consolidate a data centre by reducing the number of active physical machines. Dynamic server consolidation, as previously mentioned, can enhance energy efficiency by maximising the use of existing resources.

Rest sections of the paper are as follows. In section II, Literature review have been discussed and summary of literature review.III concludes the paper.

II. Literature review

The author [1] suggested a load balancing technique called honey bee behavior inspired load balancing (HBB-LB) that tries to achieve a well-balanced load among virtual machines in order to maximize throughput. Not only does the suggested approach balance the load, but it also considers the priority of jobs that have been withdrawn from heavily loaded Virtual Machines. The jobs that have been withdrawn from these VMs are treated like honey bees, which are the global information updaters. The job priorities are also taken into account by this algorithm. Load balancing inspired by honey bee behavior enhances overall processing performance, while priority-based balancing focuses on lowering the amount of time a job spends waiting in a VM queue. As a result, the time it takes for VMs to respond is reduced.The proposed approach shows a significant improvement in average execution time as well as a reduction in task queue waiting time.

The authors [2] presented a heuristic greedy algorithm-based VM placement technique. The author offers a VM deployment and live migration model in this technique to increase resource usage and reduce power consumption. The heuristic approach forecasts the demand and maps CPU-intensive and memory-intensive workloads to the same physical server, reducing energy consumption and balancing the workload.

The author [3] created a statistical mathematical framework for VM placement that includes all virtualization costs in the dynamic migration process. The suggested dynamic virtual machine placement approach enables VM request scheduling and lifecycle migration in Cloud datacenters to reduce active server participation and power consumption.

Based on minimum virtual machine migration, the author presented a VM allocation approach [4]. Fixed double threshold, double resource threshold, and dynamic double threshold are the three solutions developed. In the second phase, each strategy is applied and a VM is assigned to a physical datacenter. These lower and upper bound resource usage threshold policies outperform the single threshold strategy, according to the results.These solutions result in lower power consumption, fewer SLA violations, and fewer virtual machine migrations.

The major concern of a cloud service provider is to answer two questions: where to put virtual machines initially and where to migrate virtual machines when they need to be moved. The use of virtual machines (VMs) can help to reduce datacenter overloading and the participation of actual servers, resulting in better resource usage and power savings. For better performance and lower Cloud system service costs, it's critical to detect overcrowded servers quickly. In [5,] a general detection strategy for the overloaded server is proposed using logistic regression and median absolute derivation approaches. This detection algorithm can be used with any VM placement and migration algorithms to detect overcrowded servers.

The authors [6] described a task-based virtual machine placement method in which tasks are mapped to VMs depending on their demand and VMs are then put on physical machines accordingly. To save energy, the algorithm minimizes the number of active servers participating in servicing VMs. It also reduces the Cloud system's task rejection rate and makespan.

The author [7] proposes an energy consumption model that is based on statistical methods and can estimate VM power usage with a 3 percent -6 percent error rate.A workload threshold is specified for each server in this method, and if a server's workload threshold is exceeded, the VM is transferred from that overloaded server to another server to lower the overloaded server's energy usage. Without compromising QoS, this strategy can yield a significant reduction in power usage.

The author [8] uses a linear integer programming model and a bin packing model to build two exact algorithms for VM placement and consolidation that are compared to a heuristic-based best-fit approach for decreasing power consumption and VM migration cost. The combined use of these two algorithms leads in a significant reduction in power usage, according to the findings.

The author [9] offered two strategies, one for efficient power usage using the DVFS methodology and the other for VM consolidation using the DVFS technique. The first technique is utilized to determine performance deterioration as a function of power consumption, and it results in DVFS-aware workload management that saves up to 39.14 percent of energy in dynamic workload scenarios. When allocating workload to achieve QoS, the second VM consolidation technique is similarly selected dynamically.

The author [10] proposes a job consolidation algorithm based on the DVFS technique for efficient resource use on heterogeneous Cloud physical machines. The proposed algorithm will efficiently replace occupations in order to save energy.

Shutting down idle physical servers is one technique to cut energy consumption in Cloud datacenters. The dynamic round robin technique was presented in [11] as an energy-aware virtual machine scheduling mechanism. When compared to existing scheduling algorithms, the results showed that the approach saves 43.7 percent energy and 60 percent physical machine utilisation.

The authors[12] provide a model for estimating energy consumption that takes into account the ongoing tasks performed by virtual machines in order to estimate each VM's power usage. The proposed methodology additionally schedules the virtual machines to ensure their energy costs.

Most energy-efficient methods employ the VM migration technique, but in [13], the author proposed EMinTRE-LFT, an energy-aware virtual machine scheduling algorithm based on the concept that a reduction in power consumption is directly equivalent to a reduction in all physical server completion times. For simulation, the author used the OpenStack Nova scheduler and compared it to other algorithms.

Due to the dynamic and unexpected nature of Cloud user requests, cloud scheduling algorithms confront numerous obstacles. The author [14] presented a method that requires no prior knowledge of the user's request. To discover a balance between energy usage and system performance, the author did a mathematical analysis.

The author [15] proposes a real-time dynamic scheduling algorithm for scheduling distributed applications in a distributed system in order to reduce power usage. To arrive at the best solution, the suggested method employs heuristics and resource allocation strategies. With order dependent configuration between tasks for VM and power setup for different Cloud designs, it reduces power consumption and task execution time.

In [16], the authors developed a virtual machine allocation technique based on interior search for efficient energy usage and resource utilization. The suggested algorithm's model and simulation were tested on CloudSim, and the amount of energy consumed was compared to the Genetic Algorithm (GA) and Best-fit Decreasing (BFD) algorithms.

The cloud provider assigns virtual machines to the customer's application based on demand, and these virtual machines are assigned to physical equipment. For efficient resource allocation, several resource allocation algorithms rely on the VM's resource use history. The author [17] developed a QoS-aware virtual machine allocation strategy based on resource utilization history to improve service quality while lowering energy consumption.

Cloud datacenters give services to Cloud apps that use a lot of energy and emit a lot of carbon. To address this issue, the author [18]developed an energy-aware VM allocation method, which efficiently provide and schedule Cloud datacenter resources to the user's tasks, lowering datacenter energy usage and improving service quality.

Many researchers have worked on energy efficiency in Cloud Computing, but some have focused on energy efficiency in a particular type of Datacenter. The author[19] presented a power consumption method for video streaming datacenters. With the power-law characteristic, they proposed a solution for VM management. It forecasts future VM resource need based on video popularity, allocates appropriate resources for that VM, and turns down idle servers in datacenters to save power consumption. When compared to Nash and Best-fit algorithms, the results showed that our approach used less power.

SUMMARY of Literature review

This survey differed from others in that it covered both power usage and datacenter resource utilization. In this research study, we looked at various VM consolidation concepts. Different sub-problems for VM consolidation are studied in each research study, such as VM Placement, VM Selection, Host Overload Detection, and Host UnderLoad Detection. Each study focuses on one or more significant sub-problems and is successful in resolving the VM consolidation issues.

III. Conclusion

This paper offers a survey of cloud-based VM consolidation approaches. VM consolidation gives a big potential to conserve energy in the datacenter. A live VM migration is a type of VM consolidation that identifies under-utilized physical machines that can be turned off or converted to a low-power mode. Obtaining the necessary level of Quality of Service (QoS) between end users and cloud service providers, on the other hand, is crucial. We looked into the various authors' VM consolidation techniques. This study can assist readers in determining the advantages and limitations of various energy efficient algorithms that have been proposed in the literature. We will compare these algorithms in the future using the cloudsim tool to determine energy consumption and SLA violations.

Reference

[1] Dinesh Babu LD, D. B., & Krishna, P. V. (2013) "Honey bee behavior inspired load balancing of tasks in cloud computing environments", Applied Soft Computing, Vol.13, No.5, pp. 2292-2303.

[2] J. Wang, C. Huang, K. He, X. Wang, X. Chen and K. Qin, "An energyaware resource allocation heuristics for VM scheduling in cloud", Proceedings of 2013 IEEE International Conference on High Performance Computing and Communications (HPCC 2013) and 2013 IEEE International Conference on Embedded and Ubiquitous Computing (EUC 2013), pp. 587-594, 2013.

[3] X. Zheng and Y. Cai, "Dynamic Virtual Machine Placement for Cloud Computing Environments," 2014 43rd International Conference on Parallel Processing Workshops, Minneapolis, MN, pp. 121-128, 2014.

[4] D. Dad and G. Belalem, "Efficient allocation of VMs in servers of data center to reduce energy consumption," 2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech), Rabat, pp. 1-6, 2017.

[5] Y. Jararweh, M. Issa, M. Daraghmeh, M. Al-Ayyoub and M. Alsmirat, "Energy efficient dynamic resource management in cloud computing based on logistic regression model and median absolute deviation", Sustainable Computing: Informatics and Systems, vol. 19, pp. 262-274, 2018.

[6] S. K. Mishra et al., "Energy-efficient VM-placement in cloud data center", Sustain. Comput. Informat. Syst., pp. 48-50, 2018.

[7] Scheduling Algorithm of Cloud Computing System," 2013 InternationalConference on Information Science and Cloud Computing Companion,Guangzhou, pp. 219-224, 2013.

[8] C. Ghribi, M. Hadji and D. Zeghlache, "Energy Efficient VMScheduling for Cloud Data Centers: Exact Allocation and MigrationAlgorithms", 2013 13th IEEE/ACM International Symposium on ClusterCloud and Grid Computing, pp. 671-678, May 2013.

[9] P. Arroba, J. M. Moya, J. L. Ayala and R. Buyya, "DVFS-Aware Consolidation for Energy-Efficient Clouds," 2015 InternationalConference on Parallel Architecture and Compilation (PACT), SanFrancisco, CA, pp. 494-495, 2015.

[10] P. Sanjeevi and P. Viswanathan, "Towards energy-aware job consolidation scheduling in cloud," 2016 International Conference onInventive Computation Technologies (ICICT), Coimbatore, pp. 1-6,2016.

[11] C. Lin, P. Liu and J. Wu, "Energy-Aware Virtual Machine DynamicProvision and Scheduling for Cloud Computing," 2011 IEEE 4thInternational Conference on Cloud Computing, Washington, DC, 2011,pp. 736-737.

[12] N. Kim, J. Cho and E. Seo, "Energy-Based Accounting and Schedulingof Virtual Machines in a Cloud System," 2011 IEEE/ACM InternationalConference on Green Computing and Communications, Sichuan, 2011,pp. 176-181.

[13] N. Quang-Hung and N. Thoai, "Minimizing Total Busy Time withApplication to Energy-Efficient Scheduling of Virtual Machines in IaaSClouds," 2016 International Conference on Advanced Computing and Applications (ACOMP), Can Tho, 2016, pp. 141-148.

[14] Y. Chen, C. Lin, J. Huang, X. Xiang and X. Shen, "Energy EfficientScheduling and Management for Large-Scale Services ComputingSystems", IEEE Transactions on Services Computing, vol. 10, no. 2, pp.217-230, 2017.

[15] F. Juarez, J. Ejarque and R. Badia, "Dynamic energy-aware scheduling for parallel task-based application in cloud computing", FutureGeneration Computer Systems, vol. 78, pp. 257-271, 2018.

[16] M. J. Usman, A. Samad, Ismail, H. Chizari and A. Aliyu, "Energy-Efficient virtual machine allocation technique using interior searchalgorithm for cloud datacenter," 2017 6th ICT International StudentProject Conference (ICT-ISPC), Skudai, 2017, pp. 1-4.

[17] A.Horri, M. Mozafari and G. Dastghaibyfard, "Novel resource allocation algorithms to performance and energy efficiency in cloudcomputing", The Journal of Supercomputing, vol. 69, no. 3, pp. 1445-1461, 2014.

[18] A.Beloglazov, J. Abawajy and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for Cloudcomputing", Future Generation Computer Systems, vol. 28, no. 5, pp.755-768, 2012.

[19] H. Tseng, T. Yang, K. Yang and P. Chen, "An Energy Efficient VMManagement Scheme with Power-Law Characteristic in VideoStreaming Data Centers", IEEE Transactions on Parallel andDistributed Systems, vol. 29, no. 2, pp. 297-311, 2018.