Volume 4, Issue 12, December 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Green Computing in Dynamic Resource Allocation for Cloud Environment

S.M. Gayathri, K. Malathi, N.G. Bhuvaneswari Amma
Department of CSE & IIIT
Srirangam, India

Abstract*— Cloud computing allows business customers to scale up and down their resource usage based on their needs. Many of the touted gains in the cloud model come from resource multiplexing through virtualization technology. In this paper, a system that uses virtualization technology to allocate data center resources dynamically based on application demands and support green computing by optimizing the number of servers in use. The concept of "skewness" to measure the unevenness in the multi-dimensional resource utilization of a server is used. By minimizing skewness, different types of workloads can combine nicely and improve the overall utilization of server resources. A set of heuristics is developed which prevent overloading in the system effectively while saving the energy. Trace driven experiment results and demonstration of algorithm with good performance has described in the paper.

Keywords*—cloud computing, green computing, skewness, virtual machine(VM), physical machine(PM)

I. INTRODUCTION

Cloud computing is computing in which large groups of remote servers are networked to allow centralized data storage and online access to computer services or resources. There are several service models in it like: Infrastructure as a service (IaaS), Platform as a service (PaaS), Software as a service (SaaS), Unified Communications as a Service

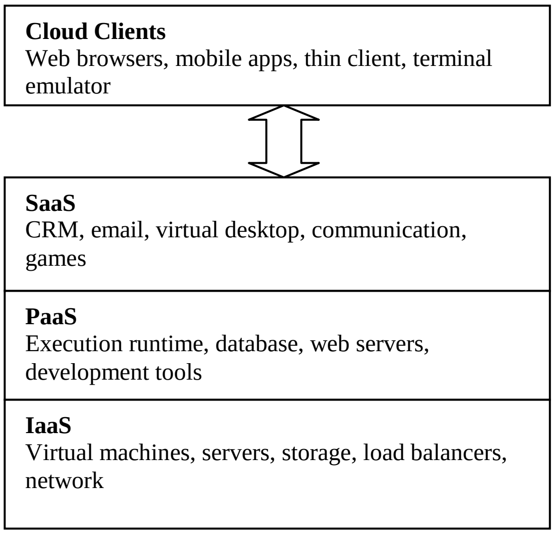

Figure 1. Service Models in Cloud

The elasticity and the lack of upfront capital investment offered by cloud computing is appealing to many businesses. Lot of discussion on the benefits and costs of the cloud model and on how to move legacy applications onto the cloud platform. The cloud service provider make the best multiplex on its virtual resources onto the physical hardware and it is important because much of the touted gains in the cloud model come from such multiplexing. Analysis have found that servers in many existing data centers are often severely under-utilized due to over-provisioning for the peak demand. The model of cloud is expected to make such practice unnecessary by offering automatic scale up and down in response to load variation. It reduces the hardware cost, it also saves on electricity which contributes to a significant portion of the operational expenses in large data centers.

Cloud computing relies on sharing of resources to achieve coherence and economies of scale, similar to a utility (like the electricity grid) over a network. Foundation of cloud computing is the broader concept of converged infrastructure and shared services[10]. In a simpler way "the cloud", also focuses on maximizing the effectiveness of the shared resources. The resources of cloud are usually not only shared by multiple users but are also dynamically reallocated per demand. These work is for allocating resources to users. For an instance, a cloud computer facility that serves European users during European business hours with a specific application (e.g., email) may reallocate the same resources to serve

North American users during North America's business hours with a different application (e.g., a web server). This approach should maximize the use of computing power thus reducing environmental damage also because less power, air conditioning, rackspace, etc., are required for various functions. By using cloud computing, multiple users can access a single server to retrieve and update their data without purchasing licenses for different applications.

The present availability of high-capacity networks, low-cost computers and storage devices as well as the widespread adoption of hardware virtualization, service-oriented architecture, and autonomic and utility computing have led to a growth in cloud computing. Cloud vendors are experiencing growth rates of 50% per annum.

II. RELATED WORK

Developers with innovative ideas for new Internet services no longer require the large capital outlaysin hardware to deploy their service or the human expense to operate it. Cloud Computing refers to both the applications delivered as services [1] over the Internet and the hardware and systems software in the datacenters that provide these services. The services have been referred to as Software as a Service (SaaS). The datacenter hardware and software is known as Cloud. When a Cloud is made available in a pay-as-you-go manner to general public, making it a Public Cloud; the service being sold is Utility Computing. It is used the term Private Cloud to refer to internal datacenters of a business or other organization which is not made available to the general public. Hence,By this Cloud Computing is the sum of SaaS and Utility Computing, but it does not include Private Clouds. People can also be users or providers of SaaS, or users and cloud providers of Utility Computing. SaaS Providers (Cloud Users) and Cloud Providers,have received less attention than SaaS Users are focused.From a hardware point of view, three aspects are new in Cloud Computing: (1) Illusion of the infinite computing resources are available on demand, by eliminating the need for Cloud Computing users to plan far ahead for provisioning. (2) Elimination of an upfront commitment by the Cloud users, thereby allowing companies to start small and increase hardware resources only when there is an increase in their needs. (3) Ability to pay for the use of computing resources on a short-term basis as needed (e.g., processors by the hour and storage by the day) and release them as needed, and hence rewarding conservation by letting machines and storage go when they are no longer useful.

The author in [2] uses feedback control theory, for adaptive management of virtualized resources, which is based on VM. In this VM-based architecture all hardware resources are pooled into common shared space in cloud computing infrastructure so that hosted application can access the required resources as per there need to meet Service Level Objective (SLOs) of application. The adaptive manager use in this architecture is multi-input multi-output (MIMO) resource manager, which includes 3 controllers and they are : CPU controller, memory controller and Input-output controller, its goal is to regulate multiple virtualized resources utilization to achieve SLOs of application by using control inputs per-VM (virtualization machine) CPU, memory and I/O allocation.

Xen is Open source virtualization software that is used to partition workstations and servers into separate virtual machines, each contains its own copy of an OS. It is developed at the University of Cambridge in the U.K. Xen is well-known [3] for its quick response and low overhead. Xen is a small, low-level "hypervisor," which is the first control software loaded when the computer starts. Virtual machine monitors (VMMs) like Xen provide a mechanism for mapping virtual machines (VMs) to physical resources .This mapping is largely hidden from the cloud users. Users with the Amazon EC2 service for do not know where their VM instances run. It is up to the cloud provider to make it sure the underlying physical machines (PMs) have sufficient resources to meet their needs. Live migration technology of VM makes it possible to change the mapping between VMs and PMs when the applications are in running state. However, a policy issue remains as how to decide the mapping adaptively so that the resource demands of VMs are met while the number of PMs used is minimized. This becomes challenging when the resource needs of VMs are heterogeneous due to the diverse set of applications they run and vary with time as the workloads grow and shrink accordingly. The capacity of PMs can also be heterogeneous because multiple generations of hardware have been co-existing in a data center.

The xend daemon (xend), which stores configuration information about each virtual machine and controls how virtual machines are created and managed.A modified version of QEMU, an open-source software program that emulates a full computer system, which includes a processor and also various peripherals which provide the ability to host operating systems in full virtualization mode.It is also referred to as a VM Guest or DomU consists of the following components: minimum of one virtual disk that contains a bootable operating system. The virtual disk can be based on file, partition, volume, or other type of block device.Virtual machine configuration information can be modified by exporting a text-based configuration file from xend or through Virtual Machine Manager.It is a combination of GUI tools, commands, and configuration files to help you manage and customize your virtualization environment. Its limitations are: (1)A policy issue remains as how to decide the mapping adaptively so that the resource demands of VMs are met while the number of PMs used is minimized.(2) No control over the business assets (data!). The main assets in every company are its data files with valuable customer information.(3) Risk of data loss due to improper backups or system failure in the virtualized environment. (4) High cost and loss of control.

Virtual machine monitors [4] provide a mechanism for mapping virtual machines (VMs) to physical resources. Users with the Amazon EC2 service do not know where their VM instances run. The cloud provider should make sure the underlying physical machines (PMs) have sufficient resources to meet their needs. Live migration technology of VM makes it possible to change the mapping between VMs and PMs while applications are running.

Design and implementation of a system that uses virtual machine technology to provide fast, transparent application migration [5]. The system that can migrate unmodified applications on unmodified mainstream Intel x86-based operating system, together with Microsoft Windows, Linux, Novell. Neither the application nor any clients communicating with the application can tell that the application has been migrated.Measurements of workloads, application downtime caused by migration is less than a second.(1)It describes the first system to provide transparent virtual machine migration of existing applications and operating systems; neither the applications nor the operating systems need to be modified. (2) It is the first paper to provide performance measurements of hundreds of virtual machine migrations of concurrently running virtual machines with standard industry benchmarks.(3) It characterizes the overheads and resources required during virtual machine migration.

McNett et al.[6] introduced extensible framework Usher, is a virtual machine management system designed to impose few constraints upon the computing environment under its management. Usher makes administrators to choose how their virtual machine environment will be configured and the policies under which they can be managed. The design allows for alternate implementations for authentication, authorization,handling of infrastructure, Scheduling of virtual machine sand logging. This design is to provide an interface whereby users and administrators can request virtual machine operations while delegating administrative tasks for these operations to modular plugins. Implementation of Usher allows for arbitrary action for any event in the system. Can be used to manage virtual clusters at two locations under very different settings, to demonstrate the flexibility of Usher to meet different virtual machine management requirements.

Virtualization concept in [7] provide significant benefits in data centers by enabling virtual machine migration to eliminate hotspots. Waldspurger [8] progress a frame work VMware ESX Server is a thin software layer designed to multiplex hardware resources efficiently among virtual machinesrunning unmodified commodity operating systems. Introduces several novel ESX Server mechanisms and policies for managing memory. And in another algorithm proposed by Chen et al. [9] can save a significant amount of energy without sacrificing user experiences. Dynamic server provisioning techniques are effective inturning off unnecessary servers to save energy. Such techniques, mostly for request-response services,can face challenges in the connection servers that host a large number of long-lived TCP connections.

III. PROPOSED SYSTEM

In this paper the design and implementation of an automated resource management system that achieves a good balance between the two goals. i) Overload avoidance: The capacity of a PM should be sufficient to satisfy the resource needs of all VMs running on it. Otherwise, the PM is overloaded and can lead to degraded performance of its VMs. ii) Green computing: The number of PMs used should be minimized as long as they can still satisfy the needs of all VMs. Idle PMs can be turned off to save energy. The development of a resource allocation system can avoid overload in the system effectively by minimizing the number of servers used. The skewness is used to measure the uneven utilization of a server. Minimizing skewness, can improve the overall utilization of servers in the face of multidimensional resource constraints. Cloudsim can be used for this implementation.

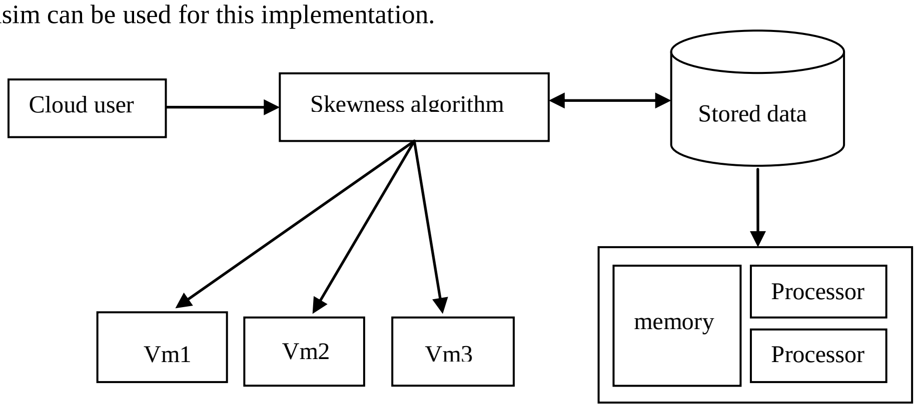

Figure 2. Block Diagram of Proposed System

To execute skewness in allocation of resources, the following steps are executed.

Algorithm used for executing skewness:

Step-1: Assume NP to be the number of servers i.e, NP = Numbers of the servers
And N be the resources i.e, N = Number of resources
Step-2: Let SK(NP) be the skewness value array
Step-3: For I = 1 to NP DO:
Step-4: For j=1 to P DO:
Step-5: Let R(J) = j th resource utilization
Step-6: SUM_sk = SUM_sk + (R(j)/average_ut-1)2
Step-7: Next j; (END LOOP)
Step-8: SK(I) = √(SUM_sk)
Step-9: Next I; (END LOOP)
Step-10: Find the biggest value in SK and store it in B
Step-11: B is the Server in as selected by skewness algorithm

$$\text{Skewness}(p) = \sqrt{\text{sum}(r_i/r^1)} \qquad \text{-Equation (1)}$$

Equation (1) is the formula used to calculate the required skewness for each resource and is stored is stored in a variable B according to the algorithm, which is then used by system in the allocation of resources. It optimizes the number of servers in use at a time. Let n be the number of PMs and m be the number of VMs in the system, respectively. The number of resources such as CPU, memory, I/O, network, etc., that should be considered is usually a small constant. Thus the calculation of the skewness and the temperature metrics for a single server takes a invariable amount of time. During load prediction, the FUSD algorithm should applied to each VM and each PM. The time complexity is O (n+m).The skewness algorithm consists of three parts such as load prediction, hot spot mitigation, and green computing.

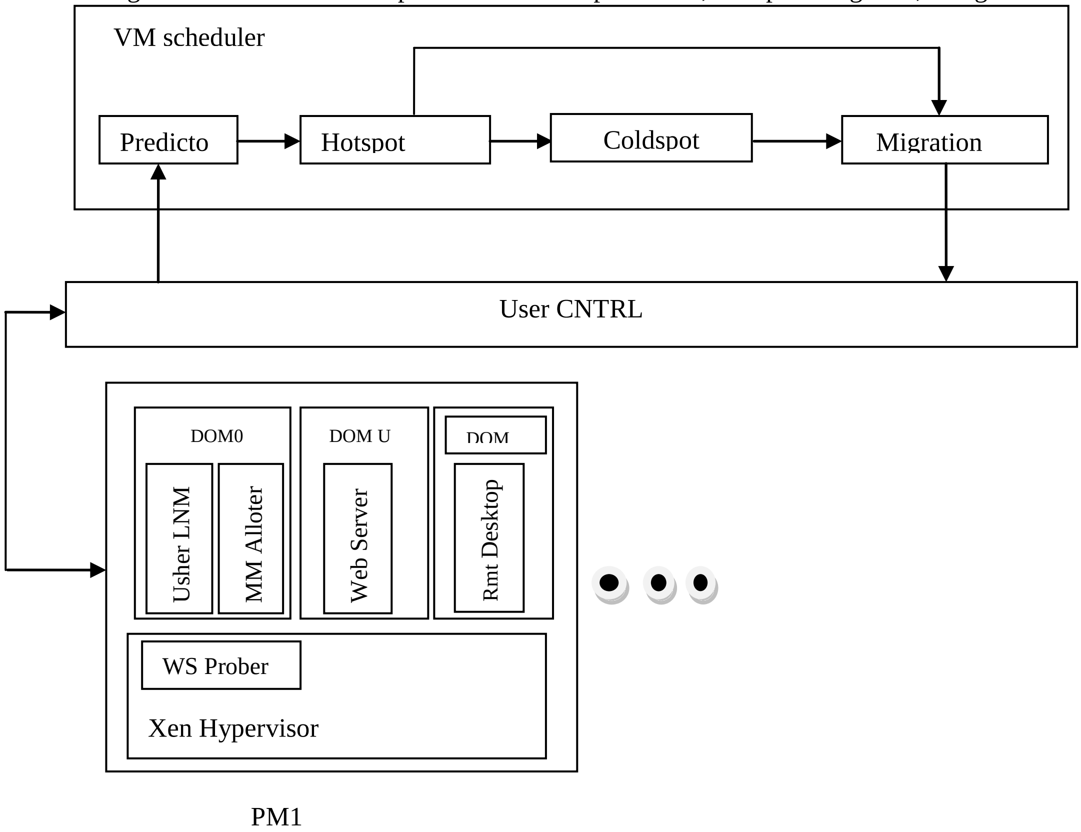

Figure 3. System Architecture

The advantages of this system are:

- A flexible, scalable infrastructure management platform has been architected and a prototype implemented
- Measurement of resource usage and end user activities lies hands of the cloud service provider.
- Opaque cost structure due to highly flexible usage of cloud services.
- Stable of cost structure.

IV. SIMULATION RESULT

The goal of the skewness algorithm is to mix workloads with different resource requirements together so that the overall utilization of server capacity is improved. In this experiment, the proposed algorithm handles a mixture of memory,CPU and network intensive workloads. Resource allocation status of three servers A, B, C has total memory allocated 500KB and resource used memory for serverA 80KB,serverB 170KB and serverC 80K. In Figure. 4 each cloud users provide cloud service Resource allocation in green computing.

Figure 4 : Host Site Information Page

In Figure.5 display Server usage and resource allocation status for user1 using Bar Chart. The cloud computing is a model which enables on demand network access to a shared pool computing resources. Cloud computing environment consists of multiple customers requesting for resources in a dynamic environment with their many possible constraints. The virtualization can be the solution for it and used to reduce power consumption by data centers.

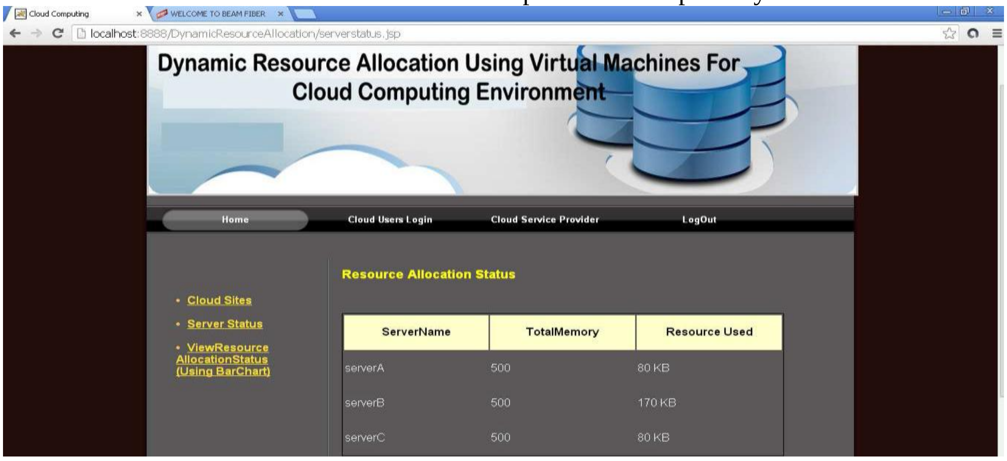

Figure 5 : Resource Allocation Status Page

The main purpose of the virtualization is that to make the most efficient use of available system resources, including energy. A data center, installing virtual infrastructure allows several operating systems and applications to run on a lesser number of servers, it can help to reduce the overall energy used for the data center and the energy consumed for its cooling. Once the number of servers is reduced, it also means that data center can reduce the building size as well. Some of the advantages of Virtualization which directly impacts efficiency and contributes to the environment include: Workload balancing across servers, Resource allocation and sharing are better monitored and managed and the Server utilization rates can be increased up to 80% as compared to initial 10-15%.

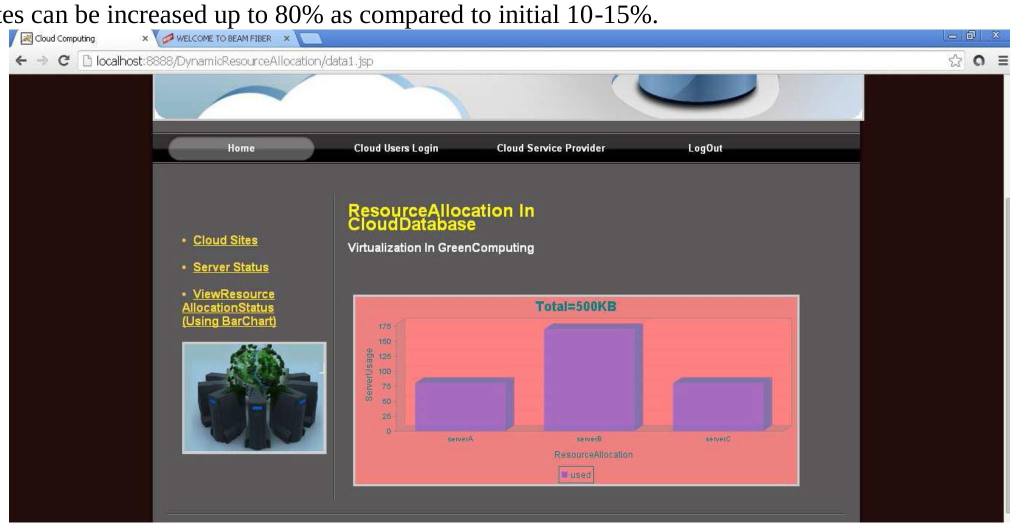

Figure 6: Resource allocation in database

V. Conclusion

Thus the design, implementation, and evaluation of a dynamic resource allocation system using green computing for cloud computing services have successfully implemented and explained in this paper. System multiplexes virtual to physical resources adaptively based on the changing demand. The skewness metric is used to combine VMs with different resource characteristics appropriately so that the capacities of servers are well utilized. The proposed algorithm achieves both overload avoidance and green computing for systems with multi-resource constraints.

References

[1] M. Armbrust et al., "Above the clouds: A berkeley view of cloud computing," University of California, Berkeley, Tech. Rep., Feb 2009.

[2] L. Siegele, "Let it rise: A special report on corporate IT," in The Economist, Oct. 2008.

[3] 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 Proc. of the ACM Symposium on Operating Systems Principles (SOSP'03), Oct. 2003.

[4] Amazon elastic compute cloud (Amazon EC2),http://aws.amazon.com/ec2/."

[5] M. Nelson, B.-H. Lim, and G. Hutchins, "Fast transparent migration for virtual machines," in Proc. of the USENIX Annual Technical Conference, 2005.

[6] M. McNett, D. Gupta, A. Vahdat, and G. M. Voelker, "Usher: An extensible framework for managing clusters of virtual machines," in Proc. of the Large Installation System Administration Conference (LISA'07), Nov. 2007.
[7] T. Wood, P. Shenoy, A. Venkataramani, and M. Yousif, "Black-box and gray-box strategies for virtual machine migration," in Proc. Of the Symposium on Networked Systems Design and Implementation (NSDI'07), Apr. 2007.
[8] C. A. Waldspurger, "Memory resource management in VMware ESX server," in Proc. of the symposium on Operating systems design and implementation (OSDI'02), Aug. 2002.
[9] G. Chen, H. Wenbo, J. Liu, S. Nath, L. Rigas, L. Xiao, and F. Zhao, "Energy-aware server provisioning and load dispatching for connection-intensive internet services," in Proc. of the USENIX Symposium on Networked Systems Design and Implementation
[10] Sumit Kumar, Sumeer Kumar, "Importance of cloud computing and its security Issues in Network Virtualization process", International Journal of Enhanced Reasearch in Management and Computer Applications, Vol. 3, September 2014, pp-(1-8).

