

VCloud: Dynamic Resource Allocation for the Cloud Users

Ranjana C, Theresa Jose

Department Of Computer Science And Engineering, Ilahia College of Engineering And Technology , MG University, India

Department Of Computer Science And Engineering Ilahia College Of Engineering and Technology MG University , India

Abstract: Cloud building means virtualization of whole system like hardware and software. That is accessible through network at any time anywhere in this world. The resources of cloud are allocated to the requested users in dynamic manner , the resources are not statically given to a particular user. In this paper we present a system that uses virtualization technology to balance the load on a physical machine. Overloading can be minimized using virtual machines to handle jobs. The green computing allows energy saving in a network. Errors while transferring data in cloud can be avoided by implementing fault tolerance.

Keywords: Cloud computing, fault tolerance, green computing, VMware workstation, virtualization.

I. Introduction

Cloud computing is the use of computing resources that are delivered as a service over a network. The scheduling of virtual machine (VM) in a cloud computing environment has become crucial due to increase in the number of users. Cloud computing is a movement away from applications needing to be installed on an individual's computer towards the applications being hosted online. Cloud resources are usually not only shared by multiple users but as well as dynamically re-allocated as per demand. This is usually done to balance load of jobs that are submitted to the service provider. Cloud computing is referred as Software as a Service that is from end user's point of view the applications are located in the cloud and software services are delivered through internet. Cloud act as Platform as a Service that is from developers point of view. The cloud providers offer an internet based platform to developers who want to create services but don't want to build their own cloud. Cloud enables Infrastructure as a Service. The cloud providers built datacenters. Power , scale , hardware , networking ,storage ,distributed systems are provided to the common cloud users. Cloud user rent storage, computation and maintenance from cloud providers. Cloud is pay as you like utility. Cloud provide data as a service. It is an infrastructure for web scale data mining and knowledge discovery.

VMware workstation provides the facility of virtualization that is we can virtualizes the physical machine using VMware workstation as well made it accessible offline over the host based physical machine (PM). VM technology allows multiple VM to run on a single physical machine. Several technical and pragmatic hurdles must be overcome when virtualizing the PC platform. PC virtualization, a personal computer, such as a desktop or portable computer, hosts several different operating systems. Server virtualization, a server computer hosts one or more, other server computers. Server virtualization makes cloud computing feasible. In technical aspect in companies for software installation license is needed. Many software's are needed in a company therefore cost for software's and its installation increases. So companies use virtualization technology. They will install software on one server and all other PCs will access that software. It is virtualization. Nowadays Citrix software is used for virtualization. Servers are of lakh's of cost, so in a company there will be a hardware that serves as a server side. The software's installed in that can be accessed and used by any computers in that company.

Cloud means connection of networks where sharing of resources is possible. So in cloud aspect the user becomes the cloud user and the server is where the request is send and processed. These all will be done in the cloud environment having lakhs of networks connected. The traditional mainframe approach runs virtual machines in a less privileged mode in order to allow the VMM (Virtual Machine Monitor) to regain control on privileged instructions, and relies on the VMM to virtualizes and interface directly to the I/O devices. Also, the VMM is in complete control of the entire machine. In the cloud environment the familiar terms are physical machine and virtual machines. Physical Machine have certain resources attached to it, like CPU , Hard disk , NIC , memory etc. These resources can be used by many users using VM concept. Resource pool is a pool where free resources are available. Based on user request these resources can be accessed. The ideal machines are identified using VM monitors (VMM). Xen Hypervisor is an example of virtual machine monitor. On demand a process needs 120 gb but the server is of 100gb as it work on the process the 20 gb is allocated virtually.

Many efforts have been made to curtail energy consumption. Hardware based approaches include novel thermal design for lower cooling power, or adopting power-proportional and low-power hardware. Dynamic Voltage and Frequency Scaling (DVFS) to adjust CPU power according to its load in data centers. Energy savings are achieved by continuous consolidation of VMs. If resource utilization of active servers is too low, some of them must be turned off to save energy. Green computing is used to optimize the numbers of servers in use. Power consumption is an important factor of cloud environment. Power off a single machine in cloud will save lot of energy instead of switch owning it if not in use. Dynamic provisioning that dynamically turns on a minimum number of servers required to satisfy application specific quality of service and load dispatching that distributes the current load among the running machines.

Our approach is motivated by two observations from real data sets collected from operating Internet services. First the total load of a typical internet services fluctuates over a day that is numbers of users accessing Windows Messenger a day. Secondly an active server, even when it is kept ideal, consumes a nontrivial amount of power. The first observation provide an opportunity to dynamically change the number of active servers, while second to shutting down machines during off-peak period provides maximum power saving. Here in this paper we try to achieve the goal of management multiple virtualization platforms and multiple VM migrations across physical machines without disruption method. Load balance is achieved using VMs when a single machine load increases. We present a system which is implementation of optimization with dynamic resource allocation dealing with virtualization machines on physical machines. The dynamic results confirmed that virtual machine which loading becomes too high it will automatically migrated to another low loaded physical machine without service disruption.

II. Related Works

The concept of virtual machine was invented by IBM, a "VM" is a fully protected and isolated copy of the underlying physical machine hardware. Intel based PC's are also increasingly being used as host to built the virtual machine. In this arena, virtual machines can be used to host applications, provide better resource utilization, and ease system manageability. Virtual machines can contain identical virtual hardware, even on hosts with different native hardware, making virtual machines freely portable between different physical machines. VMware workstation has hosted archive that allows it to co-exist with a preexisting host operating system and rely upon that operating system for device support. For developing the cloud on system: Firstly need to install VMware workstation, secondly to install VMware vcenter converter standalone client for handling data. The development procedure of cloud on core i3 using VMware workstation is presented here.

Automated resource management system that achieves a good balance between the two goals. Overload avoidance:-The capacity of PM should be able to satisfy the VM resource needs. Otherwise the PM will be overloaded. The PM count must be minimized as long as to satisfy the needs of all VMs specifies the green computing. A load prediction algorithm that can capture the future resource usage of applications. The concept of skewness is used to measure the unevenness in resource utilization. We must maximize the usage of available resources. So efficient management of resource is required for saving energy. Hot threshold means if the usage of resource is more than a specified threshold value then migration of load in VM must be done. Cold threshold is specified as if the load value on a VM is less than specified threshold then VM must go asleep saving the energy.

Migrating an active OS and all of its applications as one unit allows users to avoid many of the difficult faced by process level migration approaches. Migrating at the level of an entire VM means memory state can be transferred in a consistent fashion that is users can migrate online game or streaming media without requiring clients to reconnect. A migrating VM will include all protocol state (e.g.TCP PCBs), and will carry its IP address with it. The migration process involves certain stages they are Pre-Migration, Reservation, Iterative Pre-Copy, Stop -and - Copy, Commitment, and Activation. A key challenge in managing the migration of OS instances is what to do about resources that are associated with the physical machine that they are migrating away from. While memory can be copied directly to the new host, connections to local devices such as disks and network interfaces demand additional consideration. Managed migration is performed by migration daemons running in the management VMs of the source and destination hosts.

Thin kernel is designed to run VMs ESX server. It also multiplex hardware resources to virtual the Intel architecture. IA-32 architecture manages system hardware for high performance I/O. It is used to run unmodified commodity OS. Virtualization enables the running of multiple OS on a single machine. Guest OS needs to see a zero based memory space. Each VM gets a configurable max size of physical memory. Each Virtual Machine (VM) is isolated and protected from each other and gives an illusion of dedicated physical machine. It also allows an abstraction of server workloads. Guest OS needs to see a zero-based memory space that means virtual machines will share the memory space of physical machine. Resource rights are distributed to clients through shares.

There are several methods for distributed VM images to physical machines in a cloud computing environment no matter if it is a private cloud connected via a fast local area network or several remote cloud computing sites connected via the internet. The primary goal for efficient VM image distribution is to avoid congested links between the node where a VM image resides and the compute nodes. The communication time between the compute nodes should be minimized, whereas the actual data exchange should be maximized: distribute max data in minimal time. All important OS aspects (ie saving the state of the CPU , network , memory are covered by VM migration).

Desktops computes in enterprise environment consume lot of energy. We must save energy by lettering these machines sleep while avoiding user description. Virtualization is used for this process by migrating ideal desktop to a serve where they can remain always on. There is automated mechanism to drive the migration of the desktop computing environment between the physical desktop machines and the VM server. To reduce energy wastage by idle desktop computers in enterprise environments, the typical approach is to put a computer to sleep during long idle periods (e.g. overnight), with a proxy employed to reduce user disruption by maintaining the computer's network presence at some minimal level. To determine a desktop is ideal or active the presence of UI (User Interface) activity is initialized by user through mouse or keyboard clicks and pressed. In concept of Light green there is a controller that receives periodic updates from stubs on the desktop hypervisors on the level of user.

Datacenter administrators need to meet the SLO for hosted application. Several challenges like complex SLO that is to determine the CPU and disk allocation needed to spent fundamental transaction per unit time. Time varying resource requirements the distributed resource allows that is the tire application span multiple nodes require resources allocation access all tires to be at appropriate levels to meet end to end application SLOs. These are solved with shared virtualized infrastructure. Each physical node hosts multiple application tires running on VMs; multitier application can span multiple nodes.

A management algorithm for dynamic allocation of VM to physical services is presented. The algorithm pro-actively adapts to demand changes and migrates VMs between physical hosts. Forecasting technique is suited for handling time series of resource demands in a typical datacenter & capable of automatic adaption to various domain patterns.MFR dynamically remaps VM to PMs minimizing the number of PMs required to support a workload. The algorithm specified in this work reduces resource consumption.

III. Existing System

Existing works on autonomic management systems for virtualized server environments tackle the allocation and placement of virtual servers from different perspectives. Virtual machine monitors (VMMs) like Xen provide a mechanism for mapping virtual machines (VMs) to physical resources. Mapping is largely hidden from the cloud users. Users with the Amazon EC2 service example do not know where their VM instances runs. It's up to the cloud provider to make sure the underlying physical machines (PMs) have sufficient resources to meet their needs. The main problem of existing system is ,there are two VMs available VM1 and VM2 for a PM with size 50 and 60mb load can be handled. There are two files with f1 with 55 mb and f2 with 55mb.So the VMs is of 50 and 60 so only one file can be handled ie 55mb size file 2 can be handled by VM2 of size 60mb.The file1 remain unprocessed. This is very crucial situation in case of handling files. So an efficient handling of load balancing technique is required. There are limitations such as 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. There is no control over the business assets (data). The main assets in every company are its data files with valuable customer information. Risk of data loss due to improper backups or system failure in the virtualized environment. When network crash occurs high cost and loss of control of data is occurred.

IV. Proposed System

Cloud computing helps the cloud users to access their resource based on needs. The resources include the memory, network, CPU etc. The virtualization technology is used to achieve balanced resource allocation. The load balancing is done in accurate manner .A cloud environment is implemented by making a physical machine a virtual machine by installing VMware.VM is fully protected and isolated copy of underlying PMs hardware. VMware workstation has a hosted architecture that allows it to co-exist with a preexisting host operating system and rely upon that OS for device support.

Load balancing means balance the load on a PM. When a physical machine is overloaded its works are splitted and given to other VMs. If a large file is needed to process , it is splitted based on the size of file and is handled by different VMs. Here it shows how the migration of load is taken place when one PM load increases beyond a threshold. A load means any job submitted to process. For eg :in this context the job is to count the number of words in the text file.

If any VM is not in use it is shutdown to save energy. The concept of green computing reveals that in a cloud environment energy saving is done if any one machine is off at least for 0.5 second. To help in dynamic

allocation always opening all VMs available is not fare. ie logging VM several times takes more energy. So to save energy the VM not used must go asleep and on any user request it must wakeup and perform the job.

4.1 Advantages

A platform has been architected and a prototype is implemented. The measurement of resource usage and end-user activities lies in hands of the cloud service provider. Cloud service provider provides services to the cloud users. Opaque cost structure due to highly flexible usage of cloud service provider. This proposed system process a stable cost structure.

4.2 Architecture

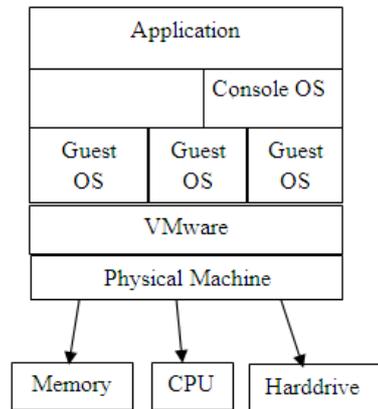


Fig1: System Architecture

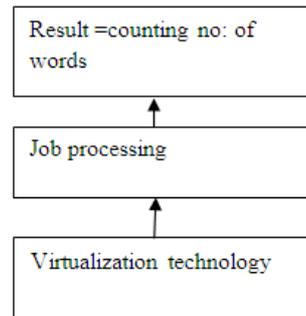


Fig 2: Processing Of Job

Fig 1: It describes thin kernel is designed to run VM. It multiplexes hardware resources virtualizes the Intel IA-32 architecture. It manages system hardware for high-performance I/O. Runs unmodified commodity operating systems. Virtualization enables the running of multiple operating systems on a single machine. Each Virtual Machine (VM) is isolated and protected from each other. The figure shows an illusion of dedicated physical machine and allows an abstraction of server workloads. Fig 2 describes, using Virtual machines the user request can be handled easily. The resources are dynamically distributed.

V. Solution Methodology

5.1 Cloud building using VMware workstation

For developing a cloud system, 1st need to install VMware workstation. 2nd install VMware vCenter converter standalone client.

1. Download VMware Workstation 10 from the website of VMware in Linux or Windows.
2. Install the download version of the VMware WS on PM.
3. Then install VMware vCenter convert standalone client over PM for data management.
4. Click in VMware virtualize a PM option for developing the VM.

Now we must use an ios of any one of operating system. Here we are using ios of Windows XP. A common folder is made in any one of the drive. It can be shared at the other VMs. Now the physical machine has been transferred as VM. Now we can run any number of VMs based on RAM size of physical machine. All the resources connected to the PM can also be used by the VMs. Now the VMs running behind the PM will run as an individual OS for users. There will be different ip address for the PM and for other VMs.

5.2 Resource Allocation

The free memory available in two VMs can be calculated from the PM. So in each VM we must 1st calculate the free memory available. The max memory and free memory of each VMs can be calculated at runtime in those VMs. Then the calculated free memory space is send to the PM.

5.2.1 Client Receive Load From VM

Here in this proposed work the VM is considered as servers and the physical machine is considered as client. So in this section it describes how the free memory space available at each VMs are sent to the PM. The memory is specified in bytes. To receive the memory from VM, a socket is opened at client that is at PM a socket is opened. The port number of socket is specified. A server socket ss1 is created and accept class is used to accept the data. Here data means the free memory sent from the VMs. The Input stream is used to receive the input from the other socket.

We are receiving data in the form of text file so to receive the memory information a file is created. To create a file, a file declaration is made using .txt extension. Output stream is used to store the data. The data is stored as an array in buffer. So from buffer the data are written to the output stream, since in buffer the data is stored as binary 0's and 1's. The while loop is used to check the availability of data. If the value is -1 then it means no data is there. File reader is used to read the file and buffered reader is used to read data from buffer. Close the output stream, inputstream used. Close the client socket opened.

5.2.2 Load the memory from VM

From the virtual machine the free memory space can be loaded. Free memory and max memory of VM is obtained by using getRuntime function. A file fl a text file is declared. File is written using file writer and buffered writer is used for writing it from stored space. So the free memory available in each VM is written by using the FileWriter. The written data is converted to string and stored in variable m1. The buffered writer is refreshed and close the buffered writer.

5.2.3 Send the calculated memory available in VM to PM

A file declaration is made using .txt extension eg: ./VM1/freememoryvm1.txt. Then open a socket specifying the machine name and port number to where the data is to be send. The length of file is calculated using fl.length and convert the length stored in byte as integer using new byte[(int)fl.length()]. Buffered inputstream is used to read the input. Then from the array data is read from starting and end point of array. The output stream is used to write the data and output stream is refreshed and closed. Then finally close the client socket.

5.3 Processing of job

A file is chosen to be processed. The aim is to count the number of words in the text file. At receiver side open the socket to receive the data to be processed and the filename is read using buffered reader. The filename is read here to do the message transfer. So in send info sender will specify ip address and two port numbers of vm1 and vm2. In receive at client side socket is opened to receive the file. The length of file is taken. In proceed sending file is done ie by taking the filename and also taking the length of file. The port number and to which VM the data is send its ip is also set here. In display the task is done ie the word count is taken. A text file is given as input and the number of words in that text is taken using string tokeniser. The string tokeniser will count the number of words in a sentence based on the space. Space act as an indicator to detect the no: of words. The sent result shows the calculated count and it is sent to the receiver side. So at client receiver the count value is obtained. The writing is done to the socket.

5.4 Demand Sharing

The overloading concept is explained here. A system is said to be overloaded when it cannot satisfy all the needs of VM attached to it. Overloading reduces the system performance. To avoid overloading demand sharing is done. In demand sharing, based on the demand from user the resources are shared. Here we consider the memory as a resource. This memory of each VM is needed to process the work. A large file is taken it is splitted based on the size of file. The size of file is read as in KB and stored in two files as f1 and f2. The file is splitted only after it is read. The size of large file is calculated in kilobytes. Then the size in kb is splitted and based on size the VMs handle these splitted text files.

5.5 Shared Allocation

Now the free memory space in two VMs are calculated. Then these splitted file are send to these VM and they are processed and corresponding result is obtained in those VMs. The result from VM1 and VM2 is sended back to PM. The PM shows the compared result from two VMs.

5.5.1 Which VM should be used to process the job

First check whether VM1 free memory is greater than VM2 free memory. If so print VM1 has maximum memory and VM1 name is stored in variable mcnam. The length of two files are checked if length of file 1 is greater than length of file 2 then splitted file1 will be allocated to VM1 and splitted file2 will be allocated to VM2. The two file are written using output stream using os1.write(mybyte1,0,mybyte1.length) and os2.write(mybyte2,0,mybyte2.length). In mybyte1(VM1) the length of file1 is stored.

The data is stored as an byte array. Zero specifies the starting value in array and mybyte1.length specifies the end value in array. And if else splitted file2 will be allocated to VM1 and splitted file1 will be allocated to VM2. Then writing of files are done using os1. Write (mybyte2,0,mybyte2.length) and os2.write(mybyte1, 0, mybyte1.length). If length of file size changes the splitted files is either allocated to lower VM or

high VM. Here we do the processing of text file only with .txt extension. In cloud storage at a time only one organization details are at a time processed.

5.6 Green Computing

Energy efficiency is increasingly important for future ICT. Rapid growth of the demand for computational power has led to the creation of large-scale data centers. Energy savings are achieved by continuous consolidation of VMs. The power consumption data for each server was obtained by first calculating the maximum power using HP's power calculator, then following the convention that average power use for midrange/high-end servers is 66% of maximum power. Hard disk arrays include supporting functionality such as cache memories, disk array controllers, disk enclosures, and redundant power supplies. In a cloud computing data center, all the storage space in the data center is consolidated and hard disk usage is centrally coordinated. Through server virtualization/consolidation, a very large number of users can share a single server, which increases utilization and in turn reduces the total number of servers required. Users do not have or need any knowledge of the tasks being performed by other users and utilize the server as though they are the only user on the server. During periods of low demand, some of the servers enter a sleep mode which reduces energy consumption. It is the responsibility of the IT manager to reduce the consumption of IT load (renewing servers, virtualizing, etc.), and the responsibility of the infrastructure manager to reduce the auxiliary consumption (more efficient equipment, free-cooling, etc.).

Here the unused VMs are power off to save energy. Green computing is nothing but reducing energy consumption and recyclability. According to this work we are using virtual machines for the purpose of recycle. i.e not a machine permanently allocated to an user that will be deallocated and assigned to any other user rotationally. And also here we are reducing energy consumption with the help of selecting VM with max free memory. If we are allocating a virtual machine with less memory then it will be slow to process the job and consume more energy that is identifying a virtual machine that is suitable to user which takes quick time to process the job.

VI. Conclusion

Cloud computing enables the business customers to access resources dynamically through virtualization technology. VMware software is used to virtualize the machine. When a machine is heavily loaded, the excess load is migrated to any other available virtual machine and the process can be executed. Dynamic resource sharing can enable effective resource allocation which reduces the cost of resource. The green computing allows energy saving by shutting down VMs not in use.

VII. Future Work

Fault tolerance service provider (SP): The entity that provides fault tolerance support to applications based on the client's requirements. A fault tolerance algorithm that handles failures at a finer granularity, and offers high-performance guarantees, consumes higher amount of resources. The task of offering fault tolerance as a service requires the service provider to realize generic fault tolerance mechanisms such that the client's applications deployed in virtual machine instances can transparently obtain fault tolerance properties. To this aim, we define ft-unit as the fundamental module that applies a coherent fault tolerance mechanism to a recurrent system failure at the granularity of a VM instance. The notion of ft-unit is based on the observation that the impact of hardware failures on client's applications can be handled by applying fault tolerance mechanisms directly at the virtualization layer than the application itself.

Here in PM the fault unit is active i.e here we are manually creating an error. The IP address of one virtual machine is given as wrong so as data is to be processed in that VM it cannot work. So fault mechanism is done. The other VM which is free in the network can be utilized to handle this process. So in cloud if any VM is stopped working or its IP address cannot be identified properly, the work assigned to this VM can be done with another free VM using Alternative path mechanism once fault tolerance had been designed. Suppose the client VM is receiving load from user, so its IP is changed by an intruder as 0.0.0.Z. Usually the format of IP address is as 192.168.1 etc. So this IP 0.0.Z cannot be located by VM. So an error message is shown that Connection Problem. Since fault tolerance is active another VM available can be used.

Acknowledgments

The authors wish to thank the faculty of the department for their support and help in completing this work.

References

- [1]. Dynamic Resource Allocation Using VM for cloud computing environment, Zhen Xiao, Senior Member, IEEE, Weijia Song, and Qi Chen, IEEE transactions on parallel and distributed system, VOL. 24, NO. 6, JUNE 2013.
- [2]. Cloud Bulding on core-i3 using Vmware Workstation, International Conference on Pervasive Computing and Communication 2012.

- [3]. Live Migration Of Virtual machine, May 2005 Christopher Clark, Keir Fraser, Steven Hand, Jacob Gorm Hansen, Eric July, Christian Limpach, Ian Pratt, Andrew Wareld
- [4]. Memory resource management in Vmware ESX server, Carl A. Waldspurger, IEEE August 2002
- [5]. Energy-Aware Server Provisioning and Load Dispatching for Connection intensive Internet services, Gong Chen, Wenbo He, Jie Liu, Suman Nath, Leonidas Rigas, Lin Xiao, Feng Zhao, August 2008
- [6]. LiteGreen: Saving Energy in Networked Desktops Using Virtualization, Tathagata Das, Pradeep Padala, Venkata N. Padmanabhan, IEEE 2007.
- [7]. Automated control of Multiple Virtual resources, Pradeep Padala, Kai-Yuan Hou, Kang G. Shin, Xiaoyun Zhu Mustafa Uysal, Zhiku, IEEE 2009.
- [8]. Dynamic Placement of VM for Managing SLA violations, Norman Bobroff, Andrzej Kochut, Kirk Beaty, IEEE 2007, IBM T.J. Watson Research Center
- [9]. Fault Tolerance Management in Cloud Computing: A System-Level Perspective, IEEE SYSTEMS JOURNAL, VOL. 7, NO. 2, JUNE 2013.