

Resource Management Using Virtual Machine Migrations

Pradeep Kumar Tiwari and Sandeep Joshi

Abstract Virtualization is a component of cloud computing. Virtualization does the key role on resource (e.g., storage, network, and compute) management and utilized the resource by isolated virtual machines (VMs). A good VMs migration system has an impact on energy efficiency policy. Good resource administration policy monitors the on-demand load management, and also manages the allocation/relocation of the VMs. The key challenges are VMs isolation and migration in heterogeneous physical servers. Threshold mechanism is effective to manage the load among VMs. Previous research shows thrashing is a good option to manage load and migrate the load among VMs during failover and high load. Linear programming (LP), static and dynamic approaches are superior to manage the load through VMs migration. This research proposes threshold-based LP approach to manage load balance and focus on dynamic resource management, load balance goals, and load management challenges.

Keywords Virtual machine · Load balance · Migration · Threshold

1 Introduction

Virtualization technology makes flexible use of resources and provides illusion of high availability of resources to end users as pay-as-use concept [1]. The objectives of effective resource management are optimizing energy consumption, flexible load balance and managing the failover cluster. Research shows an average of 10–15 %

P.K. Tiwari (✉) · Sandeep Joshi
Department of Computer Science & Engineering, Manipal University Jaipur,
Jaipur, India
e-mail: pradeeptiwari.mca@gmail.com

Sandeep Joshi
e-mail: sandeep.joshi@jaipur.manipal.edu

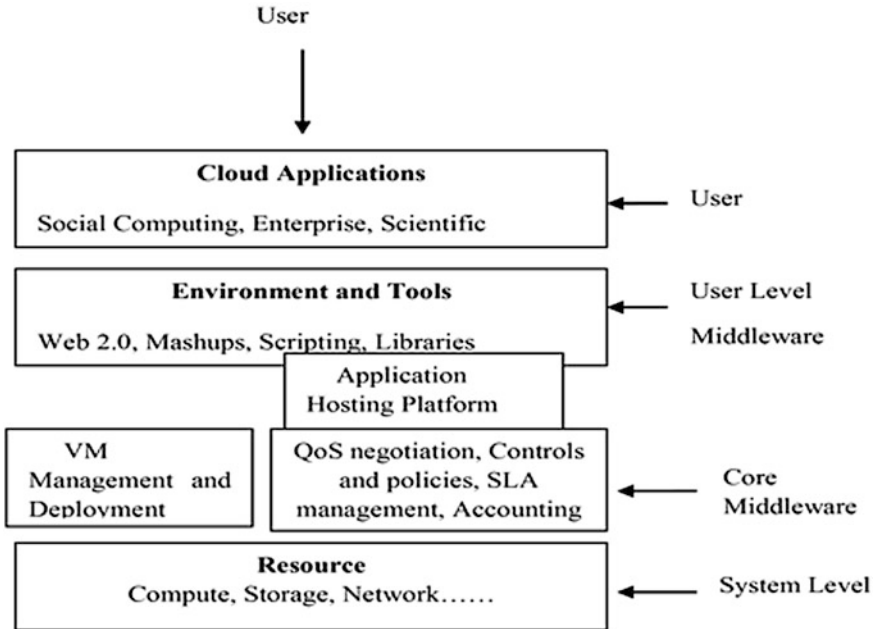


Fig. 1 Model of cloud computing system [3]

capacity is not utilized from 30 % cloud systems. Effective resource utilization can reduce the number of servers [2]. Figure 1 shows layered model of cloud computing system. Cloud computing system can be divided into four parts (User Level, Core Middleware, User Level Middleware, and User Level). Resource management (Compute, Storage, and Network) is a part of system level and it is managed by VM management, which belongs to core middle level. Resource allocation management system can be centralized or decentralized. VM migration system provides isolation, consolidation, and migration of workload. The main purpose of VM migration is to improve performance, fault tolerance, and management of resources [3]. The main focus of cloud system is selection of resources, management of load balance, and maximum utilization of available physical resources via VMs.

Resources are either statically or dynamically assigned to VMs. Static policy is a predefined resource reservation to VM according the need of end user, but in dynamic allocation resources need can be increased and decreased according to the user demands [4].

Server consolidation with migration control is managed by static consolidation, dynamic consolidation, or dynamic consolidation with migration controls.

1.1 Static Consolidation

This approach defines a pre-reserved dedicated resource allocation to VM according the need of end users. VMs resource allocation is based on the total capacity of physical system and the migration does not happen till all demands are changed.

1.2 Dynamic Consolidation

This is a periodicity-based VM migration approach and the migration is based on the current demand. If the required VMs resource demands are higher than the physical available capacity then VMs migrate to another physical server.

1.3 Dynamic Consolidation with Migration Control

This approach gives the stability during high resource demand, hotspot, and frequently changing resources demands. This approach reduces the required number of physical servers and saves the energy consumption. This approach is based on heuristic and round robin mechanism.

VMs migration depends on the resource availability of VMs. If the required load is higher than the thresh value of VM then load will be migrated to another VM. Users' resource needs, reliability criteria, and scalability terms are clearly mentioned in SLA. Providers provide the resource according to the SLA commitment. To avoid SLA violation, providers provide committed pre-reserved resources and adapt the threshold-based load management policy [5]. Three basic questions arise before VM migration. These questions are: When to migrate? Which VM to migrate? Where to migrate? Figure 2 shows the dynamic resource migration procedure steps. First step monitors the load, if the load is higher than available thresh, then it estimates the load and VMs migrate to another physical machine [1].

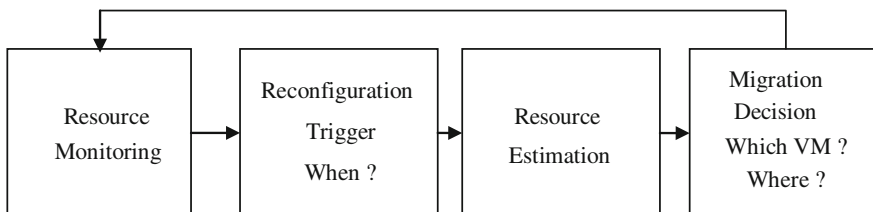


Fig. 2 Dynamic resource management steps [1]

Resource requirements may be different during frequently changing environments, but fixed load does not constrain all time. The best-suited VMs migration strategies are dynamic/heuristic approach for frequently changing environments.

2 VM Migration Management System

Dynamic VM migration management policy manages the on-demand resources availability to users. Upper and lower thresh limits can help to indentify the maximum and the minimum load limits of VMs. Researchers proposed several VM migration techniques to migrate the load from VM to VM and from physical server to physical server. Bin packing approach is good approach for offline resource management but is not effective for the optimal use of CPU. Green computing approach reduces the number of utilized servers and is energy efficient. Memory-aware server consolidation is managed by balloon filter and finger printing system to share VMs locations among heterogeneous servers [6, 7]. Thresh value shows a high capacity of VM. VM load migrates to another VM when a high resources demand arises.

Resources are managed by load monitor and VM allocation/relocation planner. Load monitor monitors the on-demand dynamic load of VMs and compares with available physical resource; relocation planner system plans relocation of VM; and VM controller controls the VM migration among the physical servers and manages the failover cluster. A high demand of resources is the cause of hotspot. Hotspot is detected when server becomes overload due to over demand of resources [8].

LP-heuristic approach [4] proposed linear programming based heuristic WFD (worst fit decreasing), BFD (best fit decreasing), FFD (first fit decreasing), and AWF (almost worst fit decreasing). This is a two-way resource management approach; first policy indentifies VMs and maps the capacity from available physical capacity and second approaches short physical server increasingly according to their capacities with respect to lexicographic order. The objective of linear programming objective is minimization of the required physical server and to map VMs resource availability from hosted physical server.

CP-heuristic [9] resource management system proposed CP (constraint programming) constraint propagation and decision-making research. CP resource allocation results are ten times much faster than integer programming resource scheduling. Heuristic approach uses first fit and best fit methodology for scheduling the short job.

The researcher's visions are green computing and minimal downtime of VM migration. Effective utilization of resources and proper on/off of utilized and non-utilized physical system can contribute to the green computing concept. Researchers continue contributing his/her effort to energy aware resource utilization approach. Energy aware data center location approach [10] proposed MBFD (modified best fit decreasing) and MM (minimization of migration) approach. MBFD optimizes the current VM allocation and chooses the most energy-efficient

nearest physical server for VM to migrate and MM approach minimizes the VM migration needs.

3 Resource Management Goals

Resource management is a semantic relationship between resource availability and resource distribution. Network, compute, and storage are the main resource components. Resource manager manages the resources according to the availability of resources and provides them to the end user on the basis of SLA agreement.

The main goals of resources distribution are (a) performance isolation, (b) resource utilization, and (c) flexible administration.

3.1 Performance Isolation

VMs are isolated from each other and resource utilization of VMs do not affect capacity of another VMs. These VMs are allocated on same physical server. Failure of VM does not affect the performance. Load will migrate to another VM in minimum downtime. Hyper-V provides quick migration and VMware provides the live migration facility. VMs have own reserved network, storage, and compute resources. Resource availability depends on SLA-based user requirements [11, 12].

3.2 Resource Utilization

Resource utilization is based on the maximum consumption of available resources and the minimum energy consumption. Resource manager must give attention to SLA-based on-demand dedicated resource requirements. Resource manager maps the highest requirements on a day, a weekly, and a monthly basis and observes the type of resource needs. This analysis can manage the rush hour resource requirements of end user. VMs resource utilization can be measured by the capacity of physical machine and unused resource can be used as reserved capacity [13].

3.3 Flexible Administration

Resource availability administrator must be able to handle high-load resource and VM migration management in a synchronized manner. VMware uses distributed resource scheduler (DRS) to manage VMs capacity (resource reservations, priorities, and limit) and VM migration. VMware's distributed power management

(DPM) system manages the power on/off management of used/unused VM and plays a vital role in energy aware resource management [14].

Internet small computer system interface (iSCSI) Internet protocol flexibly manages the target storage server and uses storage network protocol (SAN) consolidate storage into storage array. SAN can manage storage consolidation and disaster recovery.

4 Resource Management Challenges

Dynamic resource management with performance isolation, flexible administration, and on-demand resource utilization is not easy to manage simultaneously. Researchers face many problems to utilize resources.

4.1 Flexible VM Migration

Researchers proposed several VM migration methodologies, some are dedicated to resource management and others are energy efficient. Allocation and relocation of VM in heterogeneous environment with load management policy are complicated to manage. Threshold mechanism can map the load of VMs and physical server. Interface management among the VMs and physical servers is complex and tough. Researcher need to more effort to dynamic VMs migration policy [2, 8].

4.2 Storage Management

Researchers proposed [9, 10] virtualized data center management policy. The most popular techniques are pre-copy, post-copy for storage management. CP-heuristic is also a good mechanism to manage data center locations but these techniques are not good enough. Data scattered among heterogeneous locations and gathering of data into a single location in efficient response time is complex.

4.3 Hotspot Management

Hotspot is an overloaded condition of a physical machine. VMs resource access condition is greater than threshold value of the physical machine. This condition is called as hotspot. In static load balance, hotspot condition can be predetermined but in dynamic load, management policy cannot predetermine the hotspot because of its on-demand resource management policy [1].

4.4 SLA Violation

Service level agreement (SLA) defines resource usage patterns of application utilization of storage and computing resource. Cloud service provides pay-as-you-go model. Users only need dedicated pre-reserved on-time resource availability. Cloud service provider gives assurance of quality of service (QoS). Week management of resources, load imbalance, hotspot, and scattered data among heterogeneous servers are main causes of SLA violations [15].

Some service provides offer only guarantee of resource availability rather than performance. The main goal of cloud service providers is the maximum utilization of resources in minimum availability. Server heterogeneity, high resource demand, and failure of cluster affect the resource management and the availability of resources to users. Cloud providers ensure on-demand, robust, scalable, and minimal downtime access services.

4.5 Load Imbalance

End user resource requirement changes dynamically. This can lead to load imbalance in VMs. VMs use physical machine resource capacity, sometimes VMs load may be higher to a physical machine, then VM will migrate to another physical machine. Some physical machines are highly loaded and some are less loaded. It may cause discrepancy in utilizations of physical servers. Overloaded VMs downtime is higher than low-loaded ones [1].

5 Load Balance Concept and Management

High performance computing (HPC) can be achieved by DRS in minimal meantime. Threshold-based load migration with LP is a superior approach to load migration. This approach is static during the distribution of VMs ID allocation and dynamic in load migration among the VMs. This article proposes two algorithms: first algorithm allocates ID to all high-load, low-load, and average load VMs and second algorithm controls the process of load migration among VMs.

First algorithm counts and locates the VMs ID of single physical system. It classifies the VMs according to high, average, and low load. Its first array stores the ID of high-load VMs and second array stores the low-load VMs information. Low load and high load are measured by average thresh value (Ath). Average thresh value is a maximum queue length of an individual VM. VMs can be distributed according to high-load, low-load, and remaining (average load) VMs store in average array.

Algorithm 1: Virtual machine Identification

```

1: V - Total number of VMs in single host
2: For each VM [ID] -  $VM \in V$  // Assign unique ID to VMs
3: VM [ID] load == Ath (Thresh value of VM) //Check
the current load from available thrash
    if VM [ID] > Ath
        Send VM [ID] to Hth [array] //Store in high
load array
    else if VM [ID] < Ath
        Send VM [ID] to Lth [array] // Store in low
load array
    else
go to Ath array // store in maximum trash array
4. end for

```

Second algorithm specifies the load distribution form high-load to low-load VM. Algorithm can find the high-load VM from high [array] and low-load VM from low [array]. It checks for the transfer condition, if transfer condition is true then the load is transferred from high load to low load. If low-load VM queue length is full then high-load VM load will be transferred to another low-load VM. High-load VM load will come equal to Atv then high-load VM will be transferred to average array exit, if high-load VM current load is low then the VM moves to low array and then exits.

Algorithm 2: Load Balance Management

```

1: Find the VM form Hth > Ath
2: Find the VM from Lth
3: [ok = Lth [VM] < Ath < Hth [VM] /* lock the both
(Hth[VM] and Lth[VM], Check the condition and
continue transfer the load */
    if (current Lth(VM) == Ath) // during the
load transfer
        Then Lth[VM] move to Ath array
        go to step 2 // take the VM form Lth array
        repeat step 3 //
if Hth [VM] == Ath
    Move current Hth [VM] to Ath
else
    go to Lth array
4: repeat step 1 to 3
5: end

```


Algorithm manages low-load and high-load VMs. Meantime of response will be shorter and effective. Load balance must be flexible, performance-based resources utilization. Thresh valve of VM defines the high capacity of VM and does the help to find low-load and high-load VMs.

6 Conclusion

Traditional computing systems use cluster and grid computing based resource management mechanism. Cloud service provides virtualized distributed resource management policy. Good resource management policies maximize the resource utilization. Resource management policy must be able to scale the available resource and users demands. In this paper we have discussed resource management goals, VM migration policies, challenges of resource management, and load balance concept and management. This paper has proposed an effective load balance algorithm. This algorithm has proposed a way to find overloaded VMs and underloaded VMs with load transfer mechanism. Researchers have done lots of work in load balance, VM migration, and server consolidation with migration policy, but still there is a need for intelligent live migration mechanism.

References

1. Mishra, Mayank, Anwesha Das, Purushottam Kulkarni, and Anirudha Sahoo. "Dynamic resource management using virtual machine migrations." *Communications Magazine*, IEEE 50, no. 9 (2012): 34–40.
2. Ahmad, Raja Wasim, Abdullah Gani, Siti Hafizah Ab Hamid, Muhammad Shiraz, Abdullah Yousafzai, and Feng Xia. "A survey on virtual machine migration and server consolidation frameworks for cloud data centers." *Journal of Network and Computer Applications* 52 (2015): 11–25.
3. Hussain, Hameed, Saif Ur Rehman Malik, Abdul Hameed, Samee Ullah Khan, Gage Bickler, Nasro Min-Allah, Muhammad Bilal Qureshi et al. "A survey on resource allocation in high performance distributed computing systems." *Parallel Computing* 39, no. 11 (2013): 709–736.
4. Ferreto, Tiago C., Marco AS Netto, Rodrigo N. Calheiros, and César AF De Rose. "Server consolidation with migration control for virtualized data centers." *Future Generation Computer Systems* 27, no. 8 (2011): 1027–1034.
5. Beloglazov, Anton, and Rajkumar Buyya. "Adaptive threshold-based approach for energy-efficient consolidation of virtual machines in cloud data centers." In *Proceedings of the 8th International Workshop on Middleware for Grids, Clouds and e-Science*, p. 4. ACM, 2010.
6. Hu, Liang, Jia Zhao, Gaochao Xu, Yan Ding, and Jianfeng Chu. "HMDC: Live virtual machine migration based on hybrid memory copy and delta compression." *Appl. Math* 7, no. 2L (2013): 639–646.
7. Medina, Violeta, and Juan Manuel García. "A survey of migration mechanisms of virtual machines." *ACM Computing Surveys (CSUR)* 46, no. 3 (2014): 30.
8. Beloglazov, Anton, and Rajkumar Buyya. "Energy efficient resource management in virtualized cloud data centers." In *Proceedings of the 2010 10th IEEE/ACM International*

- Conference on Cluster, Cloud and Grid Computing, pp. 826-831. IEEE Computer Society, 2010.
9. Zhang, Yuan, Xiaoming Fu, and K. K. Ramakrishnan. "Fine-Grained Multi-Resource Scheduling in Cloud Datacenters." In The 20th IEEE International Workshop on Local and Metropolitan Area Networks (LANMAN 2014). 2014.
 10. Beloglazov, Anton, Jemal Abawajy, and Rajkumar Buyya. "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing." *Future generation computer systems* 28, no. 5 (2012): 755–768.
 11. Gupta, Diwaker, Ludmila Cherkasova, Rob Gardner, and Amin Vahdat. "Enforcing performance isolation across virtual machines in Xen." In *Middleware 2006*, pp. 342–362. Springer Berlin Heidelberg, 2006.
 12. Nathan, Senthil, Purushottam Kulkarni, and Umesh Bellur. "Resource availability based performance benchmarking of virtual machine migrations." In *Proceedings of the 4th ACM/SPEC International Conference on Performance Engineering*, pp. 387–398. ACM, 2013.
 13. Isci, Canturk, Jiuxing Liu, Bülent Abali, Jeffrey O. Kephart, and Jack Kouloheris. "Improving server utilization using fast virtual machine migration." *IBM Journal of Research and Development* 55, no. 6 (2011): 4–1.
 14. https://pubs.vmware.com/vsphere-50/index.jsp#com.vmware.vsphere.vm_admin.doc_50/GUID-E19DA34B-B227-44EE-B1AB-46B826459442.html, 2015.
 15. Garg, Saurabh Kumar, Adel Nadjaran Toosi, Srinivasa K. Gopalaiyengar, and Rajkumar Buyya. "SLA-based virtual machine management for heterogeneous workloads in a cloud datacenter." *Journal of Network and Computer Applications* 45 (2014): 108–120.
 16. https://pubs.vmware.com/vsphere-50/index.jsp#com.vmware.vsphere.vm_admin.doc_50/GUID-E19DA34B-B227-44EE-B1AB-46B826459442.html, 2015.
 17. Li, Kangkang, Huanyang Zheng, Jie Wu, and Xiaojiang Du. "Virtual machine placement in cloud systems through migration process." *International Journal of Parallel, Emergent and Distributed Systems* ahead-of-print (2014): 1–18.