



Efficient Resource Provisioning Using Virtualization Technology in Cloud Environment

J.Praveen Immanuel Paulraj and R.KannigaDevi

PG Student, Dept of Computer Science Engg, Kalasalingam University, Krishnankoil, Virudhunagar, India.

Assistant Professor, Dept of Computer Science Engg, Kalasalingam University, Krishnankoil, Virudhunagar, India

Abstract - Cloud computing is the practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer. Resource Allocation is the process of assigning available resources to the needed cloud applications over the internet. In computing environment, virtualization means creating a virtual version of a device or a resource, such as server, storage device, network or even an operating system in which the framework divides the resource into one or more execution environments. In this project, we present a system 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. After the resources are allocated in virtual machines, the measure of "skewness" is used to identify the unevenness in the multidimensional resource utilization of a server. To minimize skewness, different types of workloads are combined to improve the overall utilization of server resources. This will provide an efficient method for allocating resources in Cloud environment.

Index Terms: Cloud Computing, Resource Allocation, Virtualization, Skewness.

I.INTRODUCTION

Cloud computing is a model that enables ubiquitous, convenient, on-demand network access for a shared pool of configurable computing resources which can be rapidly provisioned and released with minimal management effort or service provider interaction. Cloud computing is a general term used which is used to describe a new class of network based computing that takes place over the Internet, basically a step from utility computing. It is a cumulative of numerous attempts at large scale computing with seamless access to virtually limitless resources. The user becomes responsible only for maintaining a very simple computer and makes the cloud services to take the rest of the operations. Once the user connected to the cloud services provider, they can access the application as it is installed on providers' computer. It is an emerging computing technology that is rapidly consolidating in the development and deployment of an increasing number of distributed applications. Nowadays it has become quite popular among the community of cloud users by offering of variety of resources. Some applications can access a large network of computing resources that are deployed and managed by a cloud computing platform. Developers obtain the advantages of a managed computing platform without having to commit resources for designing, building and maintaining the network.

The cloud computing technology makes the resource as a single point of access to the client and is implemented as pay per usage. As there are various advantages in cloud computing such as abstracted infrastructures, completely virtualized environments, equipped with dynamic infrastructures, free of software's and hardware installations. This allocation of resources must be made efficiently that maximizes the system utilization and overall performance. Therefore scheduling should be made in such a way that the resource should be utilized efficiently. As it is the next generation of technology which unifies everything into one. It is an on demand service as it offers dynamic flexible resource allocation for reliable and guaranteed service. In cloud computing environment multiple cloud users can request number of cloud services simultaneously. So there must be a provision for all resources that are made available to requesting in user efficient manner to satisfy their needs. Dynamic resource allocation in cloud computing is based on topology aware resource allocation, linear scheduling strategy for resource allocation and dynamic resource allocation for parallel data processing.

I. DESIGN GOALS

The resource needs of virtual machines that are heterogeneous due to the diverse set of applications they run and vary time as the workloads grow and shrink. The main goal of using this algorithm is to avoid overload. The capacity of a physical machine should be sufficient to satisfy the resource needs of all virtual machines running on it. Otherwise the private machine is overloaded and can lead to degrade the performance of its virtual machine. The design and implementation of an automated resource management system that achieves a good balance between the two goals resources is developed for system allocation which avoids the overload in the system effectively while minimizing the number of servers used. The concept of skewness is introduced to measure the uneven utilization of a server. By minimizing the skewness we can improve the overall utilization of servers in the face of multidimensional resource constraints. Also a load prediction algorithm that can capture the future usages of applications accurately without looking inside the virtual machine. The algorithm can capture the rising trend of resource patterns and helps to reduce the placement churn significantly.

II. RELATED WORKS

In cloud platforms resource allocation takes place at two levels. When an application is being uploaded in

the cloud environment the load balancer assigns the requested instances to the physical computer, attempting to balance the computational load of multiple applications to across physical computer.

Ronak Patel et.al., [1] Suggest that when an application receives multiple incoming requests these requests should be assigned to a specific application instance to balance the computational load across a set of instances of the same application. Application designers can direct requests to instance in specific availability.

Timothy wood et.al., [2] explain that in cloud computing resources allocation is the process of assigning available resources to the required cloud applications over the internet. Resource allocating starves services if the allocation is not managed precisely. Resource provisioning solves the problem by allowing the service providers to manage the resources for each individual module.

Hai Jin et.al., [3] propose that the resource allocation strategy integrates cloud provider activities for utilizing and allocating scarce resources within the limit of cloud environment. It generally requires the type and amount of resources needed by each application to a complete user job.

Rajkumar Buyya et.al., [4] explain the order and time of allocation of resources consists of physical and virtual resources. The physical resources are shared across multiple compute requests through virtualization and provisioning. The dynamic resource allocation in cloud computing have come up with new ways of facing many challenges. Generally virtualization describes the separation of a resource or request for a service from the underlying physical entity. It allows a single physical resource appear as multiple logical resources.

Sivadon Chaisiri et.al., [5] suggest that the storage virtualization is the actual storage space that use to be offered by storage area network as virtualized. Virtualizing storage space means virtualization of virtual machine disk files. It becomes useful for disaster recovery because with a virtualized infrastructure it is easy to store across over another location or even s different geographical regions.

Zibin Zheng et.al., [6] propose that the network virtualization creates networks where each application sees its own logical networks independent of their physical network. A virtual LAN is an example of a network virtualization which provides an easy, flexible and less expensive ways to manage the networks.

Sheng Di et.al., [7] explain the virtual networks make large networks more manageable by enabling a centralized configuration of devices located in physically diverse locations.

Dominique guinard et.al., [8] propose that computing network virtualization is the process of

combining hardware's and software's network resources and network functionalities into a single, software based administrative entity which is known as a virtual network.

Hong Xu et.al., [9] explain that the network virtualization involves platform virtualization combing with resource virtualization. IT is categorized as external combining many or part of networks into a virtual unit or internal providing network like functionality to the software containers on a single system. When Load prediction is disabled the algorithm simply uses the last observed load in its decision making without prediction.

Daniel Warneke et.al., [10] suggest the algorithm that tries to consolidate a physical machine as soon as its load drops below the threshold value. With prediction the algorithm correctly foresees that the load of the physical machine will increase above the threshold shortly and hence takes no action. It also reduces the placement churns by avoiding the unnecessary migrations due to the temporary load fluctuations. Consequently the number of migrations in the system with load prediction is smaller than the system without prediction.

III. SYSTEM DESIGN

This system proposes the use of virtualization technology for efficient resource provisioning in cloud computing. The virtualization technology is used in order to reduce physical server cost. In this storage resource has been taken into consideration where the virtual machine is created in the data centers to store the data. As the amount of data in one virtual machine increase, a new virtual machine is created and the data are migrated in the same data center when one particular data center is overloaded, the data are migrated to the other data centers. To avoid uneven utilization of virtual machines a skewness is measured to find the unevenness in the resource utilization in each virtual machine in data centers. Then a Load prediction algorithm is used to predict the future resource that needs efficient resource allocation to the incoming requests. To minimize the skewness different types of workloads are combined and thus improve the overall utilization of server resources.

A. VM MIGRATION

Virtualization can provide significant benefits in data centers by enabling virtual machine migration to eliminate hotspots. Server resources in a data center are multiplexed across multiple applications where each server runs one or more applications and application components may be distributed across multiple servers. The possible approach for reducing management complexity is to employ virtualization. In this the applications run on one or more virtual servers which are

mapped onto each physical server in the system. The virtualization of data center resources provides numerous benefits. It enables the applications isolation since malicious or greedy applications cannot impact on other applications which are co-located on to the same physical server. It also enables the server consolidation and provides better multiplexing of datacenter resources across the applications.

The biggest advantage of employing virtualization is the ability to flexibly remap the physical resources to the virtual servers in order to handle workload dynamics. An increase in the workload can be handled by increasing the resources allocated to the virtual servers. If idle resources are available on the physical servers or by simply migrating the virtual servers to a less loaded physical servers. Migration is transparent to the applications and all modern virtual complicated by the need to consider multiple resources for each application and physical server. Since migration is transparent to applications executing within virtual machines. The migration manager assumes that the virtual machine monitor implements a migration mechanism that is transparent to applications and uses this mechanism to automate migrate decisions.

Generally virtualization provides significant benefits in data centers by enabling virtual machine migration to eliminate hotspots. The live migration of virtual machines across distinct physical hosts provides a significant new benefit for administrations of data centers and clusters. The capability of migrating live virtual machine along distinct phsical hosts provides a significant new benefit for multiple VM-based environment in many scenarios. The combination of virtualization and migration significantly improves manageability of data centers and clusters.

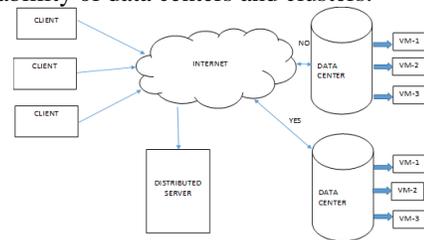


Fig 1: Data Migration

B. SKEWNESS

The concept of skewness is to quantify the unevenness in the utilization of multiple resources on a server. Let n be the no of resources we consider and rⁱ be the utilization of the ith resource. We define the skewness of the server p as

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

Where r is the average utilization of all resources for server p . All types of 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. Different types of resources can have different thresholds. The goal of skewness algorithm is to aim workloads with different resource requirements together so that overall utilization of server capacity is improved. We inject the networks load by sending the virtual machines as servers of network packets. The memory intensive applications are hence created by allocating memory on demand.

C. RESOURCE ALLOCATION

The allocated computer resources may be inadequate for big parts of the submitted job and unnecessarily increase processing time and cost. Particular tasks of a processing job can be assigned to different types of virtual machines which are automatically initiated and terminated during the job execution. The processing of frameworks typically assure the resources which they manage consists of a state set of homogeneous compute nodes. Although they are designed to deal with the individual node failures they consider the number of available machines to be constant especially when scheduling the processing jobs which are in the execution. In the cluster the compute nodes are typically interconnected through a physical high performance network. Current data processing frameworks offer to know about the network hierarchy and data sent from one node to the other has to traverse a few network switches as possible.

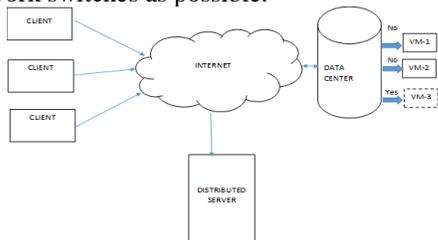


Fig 2: Virtual machine creation

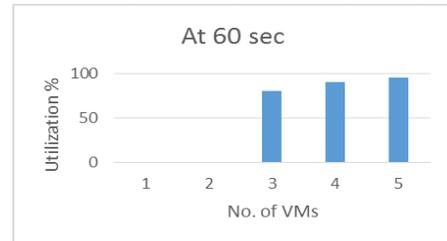
In a cloud topology the information is not exposed to the customers. Since the nodes that involved in processing a data have to transfer tremendous amounts of data through the networks. Virtual machines are migrated for administrative purposes between different locations inside the data centers without any notifications rendering the relevant network infrastructures. The only way to ensure the locality between tasks of a processing job is to execute the tasks of a processing job on the same virtual mechanisms in the cloud environment. It involves allocating fewer but more powerful virtual machines with multiple CPU cores. Data locality can be ensured by scheduling these tasks to run on a virtual machine with eight cores instead of a distinct single core

machines. With virtual machine technology we are able to the multiple operating systems on the same hardware and allow task execution over its virtual machines substrates without performance interferences. Most cloud servers built on top of centralized architectures may suffer denial of service attacks, unexpected outages and limited pooling of computational resources. Resource management becomes a major challenge for the cloud environment. Cloud operators have a variety of distinct resource management objectives to achieve virtual machines of a cloud and also impose diverse resource requirements that need to be accomplished as they run different applications owned by individual clients.

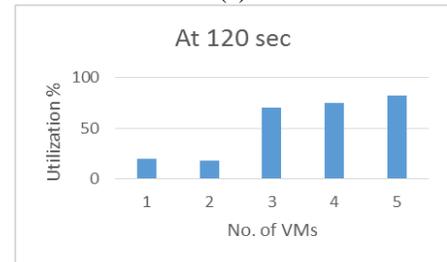
V.EXPERIMENTAL SETUP

We simulated our experiment using CloudSim. CloudSim is a Cloud computing modeling and simulation tool. It provides Cloud computing researchers to conduct new research approaches with a comprehensive experimental tool. It also supports modeling and simulation of large scale Cloud environments, which includes power management, data centers, performance, computing nodes, virtual machine provisioning and resource provisioning.

VI.PERFORMANCE EVALUATION



(a)



(b)

In our experiment, we used 2 physical machines with 5 virtual machines in it. In that PM 1 has 2 VMs and PM 2 has 3 VMs. Initially, the storage resources on VMs 1 and 2 are kept low, that is no load given to PM1 and the VMs 3,4 and 5 are overloaded storage. It is observed in Fig. 3(a). At 120 sec and 180 sec, fig. 3(b) and 3(c) shows the load in virtual machines in PM 2 are

gradually migrated to the VMs in PM 1. Then at time 240 sec, the storage resources are evenly balanced in all the available VMs is seen in fig. 3(d). Thus the under loaded VMs are utilized properly by balancing the load among the VMs.

Fig. 3 shows how the algorithm spreads the VMs to other PMs over time. It also shows the storage utilization of VMs across the PMs are balanced well as the load is increased.

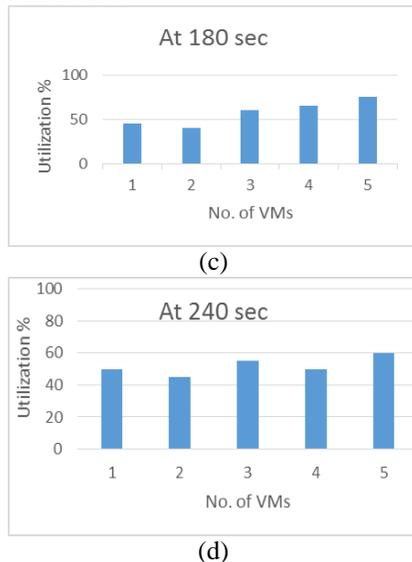


Fig 3: Load balancing between VMs

VII.CONCLUSION

We have presented the design, implementation and evaluation of a resource management system for cloud computing services. Our system multiplexes the virtual resources to the physical resources adaptively based on the changing demand. We use skewness metric to combine the virtual machines with different resource characteristics in appropriated algorithm for achieving both overload avoidance and green computing for the systems with multi resource constraints.

REFERENCES

- [1] Ronak Patel, Sanjay Patel, "Survey on Resource Allocation Strategies in Cloud Computing", International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 2, February- 2013.
- [2] T. Wood, P. Shenoy, A. Venkataramani, and M. Yousif, "Black-Box and Gray-Box Strategies for Virtual Machine Migration," Proc. Symp. Networked Systems Design and Implementation (NSDI '07), Apr. 2007.
- [3] Hai Jin, Haikun Liu, Xiaofei Liao, Liting Hu, Peng Li, "Live Migration of Virtual Machine Based on Full-System Trace and Replay", Services Computing Technology and System Lab Cluster and Grid Computing Lab School of Computer Science and Technology.
- [4] Arch Rajkumar Buyya, Saurabh Kumar Garg, and Rodrigo N. Calheiros, "SLA-Oriented Resource Provisioning for Cloud Computing: Challenges", and Solutions, 2011.
- [5] Sivadon Chaisiri, Bu-Sung Lee, "Optimization of Resource Provisioning Cost in Cloud Computing", IEEE transactions on services computing, vol. 5, no. 2, april-june 2012.
- [6] Zibin Zheng, Xinmiao Wu, Yilei Zhang, Michael R. Lyu, Jianmin Wang, "QoS Ranking Prediction for Cloud Services", IEEE transactions on parallel and distributed systems, vol. 24, no. 6, june 2013.
- [7] Sheng Di, Cho-Li Wang, "Dynamic Optimization of Multiattribute Resource Allocation in Self-Organizing Clouds", IEEE transactions on parallel and distributed systems, vol. 24, no. 3, march 2013.
- [8] Dominique Guinard, Vlad Trifa, Stamatis Karnouskos, Patrik Spiess, Domnic Savio, "Interacting with the SOA-Based Internet of Things: Discovery, Query, Selection, and On-Demand Provisioning of Web Services", IEEE transactions on services computing, vol. 3, no. 3, july-september 2010.
- [9] Hong Xu, Baochun Li, "Anchor: A Versatile and Efficient Framework for Resource Management in the Cloud", IEEE transactions on parallel and distributed systems, vol. 24, no. 6, june 2013.
- [10] Daniel Warneke, 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, june 2011.
- [11] Hyukho Kim, Woongsup Kim, Yangwoo Kim, "Predictable Cloud Provisioning Using Analysis of User Resource Usage Patterns in Virtualized Environment", Volume 121, 2010.
- [12] Sivadon Chaisiri, Bu-Sung Lee and Dusit Niyato "Optimization of Resource Provisioning Cost in Cloud Computing", 2012.
- [13] N. Bobroff, A. Kochut, and K. Beaty, "Dynamic Placement of Virtual Machines for Managing SLA Violations," Proc. IFIP/IEEE Int'l Symp. Integrated Network Management (IM '07), 2007.
- [14] Zhen Xiao, Senior Member, IEEE, Weijia Song, and Qi Chen, "Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment", PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 24, NO. 6, JUNE 2013.
- [15] A. Singh, M. Korupolu, and D. Mohapatra, "Server-Storage Virtualization: Integration and Load Balancing in Data Centers," Proc. ACM/IEEE Conf. Supercomputing, 2008.
- [16] "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.
- [17] Jiayin Li, Meikang Qiu, Jian-Wei Niu, Yu Chen, Zhong Ming, "Adaptive Resource Allocation for Pre-emptible Jobs in Cloud Systems," in 10 th International Conference on Intelligent System Design and Application, Jan. 2011, pp. 31-36.
- [18] Yazir Y.O., Matthews C., Farahbod R., Neville S., Guitouni A., Ganti S., Coady Y., "Dynamic resource allocation based on distributed multiple criteria decisions in computing cloud," in 3 rd International Conference on Cloud Computing, Aug. 2010, pp. 91-98.
- [19] 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.
- [20] Shi J.Y., Taifi M., Khreishah A., "Resource Planning for Parallel Processing in the Cloud," in IEEE 13 th International Conference on High Performance and Computing, Nov. 2011, pp. 828-833.
- [21] Aoun R., Doumith E.A., Gagnaire M., "Resource Provisioning for Enriched Services in Cloud Environment," IEEE Second International Conference on Cloud Computing Technology and Science, Feb. 2011, pp. 296-303.
- [22] T. Erl, "Service-oriented Architecture: Concepts, Technology, and Design", Upper Saddle River, Prentice Hall, 2005.
- [23] F. Chong, G. Carraro, and R. Wolter, "Multi-Tenant Data Architecture", Microsoft Corporation, 2006.
- [24] Daniel Warneke, Member, IEEE, and Odej Kao, "Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud", 2012.

- [25] Jia Rao, Member, IEEE, Yudi Wei, Student Member, IEEE, Jiayu Gong, Member, IEEE, and Cheng-Zhong Xu, “*QoS Guarantees and Service Differentiation for Dynamic Cloud Applications*”, 2012.
- [26] Chao-Tung Yang, Hsiang-Yao Cheng, Kuan-Lung Huang, “*A Dynamic Resource Allocation Model for Virtual Machine Management on Cloud*”, International Conference, GDC 2011, Held as Part of the Future Generation Information Technology Conference, FGIT 2011, Jeju Island, Korea, December 8-10, 2011.
- [27] Jian Cao, Yihua Wu, Minglu Li, “*Energy Efficient Allocation of Virtual Machines in Cloud Computing Environments Based on Demand Forecast*”, 7th International Conference, GPC 2012, Hong Kong, China, May 11-13, 2012.
- [28] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, Xen, “*The Art of Virtualization*,” Proc. ACM Symp. Operating Systems Principles (SOSP '03), Oct. 2009.
- [29] M. Nelson, B.-H. Lim, and G. Hutchins, “*Fast Transparent Migration for Virtual Machines*,” Proc. USENIX Ann. Technical Conf., 2010.
- [30] P. Padala, K.-Y. Hou, K.G. Shin, X. Zhu, M. Uysal, Z. Wang, S. Singhal, and A. Merchant, “*Automated Control of Multiple Virtualized Resources*,” Proc. ACM European conf. Computer Systems (EuroSys '09), 2009.