ISSN: 2321-7782 (Online)
Volume 2, Issue 2, February 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Paper / Case Study
Available online at: www.ijarcsms.com

A Survey on Different Virtual Machine Placement Algorithms

Chirag J. Rathod[1]
Student of Master in Computer Science & Engineering
Gujarat Technological University
Ahmedabad – India

Abstract: ***Cloud computing provide computing resources as per user requirement. Virtualization provides the facility to efficient use of hardware. Virtual machines are utilized to satisfy user requirement and it's placed on physical machine of cloud such a way that utilize hardware resources and electricity in cloud. For utilize the number of physical machine used helps in cutting down the power consumption by substantial amount. So best way to map virtual machine to physical machine such that the number of running physical machine is minimized. The virtual machine placement problem with the target of minimizing the energy consumption, which shows the utilization of resources and reducing the cost of data center in cloud. To characterize multi dimensional resource usage states of physical machine, a multi-dimensional resource cube is presented. Based on this model we describe different Virtual machine placement algorithm.***

Keywords: VM, RUV, TVC, RD, Multi-Dimensional Resource cube.

I. INTRODUCTION

Cloud computing [1] is a large scale distributed computing environment where computing resources such as memory, processing power, bandwidth, etc. are available on demand to the users. Cloud can be defined as "A Cloud is a type of parallel and distributed system The cloud computing consists of various service models such as Infrastructure as a Service(IAAS) [2], Platform as a Service (PAAS) and Software as a Service (SAAS) [3]. These models are made available to the users through virtualization techniques. The users' demands are satisfied by a set of servers hosted on virtual machines. The virtual machines utilize the resources of underlying physical machines or nodes provide and operated by organizations called 'cloud providers'. Some examples of cloud providers include Amazon EC2, GoGrid and Rackspace Cloud. The National Institute of Standards and Technology (NIST) [4] defines three service models of cloud computing:

A. Infrastructure as a Service (IaaS)

The consumer uses fundamental computing resources such as processing power, storage, networking components or middleware. The consumer can control the operating system, storage, deployed applications.

B. Platform as a Service (PaaS)

The consumer uses a hosting environment for their applications. The consumer controls the applications that run in the environment (and possibly has some control over the hosting environment), but does not control the operating system, hardware or network infrastructure on which they are running. The platform is typically an application framework.

C. Software as a Service (SaaS)

The consumer uses an application, but does not control the operating system, hardware or network infrastructure on which it's running. With advent of high capacity computers, underutilization time is increased. Adopting the use of virtual machines in such large scale environments enhances the number of 2 available servers through multiple operating system instances on a single node, there by achieving efficient hardware utilization. Virtualization is to create a virtual version of a device or resource, such as a server, storage device, network or even an operating system. A physical machine consisting of CPUs, memory, disk,

and network computing resources can be virtualized into logical virtual machines consisting of their own virtual CPUs, virtual memory, virtual disk, and virtual networking resources. Virtualization gives the advantage of running multiple operating systems on a single machine so that resource utilization of system can be increased drastically. Virtual Machine Monitor (VMM) or Hypervisor provides an interface between hardware and operating system to allow multiple guest operating systems to run on top of it. VMM allows for creation, destruction and migration [5] of virtual machines. Examples of Hypervisors are XEN [6], KVM, VMware etc. The XEN hypervisor is an open source industry standard for virtualization. However, there may be a number of underutilized nodes due to the inefficient mapping of virtual machines to physical machines. Minimizing the number of physical machines utilized helps in cutting down the power consumption drastically. The placement algorithm for virtual machines in a data center allocates various resources such as memory, bandwidth and processing power etc. from a physical machine to virtual machines such that the number of physical machines used is minimized.

II. VIRTUAL MACHINE PLACEMENT

The problem of virtual machine placement is core of cloud computing. Virtual Machine Placement [12] is a mapping between physical machines to virtual machines. Better placement strategy is to decrease running physical machines, increase the performance of infrastructure, reducing service level agreement violation rates and better administrative capabilities to the cloud provider. Placement goal [13] can be maximizing the usage of available resources or it can be saving of power being shut down some unused physical machines. Based on placement goal, algorithms will be divided into two categories._ QOS based Algorithms [14]: By providing maximum resources to virtual machines, then achieving the maximum quality of service. Maximizing the quality of a service in any situation is the goal of these algorithms power Based [14]: Power might be saving with minimize number of running physical machines. Minimize number of running physical machines without exceeding the probability of violating the user service level agreements.

III. NORMALIZED RESOURCE CUBE

Each physical machine can be represented with Normalized Resource Cube and each axis Represents different resource type.

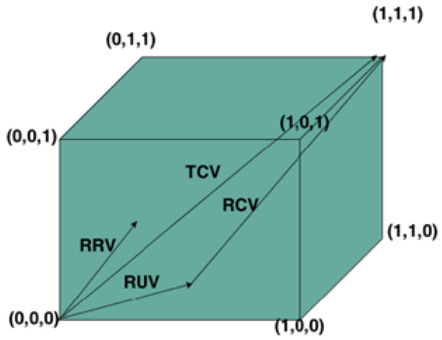

Figure 1: Resource Cube

Different vector notations of Normalized Resource cube:

TCV: Total capacitance vector of a physical machine.

RUV(t): Resource utilization vector of a physical machine at time t. Represents percentage of utilization of each resource at time t.

RCV(t): Resource capacity vector of a physical machine at time t. Represents how much percentage of resource capabilities is left.

RRV(t): Resource requirement vector of a virtual machine at time t, Represents how much percentage of resources are needed.

Initially each physical machine RUV is (0, 0, 0). If one virtual machine comes up with RRV (a, b, c), which is less than RCV (TCV-RUV) then physical machine having ability to run. After placement of this virtual machine, physical machine RUV is increased and RCV is decreased by RRV.

IV. VIRTUAL MACHINE PLACEMENT ALGORITHM

Cluster controller will collect resource utilization information from all node controllers with in the cluster. Using collected information, cluster controller will find mappings between virtual machines and node controller (physical machines).

1. First Fit [8]: While placing the VMs, first fit algorithm can be seen as a locally optimal algorithm in greedy manner. When a VM request arrives, the firstly scanned one with sufficient available resources will be selected to host the VM. To avoid unnecessary scanning, the fully loaded PMs are not scanned. If there is no PM with sufficient available resources, a new PM will start up and create the requested VM on the newly started PM. It is a greedy manner to place the VMs and, by this way, we can achieve a local optimization. Physical machines are available with their indices [12]. Each virtual machine will start searching from starting node, whenever the physical machine fulfills requirement of the virtual machine, then virtual machine will start on that node. Average time complexity of this algorithm is O(n/2).

2. Next Fit [8]: Algorithm will use variable name called next. At Initial, next is null. First virtual machine search starts from the first node. When physical machine satisfies the resource requirement, virtual machine starts on that physical machine and next is replaced by current physical machine. If there is no PM with sufficient available resources, a new PM will start up and create the requested VM on the newly started PM. For the next virtual machine search starts from next to the stored next physical machine, search continues up to desired physical machine.

3. Random Fit [8]: Randomly a physical machine is chosen for placing virtual machine. If physical machine satisfies the resource requirement, virtual machine starts on that physical machine. If it not satisfied randomly another physical machine is chosen. This process will continue for every virtual machine.

4. Least full first [16]: Find the physical machine which is least full. If the physical machine satisfy resource requirement of virtual machine then virtual machine will start. If it doesn't satisfy other physical machines also doesn't satisfy. Because virtual machine is been tested with least full physical machine. This process will continue till virtual machines exhausted. Time complexity of this algorithm is O(n).

5. Most Full First [16]: In this process, physical machines are sorted from most full to least full. Once sorted has been done first fit process will takes place. This process continues up to virtual machine allocated to any host. Time complexity of this algorithm is O(n).

6. First Fit Decreasing (Single Dimension) [17]: Both physical machines and virtual machines are sorted from high capacity to least capacity machines. Once both physical and virtual machines are arranged, First fit process takes place. Time complexity of this algorithm is O(n+m)

7. First Fit Decreasing product [18]: This is also similar to the First Fit Decreasing. Each physical machine or virtual machine is represented with different resource capacities and each resource capacity is normalized in to percentage of utilization. Example: If physical machine utilization of RAM is 512Mb and total RAM is 1024Mb then RAM utilization is 50%. Volume of physical machine or virtual machine can be defined as product of all resource types. This volume is used instead of capacity in First Fit Decreasing (Single Dimension)[13].

8. First Fit Decreasing sum [18]: This is also similar to the First Fit Decreasing. Each physical machine or virtual machine is represented with different resource capacities and each resource capacity is normalized in to percentage of utilization. Virtual or physical machines volume are defined as weighted sum of values. i.e. V = sum(wi _ Ri), where i th resource weight wi= total

demand for i th resource/total available capacity for i th resource. This volume is used instead of capacity in First Fit Decreasing (single Dimension).

9. Dot product [19]: When choosing the next virtual machine to place on an open host, Dot product algorithm will align remaining capacity with dimensions and resource demands of a virtual machine. At time t, let H(t) - vector of residual capacity of the current host. Which ever the physical machine gives maximizes the dot product , place the virtual machine. dotproduct = ån i=1WiViH(t)I Wi - i th resource weight factor, is same as calculated in first fit decreasing sum. Vi - Virtual machine i th resource demand. H(t)i - Host i th resource residual capacity.

10. Minimizing Angle [19]: All physical machines resource utilization vector (RUVi) are known and also virtual machine v resource demand is known. For each physical machine, if it satisfy the resources demand of virtual machine then find the angle between RUVi + RD vector and total capacitance vector TCV (1,1).

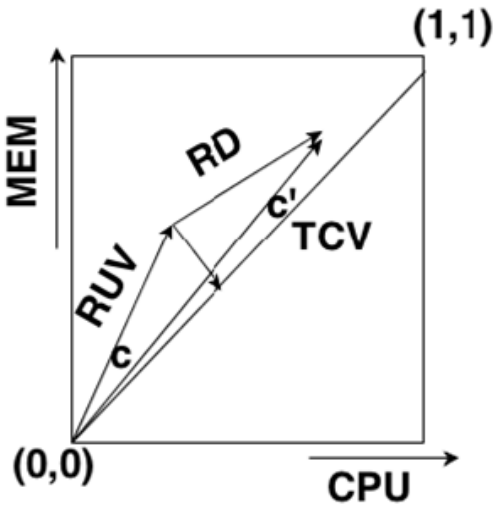

Figure 2: Minimize Angle [19]

In figure-2 the angle shown by c'. Place the virtual machine on which will give minimum angle. Place complementary virtual machines on single physical machines might balance resources on physical machine; this will lead to higher resource utilizations.

V. CONCLUSION

First fit, next fit and random fit algorithms are simple algorithms. These algorithms are do not consider the resource capacities needed for placing virtual machine and not consider the resource consumptions in physical machines. Least full first, most full first, first fit decreasing (single dimension) algorithms are designed based on CPU capacity. These algorithms consider CPU capacity needed for placing virtual machine and consider CPU consumptions in physical machines. These algorithms do not consider other resources. First fit decreasing product, first fit decreasing sum and dot product algorithms are designed for multi dimensional resources. In these algorithms, physical/virtual machines' resource volumes are calculated using all resources capacities. These volumes are used in placing virtual machines. Resource capacities are not directly involved in placing virtual machines. In minimizing angle algorithm all resources are involved directly for placing virtual machines. In this algorithm different resource consumptions are balanced so that physical machines are minimized. This algorithm concentrates on balancing resources instead of maximizing resource usage of physical machines in each dimension.

Acknowledgement

With the cooperation of my guide, I am highly indebted to Asst. Prof. Mohammed Husain Bohara, for his valuable guidance and supervision regarding my topic as well as for providing necessary information regarding review paper. I am very much thankful to Prof. G. B. Jethava for helping me in text preparation.

References

1. R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility," Future Gener. Comput. Syst., vol. 25, no. 6, pp. 599–616, Jun. 2009.
2. O. D, "Cloud computing service models , part1: Infrastructire as a service," CEO, Creative RIA, Jan. 2011, doi:10.1109/COASE.2010.5584153.
3. H. La and S. Kim, "A Systematic Process for Developing High Quality SaaS Cloud Services," in Cloud Computing, ser. Lecture Notes in Computer Science, M. Jaatun, G. Zhao, and C. Rong, Eds. Berlin, Heidelberg: Springer Berlin / Heidelberg, 2009, vol. 5931, ch. 25, pp. 278–289.
4. C. C. U. C. Group, "Cloud computing use cases white paper version 1.0," http://www.scribd.com/doc/18172802/Cloud-Computing-Use-Cases-Whitepaper, 2009.
5. S. H. J. G. H. E. J. C. L. I. P. A. W. Christopher Clark, Keir Fraser, "Live migration of virtual machines," in Proceeding NSDI'05 Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation, vol. 2. USENIX Association Berkeley, CA, USA ï£¡2005, 2007, pp. 273–286.
6. P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, "Xen and the art of virtualization," in Proceedings of the nineteenth ACM symposium on Operating systems principles, ser. SOSP '03. New York, NY, USA: ACM, 2003, pp. 164–177.
7. R. Moreno-Vozmediano, R. S. Montero, and I. M. Llorente, "Iaas cloud architecture: From virtualized datacenters to federated cloud infrastructures," Computer, vol. 45, no. 12, pp. 65–72, 2012.34.
8. K. Mills, J. J. Filliben, and C. Dabrowski, "Comparing vm-placement algorithms for on-demand clouds." in CloudCom, C. Lambrinoudakis, P. Rizomiliotis, and T. W. Wlodarczyk, Eds. IEEE, 2011, pp. 91–98.
9. S. Shang, B. Wang, J. Jiang, Y. Wu, and W. Zheng, "An intelligent capacity planning model for cloud market," Journal of Internet Services and Information Security (JISIS), vol. 1, no. 1, pp. 37–45, 5 2011.
10. A. Mehta, M. Menaria, S. Dangi, and S. Rao, "Energy conservation in cloud infrastructures," IEEE SysCon 2011, vol. 1, pp. 456–460, 2011.
11. J. Shetty, A. M. R, and S. G, "Article: A survey on techniques of secure live migration of virtual machine," International Journal of Computer Applications, vol. 39, no. 12, pp. 34–39, February 2012, published by Foundation of Computer Science, New York, USA.
12. R. G. Chris Hyser, Bret Mckee and B. J. Watson, "Autonomic virtual machine placement in the data center," in HP Labs technical report, 2007.
13. U. Bellur, C. S. Rao, and M. K. SD, "Optimal placement algorithms for virtual machines," CoRR, vol. abs/1011.5064, 2010.
14. A. Shankar and U. Bellur, "Virtual machine placment in computing clouds," CoRR, vol. abs/1011.5064, April 10, 2010.
15. W. F. de la Vega and G. S. Lueker, "Bin packing can be solved within 1+epsilon in linear time." Combinatorica, vol. 1, no. 4, pp. 349–355, 1981.
16. V. P. V. R. K. T. L. U. S. Lee, R. Panigrahy and U. Wieder, "Validating heuristics for virtual machine consolidation." in Microsoft Research. Microsoft research, p. 14 pp.
17. E. G. Coffman, Jr., M. R. Garey, and D. S. Johnson, "Approximation algorithms for np-hard problems," D. S. Hochbaum, Ed. Boston, MA, USA: PWS Publishing Co.1997, ch. Approximation algorithms for bin packing: a survey, pp. 46–93.
18. S. Srikantaiah, A. Kansal, and F. Zhao, "Energy aware consolidation for cloud computing," in Proceedings of the 2008 conference on Power aware computing and systems, ser. HotPower'08. Berkeley, CA, USA: USENIX Association, 2008, pp. 10–10. 35.
19. M. Mishra and A. Sahoo, "On theory of vm placement: Anomalies in existing methodologies and their mitigation using a novel vector based approach," in Proceedings of the 2011 IEEE 4th International Conference on Cloud Computing, ser. CLOUD '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 275–282.
20. N. Singh and S. Rao, "Energy optimization policies for server clusters," in 6th Annual IEEE Conference on Automation Science and Engineering (IEEE CASE 2010), Toronto, Canada, Aug. 2010, pp. 293–300, doi:10.1109/COASE.2010.5584153.
21. A. Silberschatz, P. B. Galvin, and G. Gagne, Applied operating system concepts, 1st ed. New York [u.a.]: Wiley, 2000.
22. B. Radojevic and M. Zagar, "Analysis of issues with load balancing algorithms in hosted (cloud) environments." in MIPRO. IEEE, 2011, pp. 416–420.
23. A. K. Norman Bobro and K. A. Beaty, "Dynamic placement of virtual machines for managing sla violations," in In Integrated Network Management. IEEE, 2007, p. 119ï£¡128.
24. A. Kochut and K. Beaty, "On strategies for dynamic resource management in virtualized server environments," in Proceedings of the 2007 15th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, ser. MASCOTS '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 193–200.
25. S. Shish and C. Tsokos, "A weighted moving average process for forecasting," Journal of Modern Applied Statistical Methods, vol. Vol 16, No. 2, 2008.
26. R. N. Calheiros, R. Ranjan, C. A. F. D. Rose, and R. Buyya, "Cloudsim: A novel framework for modeling and simulationof cloud computing infrastructures and services," CoRR, vol. abs/0903.2525, 2009.

AUTHOR(S) PROFILE

Chirag Rathod, received the BE degree in Computer Engineering from Sankalchand Patel College of Engineering Visnagar Under Gujarat Technological University in 2012, pursuing Master of Engineering degree in Computer Science and Engineering from Parul Institute of Engineering & Technology Under Gujarat Technical University, Ahmedabad.