Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 11, November 2015, pg.194 – 198

SURVEY ARTICLE

Survey on Dynamic Resource Allocation Strategy for Cloud Computing

Vinayak Sontakke[1], Paresh Patil[2], Sumedh Waghamare[3], Rajani Kulkarni[4], Mrs. M. Saravanapriya[5], Mrs. N. S. Patil[6]

Department of Computer Science, University of Pune, India

[1] vbsontakke@gmail.com; [2] paresh21p@gmail.com; [3] sumesh9495@gmail.com; [4] rajanikulkarni22@gmail.com

Abstract— *In this paper, we survey various technologies in the field of virtualization that are used for 'on demand' dynamic resource allocations.* ***Cloud Computing is one of the emerging areas in the various fields related to computer science and can be exercised for on demand dynamic resource allocation for providing reliable and guaranteed services to the consumer. Cloud systems are 'on demand' as they offer a way of providing required resources to consumers as and when they need it. It is pay-as-you-use manner service. In Cloud systems, a pool of resources is shared by several users. So making these resources available in most optimal way is a challenging task. This paper reviews various techniques for how such dynamic resource allocations can be achieved in an optimal manner.***

Keywords— Cloud Computing, Dynamic Resource Allocation, Virtual Machine, Virtualization Technology, VM placement, Xen

I. INTRODUCTION

Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. The services themselves have long been referred to as Software as a Service (SaaS). The datacenter hardware and software is what we will call a Cloud. When a Cloud is made available in a pay-as-you-go manner to the general public, we call it a Public Cloud; the service being sold is Utility Computing. We use the term Private Cloud to refer to internal datacenters of a business or other organization, not made available to the general public. Thus, Cloud Computing is the sum of SaaS and Utility Computing, but does not include Private Clouds. People can be users or providers of SaaS, or users or providers of Utility Computing. We focus on SaaS Providers (Cloud Users) and Cloud Providers, which have received less attention than SaaS Users. From a hardware point of view, three aspects are new in Cloud Computing.[2]

1. The illusion of infinite computing resources available on demand, thereby eliminating the need for Cloud Computing users to plan far ahead for provisioning.
2. The elimination of an up-front commitment by Cloud users, thereby allowing companies to start small and increase hardware resources only when there is an increase in their needs.
3. The ability to pay for 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, thereby rewarding conservation by letting machines and storage go when they are no longer useful.

Cloud computing is delivers the computing capacity and storage as a service to a community of end user. Cloud computing entrusts services with user's data, software and computation over a network. Remote accessibility enables the access to services from any location at any time. The cloud service provider can multiplex available virtual resources on to physical resources. Virtual Machine Monitor like oxen provide mechanism for mapping virtual machines. Mapping is kept hidden from users. I.e. end user will not know the location of physical resources. Virtualization is a core technology in cloud environment. Virtualization include VM placement. Cloud user requests for user request for resource. Resource may be software, operating system, applications which are integrated as VM. Service Provider provide the resource according to VM requirement. VM live migration technology makes possible to change mapping between VM and PM. [2]

II. RELATED WORK

Virtualization: Virtualization hides the physical characteristics of computing resources from their users, be they applications, or end users. This includes making a single physical resource (such as a server, an operating system, an application, or storage device) appear to function as multiple virtual resources; it can also include making multiple physical resources (such as storage devices or servers) appear as a single virtual resource [8].

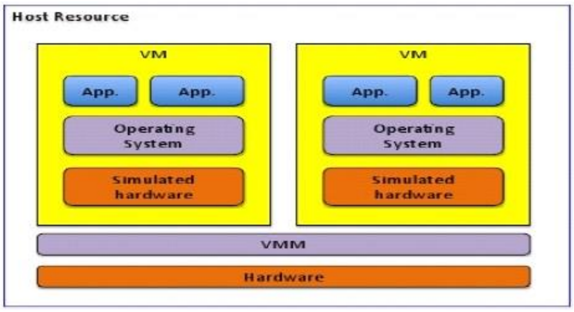

Fig. 1 Virtual Machine

Hypervisor is a software program that manages multiple instances of the same operating system on a single computer system. The hypervisor manages the system's processor, memory, and other resources to allocate what each operating system requires. [8]

Hypervisors are designed for a particular processor architecture and may also be called **virtualization managers**

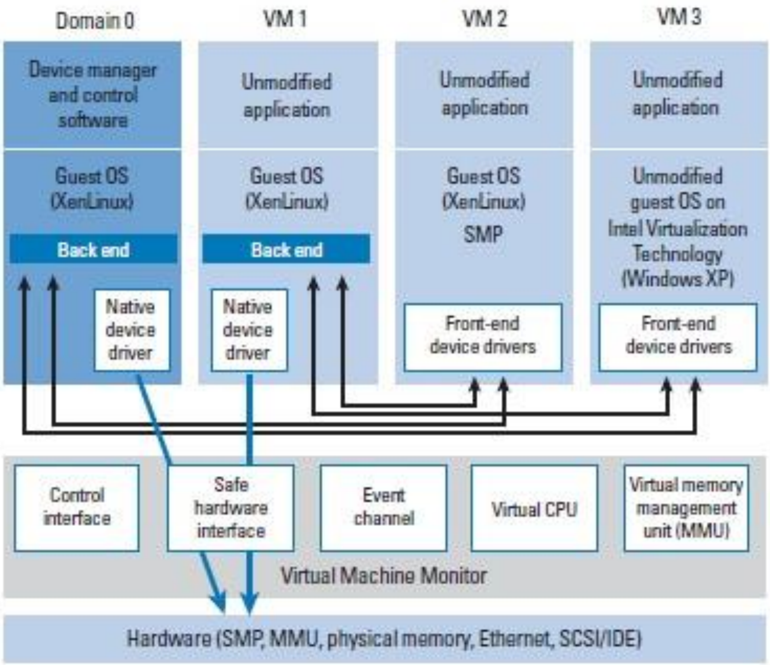

Fig. 2 Xen Architecture

A. Overview of Xen

Recent advances in virtualization technologies enabling data centers to consolidate servers, normalize hardware resources, and isolate applications on the same physical server are driving rapid adoption of server virtualization in Linux environments. Xen 3.0 architecture is near native speed virtualization software for Intel x86 architectures. Virtualization software abstracts the underlying hardware by creating an interface to *virtual machines* (VMs), which represent virtualized resources such as CPUs, physical memory, network connections, and block devices. Software stacks including the OS and applications are executed on top of the VMs. Several VMs can run simultaneously on a single physical server. Multiplexing of physical resources between the VMs is enforced by a VM monitor, which is also designed to provide the required translations of operations from the VMs to the physical hardware.

Xen is an open source virtualization software based on para virtualization technology. The architecture of Xen 3.0 hosting four VMs (Domain 0, VM 1, VM 2, and VM 3) is shown in figure 2 below. This architecture includes the Xen Virtual Machine Monitor (VMM), which abstracts the underlying physical hardware and provides hardware access for the different virtual machines. Domain 0 has special role in VM. Only Domain 0 can access the control interface of the VMM, through which other VMs can be created, destroyed, and managed. Management and control software runs in Domain 0. Administrators can create virtual machines with special privileges—such as VM 1—that can directly access the hardware through secure interfaces provided by Xen. Administrators can create other virtual machines that can access the physical resources provided by Domain 0's control and management interface in Xen. [10]

B. Types of Virtualization

Para-virtualization: VM does not simulate hardware Use special API that a modified guest OS must use hyper calls trapped by the Hypervisor and serviced Para virtualization involves modifying the OS kernel to replace non-virtualizable instructions with hyper calls that communicate directly with the virtualization layer hypervisor. The hypervisor also provides hyper call interfaces for other critical kernel operations such as memory management, interrupt handling and time keeping. Para virtualization is different from full virtualization, where the unmodified OS does not know it is virtualized and sensitive OS calls are trapped using binary translation. Para virtualization cannot support unmodified OS. [10]

OS-level virtualization: OS allows multiple secure virtual servers to be run guest OS is the same as the host OS, but appears isolated apps see an isolated OS kernel of an OS allows for multiple isolated user-space instances, instead of just one. Each OS instance looks and feels like a real server OS virtualization virtualizes servers on the operating system (kernel) layer. This creates isolated containers on a single physical server and OS instance to utilize hardware, software, and data center and management efforts with maximum efficiency. OS-level virtualization implementations that are capable of live migration can be used for dynamic load balancing of containers between nodes in a cluster. [10]

Application level virtualization: A fully virtualized application is not installed in the traditional sense, although it is still executed as if it is (runtime virtualization) Full application virtualization requires a virtualization layer. E.g. JVM [10]

III. ALLOCATION STRATEGIES

Static Allocation: In static allocation the cloud user has to make prior request for the resources. By doing so user knows what resources are required and how many instances of the resources are needed ahead of using the system. But the drawback in this case is it leads to underutilization or overutilization of resources depending on the time the application is run.[6]

Dynamic Allocation: Cloud resources are requested by the cloud user on the fly or as and when the application needs. Here underutilization and overutilization of resources is avoided as much as possible. But the requested resources might not be available when requested on the fly. The service provider has to make allocation from other participating cloud data center.

There are so many algorithms for scheduling in cloud computing. The main advantage of scheduling algorithm is to obtain a high performance. The main examples of scheduling algorithms are FCFS, Round-Robin, Min-Min algorithm, Max-Min algorithm and meta- heuristic algorithms (ACO, GA, Simulated annealing, PSO, Tabu search and many more) [6].

1 Round Robin Algorithm (RR): In this datacenter controller uses the concept of time quantum or slices it assigns the requests to a list of VMs on a rotating basis. The resources of the service provider are provided on the basis of the time quantum. The first request is allocated to a VM randomly from the group and then the datacenter controller assigns the subsequent requests in a circular order. Once the VM is assigned the request, the VM is moved in circular motion to the end of the list. In Round Robin Scheduling if time quantum is very large then Round Robin Scheduling Algorithm is same as the FCFS Scheduling and if the time quantum is extremely too small then Round Robin Scheduling is called as Processor so time quantum play an important role in RR algorithm. In this RR algorithm; there is a better allocation concept known as **Weighted Round Robin Allocation** in which we can assign a weight to each of VM so that if one VM is capable of handling twice as much load, then powerful server gets a weight of 2. In this cases, the data center controller will assign two requests to the powerful VM for each request assigned to a weaker one. [9]

2 Throttled Load Balancer (TLB): In this algorithm the load balancer maintains the record of each state (busy or ideal) in an index table of virtual machines. First the client or server makes a request to data center to find a suitable virtual machine to perform the recommended job. The data center queries the load balancer for allocation of the VM. The load balancer check the index table from top until the first available VM is found, if the VM is found, the load data center communicates the request to the VM identified by the id. Further, if appropriate VM is not found, the load balancer returns -1 to the data center. When the VM completed the allocated task, a request is acknowledged to data center, which is apprised to load balancer to de- allocate the same VM whose id is already communicated. The total execution time is estimated in three phases. In the first phase the formation of the virtual machines and scheduler will be idle waiting to schedule the jobs in the queue ,in second phase once jobs are allocated, the virtual machines in the cloud will start processing, and finally in the third phase the destruction of the virtual machines. The throughput of the model can be estimated. as the total number of jobs executed within a required time span without considering the any destruction time. This algorithm will improve the performance by providing the resources on-demand, by reducing the rejection in the number of jobs submitted and resulting in increased number of job executions.[9]

3 Active Monitoring Load Balancer (AMLB): The AMLB maintains information about each VMs and also the number of requests currently allocated to which VM. When a request is to allocate a new VM arrives, first it identifies the least loaded VM. If there are more than one loaded VM, then the first identified is selected. Active VM Load Balancer returns VM id to datacenter controller. The datacenter controller sends the request to the VM identified by that id. This algorithm is quite similar with Weighted Round Robin Algorithm of cloud computing in order to achieve better response time and processing time. [9]

4 Adaptive Resource Allocation (ARA): We propose a new load balancing algorithm, called as ARA, adaptive resource allocation in cloud systems, which improve the overall system performance, availability and counteract the effect of business. The main contributions of this algorithm are (1) to present an on/off prediction approach which accurately forecasts changes in user demands by the knowledge of burstiness in workloads; and (2) to develop a smart load balancer, "random" (i.e., randomly select one among all sites) and which on-the-fly shifts between the schemes that are "greedy" (i.e., always select the best site) and based on the predicted information. ARA reduces the response times by optimizing the dispatch of loads across computing sites and adapts a smart site selection for cloud users under both bursty and non-bursty workloads. Amazons EC2 reveal the effectiveness of ARA in a real cloud environment. Under non-bursty conditions the "greedy" methods will always select the best site, obtain better performance than the "random" ones. But under bursty conditions we distributing jobs randomly among all computing sites. ARA algorithm tunes the load balancer by adjusting the trade-off between greediness and randomness in the selection of sites. [9]

5 Skewness Algorithm: Skewness is the measure of unevenness resource utilization of a server. By minimizing the *skewness*, we can combine different types of workloads nicely and improve the overall utilization of server resources [1].

6 Preemptable shortest job next scheduling algorithm (PSJN): This algorithm is proposed in a private cloud. In this paper they combine the pre-emption technique of Round-robin algorithm with shortest process next (PSN).This algorithms gives cost benefits and improve the response time and execution time. [9]

7 Optimized Activity based Costing algorithm: In this paper implementation of the optimized algorithm is compared with the traditional task scheduling algorithm. The main goal of this optimized algorithm is to gain more profit as compare to the traditional ones. ABC is the scheduling algorithm of measuring the object's cost as well as the performances of activities. [9]

8 Improved Cost Based algorithm: This algorithm improves the traditional cost-based scheduling algorithm for making appropriate mapping of tasks to resources. It grouped tasks according to the processing capabilities of available resources.[9]

9 User-Priority Guided Min-Min scheduling Algorithm: In this paper an improved load balanced algorithm is introduced on a base of Min-min algorithm in order to minimize the make span and maximize the utilization of resource. [9]

10 Scheduling algorithm based on QOS: In this paper the proposed algorithm is based on quality of service driven. It computes the priority of tasks on the basis of different attributes of tasks and after that apply sorting on tasks onto a service which can further complete the tasks. [9]

IV. CONCLUSIONS

Cloud computing offers a way for using the available resources in an optimal manner. Resources can be made available to users in an on demand basis. Such emerging technologies are being progressively used in enterprise and business markets. The cloud computing offers business customers to scale up and down their resource usage based on need. The main advantage in the cloud model comes from resource multiplexing through virtualization technology. Virtualization technology is used to allocate data center resources dynamically based on application needs and support green computing by optimizing the number of servers in use.

REFERENCES

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

[2] M. Armbrust et al., "Above the Clouds: A Berkeley View of Cloud Computing," technical report, Univ. of California, Berkeley, Feb. 2009.

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

[4] C.A. Waldspurger Proposed a "Memory Resource Management in Vmware ESX Server," Proc. Symp. Operating Systems Design and Implementation (OSDI '02),Aug. 2002.

[5] G. Chen, H. Wenbo, J. Liu, S. Nath, L. Rigas, L. Xiao, and F. Zhao a "Energy- Aware Server Provisioning and Load Dispatching for Connection-Intensive Internet Services," Proc. USENIX Symp. Networked Systems Design and Implementation (NSDI '08), Apr. 2008.

[6] Jiayin Li, Meikang Qiu, Jian-Wei Niu, Yu Chen, Zhong Ming, "Adaptive Resource Allocation for Pre-empt able Jobs in Cloud Systems," in 10th International Conference on Intelligent System Design and Application, Jan. 2011, pp. 31-36.

[7] Goudarzi H., Pedram M., "Multi-dimensional SLA based Resource Allocation for Multi-tier Cloud Computing Systems," in IEEE International Conference on Cloud Computing, Sep. 2011, pp. 324-331.

[8] Chieu T.C., Mohindra A., Karve A.A., Segal A.,"Dynamic Scaling of Web Applications in a Virtualized Cloud Computing Environment," in IEEE International Conference on e-Business Engineering, Dec. 2009, pp. 281-286.

[9] Prabhjot Kaur and Dr. Pankaj Deep Kaur, "Efficient and Enhanced Load Balancing Algorithms in Cloud Computing, Feb 2015

[10] Tim abels, Puneet Dhawan, and Balasubramanian Chandrasekaran, "An Overview of Xen Virtualization" , August 2005.

