International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A Survey on Performance and Energy Management in Cloud Computing

Pratima D. Nerkar[1], Suresh B. Rathod[2]

[1]Sinhagad Academy of Engineering, Pune University, Kondhwa, Pune, Maharashtra, India

Abstract: ***Cloud computing is the latest distributed computing paradigm and it offers tremendous opportunities to solve large-scale scientific problems. Energy awareness and real time service management are big challenges in cloud datacenter. Reducing Energy consumption has been an essential technique for Cloud resources and data centers not only to decrease operating costs, but also to improve the system reliability. Many techniques such as Scheduling, live migration, and resource allocation have been proposed to reduce energy consumption. Concepts like DVFS, Dynamic MIPS adjustment and live migration contributed to reduction in energy consumption. Several schemes reduce power consumption by real-time services, and suggest power-aware profitable provisioning of real-time services.***

Keywords: Cloud computing, DVFS, live migration, Dynamic MIPS.

1. Introduction

Cloud computing refers to applications and services that run on a distributed network using virtualized resources and accessed by common Internet protocols and networking standards[2]. It is distinguished by the notion that resources are virtual and limitless and that details of the physical systems on which software runs are abstracted from the user**.** Cloud computing is nothing but a specific style of computing where everything from computing power to infrastructure, business apps are provided "as a service". Cloud computing is not a new technology [4]. It's just a way of using old services effectively. According to definition of cloud and its application has increased its demand day-by-day. With this increase in demand cloud computing has become favourite computing today. With more and more suppliers began offering cloud computing services, these services are convenient to users but consuming a lot of energy. Thus, how to save the energy of the data center without affecting both the economic efficiency and system performance is an important issue. Resource provisioning [14] in minutes, on demand scalability [5], pay per use these features have motivated the users to increase demand and which has lead to the increase in resources in data center in order to fulfill these demands. Data centers consume from 10 to 100 times more energy per square foot than typical office buildings. They can even consume as much electricity as a city. The main part of power consumption [11] in data centers comes from computation processing, disk storage, network, and cooling systems. Lowering the energy usage of data centers becomes a challenging issue because computing applications and data are growing so quickly that increasingly larger servers and disks are needed to process them fast enough within the required time period. Thus, data center resources need to be managed in an energy-efficient [12] manner to drive Green Cloud computing. The pay-as-you-go mechanism in Cloud computing assures Service Level Agreements (SLAs) [6] between customers and Cloud providers. SLAs specify the negotiated agreements on the Quality of Service (QoS), such as deadline constraints. Thus, data centers [15] must minimize power consumption without violating the SLAs [1]. As many applications require deadline constraints, this paper focuses on energy-aware management of real-time Cloud services, such financial analysis, real-time distributed databases, etc. One of big challenges in data centers is to manage system resources in a energy-efficient way. Virtualization technology can simulate a variety of different platforms and manage the resources of the system. By applying the virtualization technology, in accordance with the requirements of the users to configure a virtual machine, both the computing environment and resource management problems can be solved. Thus many techniques have been proposed such as Scheduling technique [1], virtual machine migration [1]and resource allocation[1], in order to reduce the CPU utilization and energy consumption.

2. Related Work

2.1 Dynamic Voltage Frequency Scaling (DVFS)

Dynamic Voltage and Frequency Scaling (DVFS) [1] is an essential part of controlling the power consumption of any computer system, ranging from mobile phones to servers. DVFS efficiency relies on hardware-software co-optimization, thus using existing hardware cannot reveal the full optimization potential beyond the current implementation's characteristics Dynamic Voltage and Frequency Scaling (DVFS) is one of the most important techniques for managing the dynamic power consumption of a system. From mobile devices to large data centers, scaling frequency down when it is not critical for performance also allows voltage to be scaled down. Due to the dependence between power consumption and voltage frequency, this technique can provide significant energy savings, with limited or no performance loss dynamically adjust the voltage and frequency of the processor in execution time without having to restart the power supply, system voltage and frequency can be adjusted. Power saving technology by reducing the voltage supply CPU voltage can be lowered, but the execution speed of the work will be reduced Three power-aware VM provisioning schemes:

1)Lowest-DVFS for VM Provisioning: Adjusts the processor speed to the lowest level at which HRT-VMs meet their deadlines.

2) δ-Advanced-DVFS for VM Provisioning It operates the processor speed δ% faster in order to increase the possibility of accepting incoming HRT-VM requests.
3) Adaptive-DVFS for VM Provisioning Adaptive-DVFS scheme manages the average arrival rate, the average service rate, and the average deadline for the last service request. It adjusts the processor scale.

2.2 Dynamic Energy generation

DVFS [13] is able to reduce the power consumption of a CMOS integrated circuit, such as a modern computer processor, by reducing the frequency at which it operates, as shown by

$$P = C*f*V^2 + S_{static} \quad (1)$$

Where C is the capacitance of the transistor gates (which depends on feature size), f is the operating frequency and V is the supply voltage. The voltage required for stable migration operation is determined by the frequency at which the circuit is clocked, and can be reduced if the frequency is also reduced. This can yield a significant reduction in power consumption because of the relationship shown above equation (1).

2.3 Virtualization

Cloud computing virtualizes systems by pooling and sharing resources as shown in Figure [1]. Systems and storage can be provisioned as needed from a centralized infrastructure, costs are assessed on a metered basis, multi-tenancy is enabled, and resources are scalable with agility. Virtualization has gained traction in a wide variety of contexts. The rise of Cloud Computing and the wide adoption of the Open Flow API in computer networks are just a few examples of how virtualization has changed the foundations of computing. In general, the term "virtualization"[9] refers to the process of turning a hardware-bound entity into a software-based component. The end result of such procedure encapsulates an entity's logic and is given the name of Virtual Machine (VM) [16]. The main advantage of this technique is that multiple VMs can run on top of a single physical host, which can make resource utilization much more efficient. Of particular interest are those VMs with high availability requirements, such as the ones deployed by cloud providers, given that they generate the need to minimize the down time associated with routine operations.

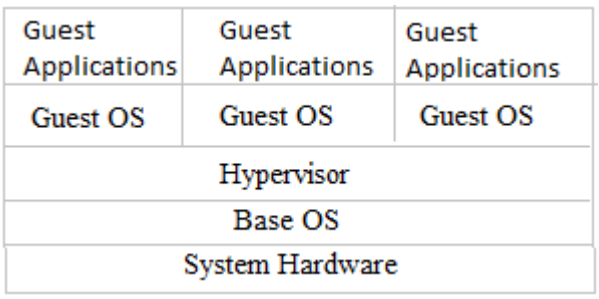

Figure 1: Components of Server Virtualization

2.4 Virtual machine migration

Virtual machine migration [8] enables load balancing, hot spot mitigation and server consolidation in virtualized environments. Live VM migration can be of two types - adaptive, in which the rate of page transfer adapts to virtual machine behavior (mainly page dirty rate), and non-adaptive, in which the VM pages are transferred at a maximum possible network rate. In either method, migration requires a significant amount of CPU and network resources, which can seriously impact the performance of both the VM being migrated as well as other VMs. This calls for building a good understanding of the performance of migration itself and the resource needs of migration. Such an understanding can help select the appropriate VMs for migration while at the same time allocating the appropriate amount of resources for migration. While several empirical studies exist, a comprehensive evaluation of migration techniques with resource availability constraints is missing. technique to employ under a given set of conditions. In this work, we conduct a comprehensive empirical study to understand the sensitivity of migration performance to resource availability and other system parameters (like page dirty rate and VM size). The study reveals several shortcomings of the migration process. We also quantified the impact of migration on the performance of applications running on the migrating VM and other co-located VMs. It allows a virtual machine running on a physical machine to be migrated to another physical machine .It is classified into Offline Before migration, current user's state suspended or shut down Users cannot take any action**.** In Real-time migration it is not necessary to shut down the original virtual machine task can be migrated at the user unaware situation. The advantages of the real-time migration include load balancing, power efficiency, and convenient maintenance.

2.5 Virtual Machine Monitor (VMM) / Hypervisor

The software component called Hypervisor [3] allows multiple operating systems to share a single hardware host. It is an abstraction layer between host machine hardware and virtual machine OS. (Guest OS). Each guest OS appears to have the host's processor, memory, and other resources all to itself. However, the hypervisor is actually controlling the

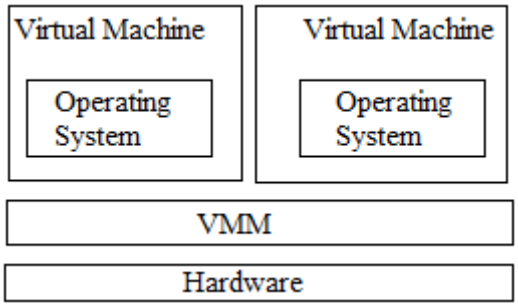

Figure 2: VMM Architecture

host processor and resources, allocating what is needed to each operating system in turn and making sure that the guest operating systems (called virtual machines) cannot disrupt each other as shown in Figure [2].

There are two types of hypervisors:
Type 1 (or native, bare metal) hypervisors as shown in Fig. [3] run directly on the host's hardware to control the hardware and to manage guest operating systems. A guest operating system thus runs on another level above the hypervisor. Examples - Oracle VM Server for SPARC, the Citrix XenServer, KVM, VMware ESX/ESXi, and Microsoft Hyper-V hypervisor.

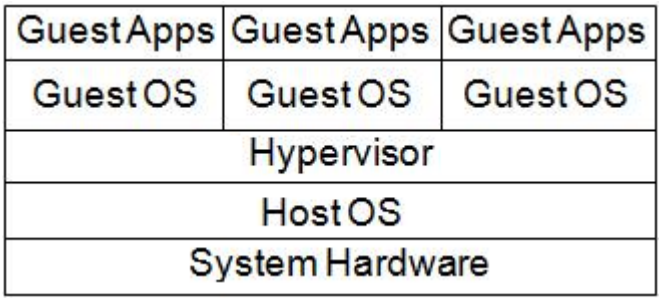

Figure 3: Type 1

Type 2 (or hosted) hypervisors refer Figure 3 run within a conventional operating system environment. With the hypervisor layer as a distinct second software level, guest operating systems run at the third level above the hardware. Examples – Vmware [3] Workstation and VirtualBox

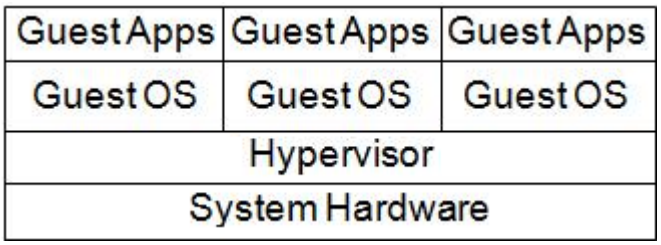

Figure 4: Type 2

2.6 Virtualization Techniques

With the problems and solutions mentioned in the previous section in mind, we now take a look at two techniques to realize system virtualization [17],[9] Figure 5.

2.6.1 Paravirtualization (Type - I)

The paravirtualization approach allows each guest to run a full operating system. But these do not run in ring 0. Due to that all the privileged instructions can't be executed by a guest. In order of that, modifications to the guest operating systems are required to implement an interface. This is used by the VMM to take over control and handle the restricted instructions for the VM. The paravirtualization approach promises nearly to native performance but lacks in the support for closed source operating systems. To apply the mentioned modifications, the source code of the kernel of an operating system has to be patched. Thus, running Microsoft Windows in a VM is impossible using paravirtualization [9].

2.6.2 Fullvirtualization (Type – II)

This approach allows to operate several operating systems on top of a hosting system, each running into its own isolated VM. The VMM uses hardware support as described to operate these, which allows to run the guest operating systems without modifications. The VMM provides I/O devices for each VM, which is commonly done by emulating older hardware. This ensures that a guest OS has driver support for these devices. Because of the emulated parts fullvirtualization is not as fast as paravirtualization. But if one needs to run a closed source OSs, it is the only viable technique to do so.

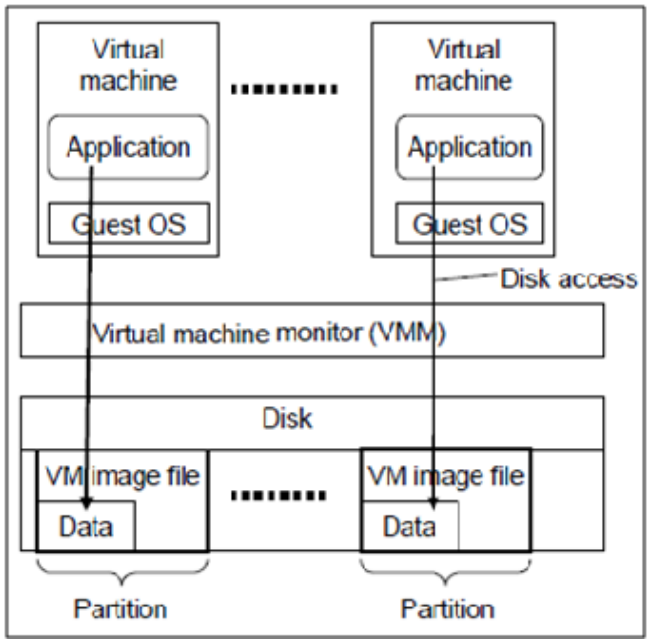

Figure 5: Virtualization Technique

2.7 Real time service

The steps for a real-time service are as follows Figure 6.

- Requesting a virtual platform
- Generating a RT-VM [11] from real-time applications
- Requesting a real-time VM
- Mapping physical processors
- Executing the real-time applications

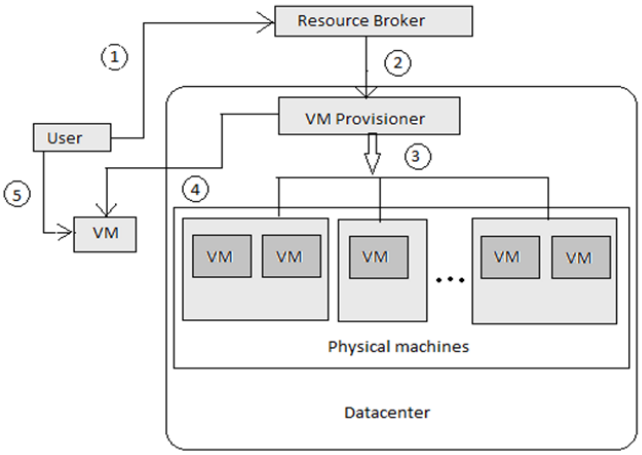

Figure 6: Real time service framework

2.8 Dynamic MIPS Adjustment

CPU frequency and MIPS [1] directly contribute to the dynamic energy consumption In order to reduce energy consumption we have to calculate the new CPU utilization rate of Vitual machine. It can be calculated in following way:
.new CPU utilization in MIPS of VM =Upper bound MIPS – Current CPU utilization in MIPS

Using above we can calculate the MIPS rate of specific VM as follows:
new MIPS rate to specific VM =(new CPU utilization +(remaining MIPS of current service/remaining time of deadline at time t))/ current allocated MIPS rate for the service on VM.

2.9 Live Migration

Live migration [10] ,[7] moves running virtual machines from one physical server to another with no impact on virtual machine availability to users. By pre-copying the memory of the migrating virtual machine to the destination server, live migration minimizes the transfer time of the virtual machine. A live migration is deterministic, which means that the administrator, or script, that initiates the live migration determines which computer is used as the destination for the live migration. The guest operating system of the migrating virtual machine is not aware that the migration is happening, so no special configuration for the guest operating system is needed. After initiating a live migration, the following process occurs:
Live migration setup occurs.

- Memory pages are transferred from the source node to the destination node.
- Modified pages are transferred.
- The storage handle is moved from the source server to the destination server.
- The virtual machine is brought online on the destination server.
- Network cleanup occurs.

3. Conclusion

In this paper ,with the combination of dynamic MIPS adjacement and DVFS concept along with live migration contribute to reduce energy consumption .In future some reallocation algorithm for CPU can be formed to reduce CPU energy consumption.

References

[1] Worachat Chawarut, Lilakiatsakun Woraphon," Energy-Aware and Real-time Service Management In Cloud Computing,"IEEE,2013
[2] "Cloud computing," http://www.ibm.com/cloud-computing/us/en/whatis-cloud computing.html.
[3] Zhiming Shen, Zhe Zhang, Andrzej Kochut, Alexei Karve, Han Chen, Minkyong Kim Hui Lei, Nicholas Fuller, "VMAR: Optimizing I/O Performance and Resource Utilization in the Cloud".
[4] "Cloud Computing," http://en.wikipedia.org/wiki/Cloud computing.
[5] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. "Above the Clouds: A Berkeley View of Cloud Computing," Technical Report EECS-2009-28, EECS Department, Univ.of California, Berkeley, 2009.
[6] K. H. Kim, W. Y. Lee, J. Kim, R. Buyya. "SLA-Based Scheduling of Bag-of- Tasks Applications on Power-Aware Cluster Systems," IEICE Transactions on Information and Systems, Issue 12, pp. 3194-3201, 2010.
[7] P. Padala, "Understanding live migration of virtual machines." Available: http://tinyurl.com/24bdaza, Jun. 2010.
[8] D. Breitgand, G. Kutiel, and D. Raz, "Cost-aware live migration of services in the cloud," in 2011 USENIX Workshop on Hot Topics in Management of Internet, Cloud, and Enterprise Networks and Services.
[9] VMWare, "VMmark Virtualization Benchmarks," http://www.vmware.com/products/vmmark/, Jan. 2010.
[10] S.Hacking and B. Hudzia, "Improving the live migration process of large enterprise applications," inProceedings 2009 International Workshop on Virtualization Technologies in Distributed Computing.
[11] K. H. Kim, A. Beloglazov and R. Buyya. "Power-aware provisioning of virtual machines for real-time Cloud services," Concurrency and Computation: Practice & Experience archive Volume 23 Issue 13, pp.1491-1505 2011
[12] A. Beloglazov, R. Buyya. "Energy Efficient Allocation of Virtual Machines in Cloud Data Centers," 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, pp. 577-578, 2010.
[13] K. H. Kim, A. Beloglazov and R. Buyya, "Power-aware provisioning of Cloud resources for real-time services," MGC '09 Proceedings of the 7th International Workshop on Middleware for Grids, Clouds and e- Science. ACM. 2009.
[14] L. T. Lee, K. Y. Liu and H. Y. Huang, "Dynamic resource management for energy saving in the cloud computing environment," Second International Symposium on Information and Automation (ISIA 2012), 2012.
[15] Carlo Mastroianni, Michela Meo, and Giuseppe Papuzzo," Probabilistic Consolidation of Virtual Machines in Self-Organizing Cloud Data Centers," IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 1, NO. 2, JULY-DECEMBER 2013
[16] Mladen A. Vouk," Cloud Computing – Issues, Research and Implementations," Journal of Computing and Information Technology - CIT 16, 2008, 4, 235–246
[17] T.Swathi, K.Srikanth, S. Raghunath Reddy," VIRTUALIZATION IN CLOUD COMPUTING," IJCSMC, Vol. 3, Issue. 5, May 2014, pg.540 – 546

Author Profile

Miss Pratima Devidas Nerkar completed B.E. in Computer Engineering in 2011 from K.K.Wagh C.O.E, Nashik.

Prof. Suresh Baliram Rathod completed B.E. in Information Technology in 2007 from STBCE; Tuljapur.He has also completed his M.E. from SCOE, Pune, Maharashtra, India