IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 10, December 2014.
www.ijiset.com
ISSN 2348 – 7968

A Review-
Dynamic Resource Allocation using Virtual Machines for Cloud Computing Environment

Mahesh Gaykar, Pankaj Deshmukh, Navnath Tapkir, Rupali Shende

Department Computer Science and Engineering ,MIT Acadamy Of Engineering, Pune, India

Department Computer Science and Engineering MIT Acadamy Of Engineering, Pune, India

Department Computer Science and Engineering MIT Acadamy Of Engineering, Pune, India

Department Computer Science and Engineering MIT Acadamy Of Engineering, Pune, India

Abstract

A cloud computing infrastructure is a complex system with a large number of shared resources. Cloud resource management requires complex policies and decisions for multi-objective optimization. In Cloud computing multiple cloud users can request number of cloud services simultaneously. In this paper virtualization is used to allocate resources based on client needs and also supports green computing concept. It offers the virtualized resources to the cloud users. The concept Skewness algorithm is used to measure the resource utilization of server and minimizing the Skewness can improve overall utilization of server. Overload avoidance is maintained using skewness and virtualization. This paper provides detailed description of the dynamic resource allocation techniques in cloud for cloud users

Keywords—cloud computing, virtualization, Green Computing, Skewness, Migration

I. INTRODUCTION

Cloud computing [1] is the next stage in evolution of the internet. The cloud provides the means through which everything from computing power to computing infrastructure, applications, business processes to personal collaboration can be delivered to you as a service wherever and whenever you need. Cloud computing is the next generation in computation.

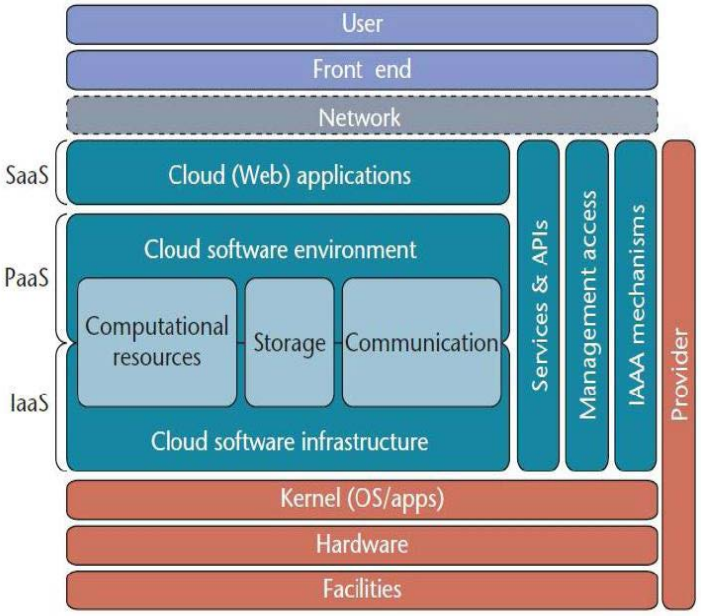

Figure 1: Cloud Computing Architecture

It is a new term currently being popular in the market. Cloud computing is emerging as a new paradigm of large scale distributed computing. Now a days, due to rapidly changing requirements of users and increase in demand of services and applications requiring large heavy and powerful resources has led to the development of cloud computing.

The strategies for cloud resource management associated with the three cloud delivery models, Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service

ISSN 2348 – 7968

(SaaS), differ from one another. The end users can use these resources over a network on-demand basis in pay-as-you-say manner.

In infrastructure as a service (IaaS), a cloud based virtual server provides networking and mass storage services and other infrastructure services. In platform as a service (PaaS), providing computable platform where the user can host and develop applications and services.
In Software as a Service (SaaS), an application is hosted as a service to customers who access it via the Internet. When the software is hosted off-site, the customer doesn't have to maintain it or support it. On the other hand, it is out of the customer's hands when the hosting service decides to change it.

Cloud computing is offered in different forms:
- ✓ Public clouds
- ✓ Private clouds
- ✓ Hybrid clouds, which combine both public and private

.

1) Public Cloud:
Services are open for public use. Different types of public cloud service providers are Amazon AWS, Microsoft and Google.

2) Private Cloud:
Operated solely for a single organization. It may be managed by the organization or a third party, and may exist on-premises or off-premises.

3) Hybrid Cloud:
The cloud infrastructure is a combination of two or more clouds (private, community, or public) that they are unique entities but are bound together by standardized technology that enables data and application portability.

Cloud computing is a collection of different virtual machines(VM) which are installed on the physical machine(PM).Each PM may have more than one VMs.
Live virtual machine migration is a unique capability of system virtualization which allows applications to be transparently moved across physical machines with a consistent state on current PM.
In cloud computing, Resource Allocation is the process of assigning available resources to the needed cloud applications over the internet using one resource allocation algorithm among all. The dynamic resource allocation based on distributed multiple criteria decisions.
The two main goals that can be achieved are

1. The capacity of PM should be able to satisfy the needs of the VM's running.
2. To save Power and well utilization of PMs. The number of PM's should be minimized.

The proposed contributions are:
1] A resource allocation system that can avoid overload in the system effectively while minimizing the number of servers used.
2] The concept of "skewness" to measure the uneven utilization of a server. By minimizing skewness, one can improve the overall utilization of servers in the face of multi-dimensional resource constraints.
3] A load prediction algorithm that can capture the future resource usages of applications accurately without looking inside the VMs. The algorithm can capture the rising trend of resource usage patterns and help reduce the placement churn significantly.

Thus we are going to discuss various resource allocation algorithms, skewness algorithms and different terms like cold spot, Hot spot**.**

II. VIRTUALIZATION

The base term of cloud computing is Virtualization [2]. Virtualization separates resources and services from the underlying physical delivery environment. When you think about cloud management, it's important to separate resources from their physical implementations. Without virtualization, the cloud becomes very difficult to manage. Virtualization is so important for cloud computing because it is possible to simplify many aspects of computing.Virtualization is creating a virtual machine as like hardware system, an operating system or a network resource. In virtualization, many applications and operating systems (OSes) are supported in a single physical system by partitioning (separating) the available resources.Here the CPU is shared among the operating systems. Memory is shared among the virtual machines for resource allocation and utilization purpose. Virtualization is a huge term that refers to computing elements which are running on a virtual machine rather than on a real machine. Thus one can efficiently utilizing the resources according to the application need by using the concept of Virtualization. Virtualization technology can expand the capacity of the hardware, and simplify software re-configuration process. One of the benefits of virtualization is the way that it abstracts hardware assets, in essence allowing a single piece of hardware to be used for multiple tasks. The virtualization technology is a key technology for cloud computing. On a cloud computing platform, dynamic resources can be effectively managed using virtualization technology.

ISSN 2348 – 7968

III. DIFFERENT TERMS FOR RESOURCE ALLOCATION

A. Skewness :

Skewness is used to identify the uneven utilization of multiple resources. By minimizing the skewness it leads efficient utilization of resources as well as elevates the performance and rescue power too. Key concepts of Skewness[2]: hot spot migration, load prediction, and green computing.

B. Black Box Monitoring :

In black box huge number of migration of VMs can take place by increasing the number of resource request (load) on virtual machine. In this concept railing [8] processor, network and memory usage of each virtual server by the monitoring engine. Here monitoring, load prediction and profile generation are difficult process. Also different set of CPU, Memory, Network monitoring is done.

C. Gray Box Monitoring :

Gray-box monitoring is light-weight monitoring process that is run inside each virtual server [9]. Gray-box monitoring supports statistics of memory, network, and CPU usage in different platforms. Some of processes are the memory usage monitoring, enabling upbeat detection and mitigation of memory hotspots.

D. Vector Dot :

In vector dot scheduling POLYPHONY is used to virtualize the system. An end-to-end view of the SAN including usage characteristics and performance is given by POLYPHONY. It ameliorates the utilization of resource includes data canter network bandwidth, I/O bandwidth and physical servers. Here is virtualization storage migration [3] is done instead of migration of machine migration. To calibrate current utilization of resource, extended vector product (EVP) is used.

E. Term -Green Scheduling Algorithm :

Green scheduling is used to arbitrate which server to be in running state. It will automatically turn on or turn off [4] the physical machines based on load and allocated virtual machines. Server should be in four states: OFF, ON, SHUTTING, RUNNING.

F. Term -Benchmark Algorithm :

The benchmark algorithm is used to contrast the performance of different empirical resource allocation algorithms. Performance of algorithms is based on CPU utilization, that take in consideration the utilization threshold in addition to VM migration .CPU utilization is stumble upon Mean absolute deviation (MAD) and inter-quartile range.

G. Term -Control Algorithm :

In the control algorithm a different set of techniques are used to forecast unsteady workloads of the system. In this two set of process are used Optimal Markov Host Overload Detection (MHOD-OPT) with Markov Host Overload Detection (MHOD).

H. Priority Algorithm :

In a cloud computing environment, multiple customers are submitting the appeal for the same resource on different constraint. For example in a high performance applications which deal with scientific imitations such as weather forecasting, monsoon prediction etc. Which entails large amount of computing resources such as processors, servers, storage etc? Most of the users are requesting the resources among these. So in such a situation cloud administrator get confuse to which user among most has to allocate the resource which is requested by more than one user. The proposed priority algorithm aids cloud admin to commit priority among the users and designate resources efficiently according to priority. This resource allocation technique is more efficient than grid and utility computing because in those systems there is no priority among the user request. And here cloud administrator is arbitrarily taking decision and he is able to providing priority to those user who have proposed their job first which is based on first come first serve technique. The cloud admin can able to take verdict easily based on different parameters to decide priority among different user request.

In order to sprint specific model large computational resources such as server, memory, processors, software etc are demanded. There are two types of job execution some are executing concurrent manner and some are sequential manner. In such a situation job type is also very significant parameter. In a cloud environment there are different types of users like the user is internal (internal cloud) to a cloud or he is external to cloud (public cloud) is also one more important parameter to be consulted during job submission. So thus the developed priority algorithm deliberates in detail how efficiently it will aid cloud admin to compute priority among the user requests.

ISSN 2348 – 7968

IV. SKEWNESS:

Skewness [2] is used to gauge the unevenness resource utilization of resources of a server. To minimizing skewness, One can merge several types of workloads that the efficient utilization of resources and improve performance. The Load predication in skewness algorithm consummates periodically to assign resources based on the predicted future resource demands of VMs. Thus brew prediction based on the past external behaviors of VMs. To measure the load every minute and forecast the load in the next minute.

.

A. Skewness Calculation Formula

Let n be the number of resources we consider and ri be the utilization of the i-th resource [2]. We define the resource skewness of a server p as

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

By minimizing the skewness, combine different types of workloads nicely and improve the overall utilization of server resources.

We represent skewness in following simple form,

P= r/R -1

Where
p = skewness to determine the usage level of memory by a VM
r – Memory usage of the current VM under evaluation
R – Average of memory usage by all VMS
If p = 0, the Memory utilization is normal
P >0, then the memory usage is more compared to others VMS
P<0, then the memory usage is less compared to other VMs

B. Analysis of skewness algorithm

The skewness [2] algorithm consists of three parts: load prediction, hot spot mitigation, and green computing. Let n and m be the number of PMs and VMs in the system, respectively. The number of resources (CPU, memory, I/O, etc.) that need to be considered is usually a small constant.

Algorithm allocates resources based on the demands of VM. Here we define the server a hotspot and if the utilization exceeds the above the hot threshold then it shows that the server is overloaded and Vm's are moved away. The temperature is about to zero when the server is not a hot spot.

We define a cold spot when the utilization of there sources are below the cold threshold which indicates that the server is idle and it has to be turned of in order to save energy. This is done when mostly all servers are actively used below the green computing threshold else it is made inactive.

C. Load Predictioin:

Load prediction [2] algorithm execute periodically in skewness algorithm to predict feature resource demand on VM. The proposed system defines a server as a hot spot if the utilization of any of its resources is above a hot threshold. This indicates that the server is overloaded and hence some VMs running on it should be migrated away.

D. Hot spot:

The proposed system presents a server as a hot spot if the utilization of any of its resources is above a hot threshold. This indicates that the server is overloaded and hence some VMs running on it should be migrated away on another PMs.

E. Cold spot:

The proposed system presents a server as a cold spot if the utilizations of all its resources are below a cold threshold. This shows that the server is mostly in idle state and there is a need that PM get turn off to save energy.

F. Hot spot mitigation

The arrangement of the hot spots in sorted order so that one can eliminate them else keeps the temperature low. The motive is to move away the VM's that can reduce the server's temperature. Among all, select the one which can reduce Skewness.

G. Green computing

By the purpose of saving energy we have to turn off some servers when the resource utilization of active servers is too low. This is exercised in green computing algorithm. The challenge here is to reduce the number of active servers during low load without decreasing performance either now or in the future. The algorithm is practiced when utilization of all active servers are below the green computing threshold.

When the resource utilization of all active servers is too low, some of them can be turned off to save energy. This is practiced in our green computing algorithm. The challenge here is to reduce the number of active servers during low load without decreasing performance either now or in the future. The algorithm is appealed when the average utilizations of all resources on active servers are below the

green computing threshold. We sort the cold spots in ascending order based on their memory size. Since the system needs to emigrate away all its VMs before we can shut down an under-utilized server, we describe the memory size of a cold spot as the collective memory size of all VMs running on it. Keep in mind that our model guesses all VMs connect to share back-end storage. Hence, the cost of a VM live migration is ascertained mostly by its memory footprint.

Advantages of Skewness Algorithm

1. Skewness is used to quantify the unevenness in the utilization of multiple resources on a server which improves the performance.
2. Skewness algorithm is to mix workloads with different resource requirements together so that the overall utilization of server capacity is improved. By above technique proposed algorithm handles a mix of CPU, memory, and network intensive workloads.

V. CONCLUSION

This paper presents the hypothetical study of heterogeneous dynamic resource allocation techniques in cloud computing environment. Cloud computing can resolve obscure set of tasks in shorter time by proper resource utilization. The proposed system multiplexes virtual to physical resources adaptively based on the changing needs. To use the skewness to merge VMs with different resource characteristics judiciously so that the capacities of servers are best utilized. The algorithm attains both overload avoidance and green computing for systems with multi-resource constraints.

REFERENCES

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

[2] Zhen Xiao, Senior member, IEEE, weijia song and Qi chen "*Dynamic Resource allocation using Virtual Machines For Cloud Computing Environment* ," IEEE Transaction on parallel and distributed systems, vol.24, No.6 june 2013.

[3] Daniel Warneke and Odej Kao, "Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud", IEEE Transactions On Parallel And Distributed Systems, vol 22,no.6, pp. 985-997,2011.

[4] Gong Chen, Wenbo He, Jie Liu, Suman Nath, Leonidas Rigas, Lin Xiao, and Feng Zhao,"Energy-Aware Server Provisioning and Load Dispatching for Connection-Intensive Internet Services", Proceedings. USENIX Symposium. Networked Systems Design and Implementation, 2008.

[5] Sheng Di, Cho-Li, "Dynamic Optimization of Multi-attribute Resource Allocation in Self-Organizing Clouds", IEEE Transactions on Parallel and Distributed Systems, vol. 24, pp. 205-225, 2012.

[6] "Adaptive Management of Virtualized Resources in Cloud Computing Using Feedback Control," in First International Conference on Information Science and Engineering, April 2010, pp. 99-102.

[7] Ying Song, Yuzhong Sun, Member, IEEE, and Weisong Shi, Senior Member, IEEE*A" Two-Tiered On-Demand Resource Allocation Mechanism for VMBased Data Centers"*, IEEE transactions on services computing, vol. 6, no. 1, january-march 2013.

[8] Timothy Wood, Prashant Shenoy, Arun Venkataramani, and Mazin Yousif ,"Black-box and Gray-box Strategies for Virtual Machine Migration", Proceedings of National conference on distributed Systems, pp.229-242,2012.

[9] T. Wood ,P. Shenoy, A.Venkataramani and M.Yousif ,"Black-Box and Gray-Box Strategies for Virtual Machine Migration", Proc. Symposium. Networked Systems Design and Implementation, 2007.