International Journal on Applications in Information and Communication Engineering
Volume 1: Issue 8: August 2015, pp 11-15. www.aetsjournal.com ISSN (Online) : 2394-6237

Effective Resource Allocation For Dynamic Workload In Virtual Machines Using Cloud Computing

J.Stalin , R.Kanniga Devi

Abstract— In cloud computing, the business class customers perform scale up and scale down of resources based on their needs. The resource multiplexing in cloud computing technology can be achieved better in virtualization technology. Here we present data center resources which uses virtualization technology based on the resource demands and optimize the number of servers we use. In this paper the concept of skewness is been introduced in order to measure the un evenness of resource utilization and load in the server. If the skewness then the overall utilization of server resources is improve. We develop heuristics sets in order to prevent overload in the system hence, the energy has been saved. By offering automated scale up and scale down in response to load variation, we reduce the hardware cost, operational expenses in large data centers and also save energy.

Index Terms— Multiplexing, Virtualization technology, Skewness, Resource utilization.

I. INTRODUCTION

Cloud computing refers to applications and services offered over the Internet. These services are offered from data centers all over the world, which collectively are referred to as the cloud. The idea of the cloud simplifies the many network connections and computer systems involved in online services. Any user with an Internet connection can access the cloud and the services it provides. Since these services are often connected, users can share information between multiple systems and with other users. Cloud computing also reduces the cost and improves the energy and management.

Several trends are opening up the era of Cloud Computing, which is an Internet-based development and use of computer technology. The ever cheaper and more powerful processors, together with the software as a service (SaaS) computing architecture are transforming data centers into pools of computing service on a huge scale. The increasing network bandwidth and reliable yet flexible network connections make it even possible that users can now subscribe high quality services from data and software that reside solely on remote datacenters. Moving data into the cloud offers great convenience to users since they don't have to care about the complexities of direct hardware management. The Cloud Computing vendors, Amazon Simple Storage Service (S3) and Amazon Elastic Compute Cloud (EC2) are both well-known examples. While these internet-based online services do provide huge amounts of storage space and customizable computing resources, this computing platform shift, however, is eliminating the responsibility of local machines for data maintenance at the same time.

As a result, users are at the mercy of their cloud service providers for the availability and integrity of their data. On the one hand, although the cloud infrastructures are much more powerful and reliable than personal computing devices, broad range of both internal and external threats for data integrity still exist. Examples of outages and data loss incidents of note worthy cloud storage services appear from time to time. Since users may not retain a local copy of outsourced data, there exist various incentives for cloud service providers (CSP) to behave unfaithfully towards the cloud users regarding the status of their outsourced data. However, the fact that users no longer have physical possession of data in the cloud prohibits the direct adoption of traditional cryptographic primitives for the purpose of data integrity protection. Hence, the verification of cloud storage correctness must be conducted without explicit knowledge of the whole data files. Meanwhile, cloud storage is not just a third party data warehouse. The data stored in the cloud may not only be accessed but also be frequently updated by the users, including insertion, deletion, modification, appending, etc.

Cloud computing is the delivery of computing as a service rather than a product, whereby shared resources, software, and information are provided to computers and other devices as a utility (like the electricity grid) over a network. Cloud computing provides computation, software, data access, and storage services that do not require end-user knowledge of the physical location and configuration of the system that delivers the services. Parallel to this concept can be drawn with the electricity grid, wherein end-users consume power without needing to understand the component devices or infrastructure required to provide the service.

Cloud computing has been changing how most people use the web and how they store their files. It's the structure that runs sites like Facebook, Amazon and Twitter and the core that allows us to take advantage of services like GoogleDocs and Gmail. But how does it work. Most websites and server-based applications run on particular computers or servers. What differentiates the cloud from the way those are setup is that the cloud utilizes the resources from the computers as a collective virtual computer, where the

J.Stalin ,PG Scholar, Department of Computer Science and Engineering, Kalasalingam University, India (Email: stalinmdu7@gmail.com)

R.Kanniga Devi ,Assistant Professor, Department of Computer Science and Engineering, Kalasalingam University, India(Email: rkannigadevi@gmail.com)

applications can run independently from particular computer or server configurations. They are basically floating around in a cloud of resources, making the hardware less important to how the applications work.

The selection of cloud deployment model depends on the different levels of security and control required. The Private cloud infrastructure is operated solely for a single organization with the purpose of securing services and infrastructure on a private network. This deployment model offer the greatest level of security and control, but it requires the operating organization to purchase and maintain the hardware and software infrastructure, which reduces the cost saving benefits of investing in a cloud infrastructure.

AWS is located in 9 geographical Regions- US East (Northern Virginia), US West (Northern California), US West (Oregon),AWS GovCloud (US) Region, São Paulo (Brazil), Ireland, Singapore, Tokyo and Sydney. There is also a GovCloud in the USA provided for US Government customers. Each Region is wholly contained within a single country and all of its data and services stay within the designated Region. Each Region has multiple Availability Zones, which are distinct data centers providing AWS services. Availability Zones are isolated from each other to prevent outages from spreading between Zones. Several services operate across Availability Zones (e.g. S3, DynamoDB) while others can be configured to replicate across Zones to spread demand and avoid downtime from failures.

II. PROBLEM STATEMENT

Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment

According to this paper [1], the 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. So that, to introduce the concept of skewness to measure the unevenness in the multidimensional resource utilization of a server. By minimizing skewness, users can combine different types of workloads nicely and improve the overall utilization of server resources. 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. It is up to the cloud provider to make sure the underlying physical machines (PMs) have sufficient resources to meet their needs. VM live migration technology makes it possible to change the mapping between virtual machines and physical machines while applications are running. To present the design and implementation of an automated resource management system that achieves a good balance between the two goals. The two goals are overload avoidance and green computing.

A. Xen and the Art of Virtualization

Xen's approach [2], is to target at hosting up to 100 virtual machine instances simultaneously on a modern server. The virtualization approach taken by Xen is extremely efficient. So that, to allow operating systems such as Linux and Windows XP to be hosted simultaneously for a negligible performance overhead. Modern computers are sufficiently powerful to use virtualization to present the illusion of many smaller virtual machines (VMs), each running a separate operating system instance. This has led to a resurgence of interest in VM technology. So that, to present Xen, a high performance resource-managed virtual machine monitor (VMM) which enables applications such as server consolidation, co-located hosting facilities, distributed web services, secure computing platforms and application mobility. Full virtualization was never part of the x 86 architectural designs. Certain supervisor instructions must be handled by the VMM for correct virtualization, but executing these with insufficient privilege fails silently rather than causing a convenient trap. Efficiently virtualizing the x86 MMU is also difficult. These problems can be solved, but only at the cost of increased complexity and reduced performance.

B. Live Migration of Virtual Machines

This paper presents [3] the migrating operating system instances across distinct physical hosts is a useful tool for administrators of data centers and clusters. It allows a clean separation between hardware and software, and facilitates fault management, load balancing and low-level system maintenance. Migrating an entire OS and all of its applications as one unit allows us to avoid many of the difficulties faced by process-level migration approaches. The narrow interface between a virtualized OS and the virtual machine monitor (VMM) makes it easy avoid the problem of residual dependencies in which the original host machine must remain available and network-accessible in order to service certain system calls or even memory accesses on behalf of migrated processes.

C. Memory Resource Management in Vmware ESX Server

VMware ESX Server [4], is a thin software layer designed to multiplex hardware resources efficiently among virtual machines running unmodified commodity operating systems. A ballooning technique reclaims the pages considered least valuable by the operating system running in a virtual machine for managing memory. The design of ESX Server differs significantly from VMware Workstation, which uses a hosted virtual machine architecture that takes advantage of a pre-existing operating system for portable I/O device support. ESX Server manages system hardware directly, providing significantly higher I/O performance and complete

control over resource management.ESX Server maintains a map data structure for each VM to translate physical page numbers (PPNs) to machine page numbers (MPNs). ESX Server supports over commitment of memory to facilitate a higher degree of server consolidation than would be possible with simple static partitioning.

III. APPROACH

To present the design and implementation of an automated resource management system that achieves a good balance between the two goals. The two goals are Overload avoidance and Green computing.

A. Overload avoidance

The capacity of a physical machine should be sufficient to satisfy the resource needs of all virtual machines running on it. Otherwise, the physical machine is overloaded and can lead to degraded performance of its virtual machines.

B. Green computing

The number of physical machines used should be minimized as long as they can still satisfy the needs of all virtual machines. Idle physical machines can be turned off to save energy.

To handle multiple objectives like power and performance during virtual machine placement by minimizing both under provisioning and over provisioning problems under the demand. The skewness algorithm achieves both overload avoidance and green computing for systems with multi-resource constraints and also capacities of servers are well utilized.

C. Skewness Algorithm

Skewness is the measure of uneven resource utilization of a server. Let *n* be the number of resources present in server and *ri* be the utilization of the *i*-th resource. We define the resource skewness of a server p as the following equation

$$\text{Skewness (p)} = \sqrt{\sum_{i=1}^{n} (\frac{r_i}{r-} - 1)2}$$

where *r* is the average utilization of all resources for server *p*. In practice, not all types of resources are performance critical and hence we only need to consider bottleneck resources in the above calculation. By minimizing the *skewness*, we can combine different types of workloads nicely and improve the overall utilization of server resources.

We define a server as a cold spot if the utilizations of all its resources are below a cold threshold. This indicates that the server is mostly idle and a potential candidate to turn off to save energy. However, we do so only when the average resource utilization of all actively used servers (i.e., APMs) in the system is below a green computing threshold. A server is actively used if it has at least one VM running. Otherwise, it is inactive. Hot spot Mitigation

To reduce the temperature of the hot spots to less or equal to warm threshold. The nodes in hot spots are sorted by quick sort in the descending order. VM with the highest temperature should be first migrated away. Destination is decided based on least cold node. After every migration the status of each node is updated. This procedure continues until all hot spots are eliminated. The VM which is removed from the identified hot spot can reduce the skewness of that server the most. For each VM in the list, if a destination server can be found to accommodate it then that server must not become a hot spot after accepting this VM. Among all such servers, we select one whose skewness can be reduced the most by accepting this VM. Note that this reduction can be negative which means we select the server whose skewness increases the least. If a destination server is found, then the VM can be migrated to that server and the predicted load of related servers was updated. Otherwise, move on to the next VM in the list and try to find a destination server for it. As long as a destination server was found for any of its VMs, it can be considered as this run of the algorithm a success and then move on to the next hot spot. Note that each run of the algorithm migrates away at most one VM from the overloaded server.

D. Cold spot Mitigation

To reduce the power consumption, the servers that are under-utilized are switched off. We sort the cold spots in ascending order and select the node with the least temperature. The VMs in these least loaded cold spot servers should be migrated away to another cold spot without raising the temperature of the destination above the warm threshold. If such cold spot destinations are not available, then move the load in the cold spot to a warm spot without raising the temperature of the destination above the warm threshold.

system design

E. System Architecture

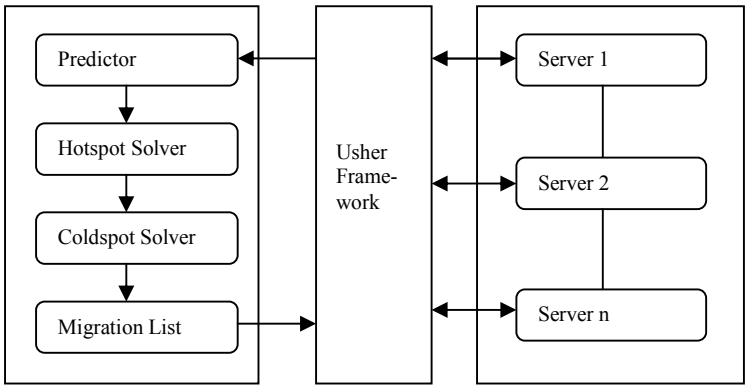

Fig.1 System Architecture

The major part of the project development sector considers and fully survey all the required needs for developing the project. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what

are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations. Generally algorithms shows a result for exploring a single thing that is either be a performance, or speed, or accuracy, and so on. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system. System architecture can comprise system components, the externally visible properties of those components, the relationships (e.g. the behavior) between them.

F. Modules

1. Virtual Machine Creation
2. Resource Allocation
3. Skewness Implementation
4. Energy Monitoring

1) Virtual Machine Creation

Virtualization in computing is the creation of a virtual (rather than actual) Version of something, such as a hardware platform, operating system, and a storage device or network resources.VM live migration is a widely used technique for dynamic resource allocation in a virtualized environment. The process of running two or more logical computer system so on one set of physical hardware. Dynamic placement of virtual servers is used to minimize SLA violations. When user creates a virtual machine, a cloud service is automatically created to contain the machine. User can create multiple virtual machines under the same cloud service to enable the virtual machines to communicate with each other, to load-balance between virtual machines and to maintain high availability of the machines. User can manage the availability of their application that uses multiple virtual machines by adding the machines to an availability set. Availability sets are directly related to fault domains and update domains. A fault domain in Windows Azure is defined by avoiding single points of failure, like the network switch or power unit of a rack of servers. In fact, a fault domain is closely equivalent to a rack of physical servers. When multiple virtual machines are connected together in a cloud service, an availability set can be used to ensure that the machines are located in different fault domains.

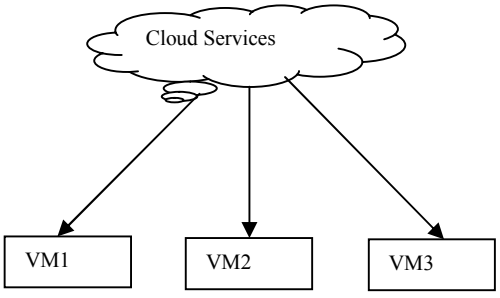

Fig.2 Virtual Machine Creation

2) Resource Allocation

Dynamic resource management has become an active area of research in the cloud computing paradigm. Cost of resources varies significantly depending on configuration for using them. Hence efficient management of resources is of prime interest to both cloud providers and cloud users. The success of any cloud management software critically depends on the flexibility, scale and efficiency with which it can utilize the underlying hardware resources while providing necessary performance isolation. Successful resource management solution for cloud environments needs to provide a rich set of resource controls for better isolation, while doing initial placement and load balancing for efficient utilization of resources.

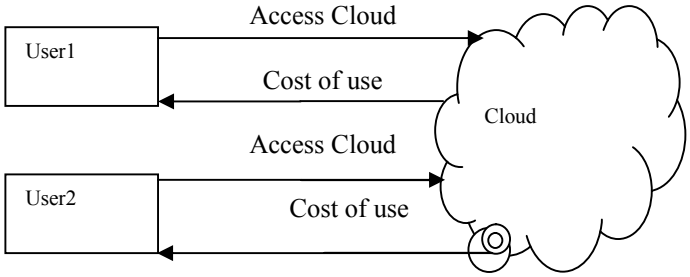

Fig.3 Resource Allocation

3) Skewness Implementation

Skewness is used to measure the uneven utilization of a server. By minimizing skewness, user can improve the overall utilization of servers in the face of multidimensional resource constraints. In case, to select the VM whose removal can reduce the skewness of the server? For each VM in the list, to see if user can find a destination server to accommodate it. The server must not become a hot spot after accepting this VM. Among all such servers, user select one whose skewness can be reduced the most by accepting this VM. All things being equal, to select a destination server whose skewness can be reduced? Skewness algorithm is to mix workloads with different resource requirements together so that the overall utilization of server capacity is improved.

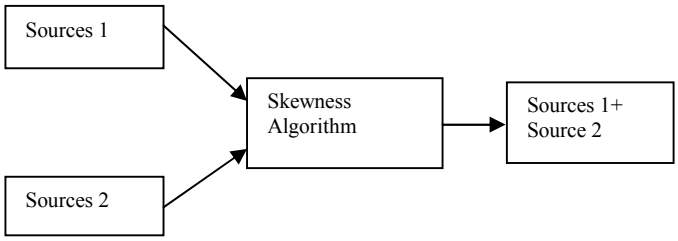

Fig.4 Skewness Implementation

4) Energy Monitoring

Multiple virtual machines can be dynamically started and stopped on a single physical machine according to incoming requests, hence providing the flexibility of configuring various partitions of resources on the same physical machine to different requirements of service requests. By dynamically migrating virtual machines across physical machines, workloads can be consolidated and unused resources can be switched to a low-power mode, turned off or

configured to operate at low-performance levels in order to save energy. Energy Monitor is used to observe energy consumption caused by virtual machines and physical machines and provides information about energy consumption to the virtual machine manager to make energy-efficient resource allocation decisions.

IV. CONCLUSION

The dynamic resource allocation is growing need of cloud providers for more number of users and with the less response time. Hence the on-demand resource allocation based SLA as per defined task priority helps to satisfy the efficient provisioning of cloud resources to multiple cloud users. In this paper, to present the design, implementation, and evaluation of a resource management system for cloud computing services. This system multiplexes virtual to physical resources adaptively based on the changing demand. To use the skewness metric to combine virtual machines with different resource characteristics appropriately so that the capacities of servers are well utilized. This algorithm achieves both overload avoidance and green computing for systems with multi-resource constraints.

V. FUTURE WORK

Our proposal utilizes the resources very efficiently. It will be very effective if it also concentrate on the processing time and processing speed. When compared to existing system our proposal can make an acceptable change in the processing time. In future, Response time is also need to be considered effectively.

REFERENCES

[1] Zhen Xiao, Senior Member, IEEE, Weijia Song, and Qi Chen proposed a "Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment", IEEE Transactions on Parallel and Distributed System, vol.24, No.6, June 2013.

[2] P.Barham,B.Dragovic,K.Fraser,S.Hand,T.Harris,A.Ho,R.Neugebauer,I.Pratt,and A. Warfield Proposed a "Xen and the Art of Virtualization," Proc. ACM Symp.Operating Systems Principles(SOSP'03),Oct.2003.

[3] C. Clark, K. Fraser, S. Hand, J.G. Hansen, E. Jul,C. Limpach, I.Pratt, and A. Warfield Proposed a "Live Migration of Virtual Machines,"Proc. Symp. Networked Systems Design and Implementation (NSDI '05), May 2005.