A Review Paper On Dynamic Resource Allocation In Cloud Environment

Rekha Rani[#1], Sumit Dahiya[#2]

[1] Assistant Professor , Department of Computer Science
South Point Institute of Engineering & Management affiliated to
Deenbandhu Chhotu Ram University of Science & Technology (DCRUST), Sonepat

malik.rekhu@gmail.com

[2] Student, M TECH(CSE)
South Point Institute of Engineering & Management affiliated to
Deenbandhu Chhotu Ram University of Science & Technology (DCRUST), Sonepat

dahiyasumit3148@gmail.com

Abstract— Cloud computing has become a new age technology that has got huge potentials in enterprises and markets. Clouds can make it possible to access applications and associated data from anywhere. Resource allocation is one of the important challenges in cloud computing environment. It depends on how to allocate the resource to the particular task. Resource allocation can be done by two methods. One of the methods statically allocates the resources and other dynamically allocates the resources. With static resource allocation the cloud user has to make prior request for the resources. With dynamic resource allocation the cloud resources are requested by the cloud user on the fly or as and when the application needs. This paper provides review of various dynamic resource allocation techniques.

Keywords— Cloud Computing, Resource Allocation, Resource Provisioning

I. INTRODUCTION

Cloud computing [1] comprises of 2 components ―the front end and the back end. The front end includes client's devices and applications that are required to access cloud. And the back end refers to the cloud itself. The whole cloud is administered by a central server that is used to monitor client's demands (Fig 1).
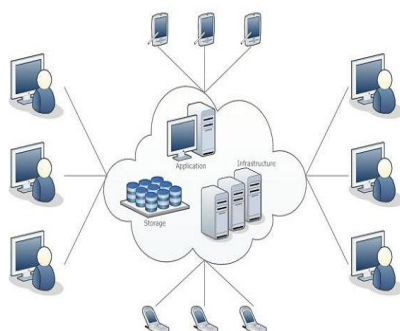

Figure 1: Cloud Computing

Cloud Computing has numerous advantages [2]. Some of them are listed below:

▪ One can access applications as utilities, over the Internet.
▪ Manipulate and configure the application online at any time.
▪ It does not require to install a specific piece of software to access or manipulate cloud application.
▪ Cloud Computing offers online development and deployment tools, programming runtime environment through **Platform as a Service model**.

▪ Cloud resources are available over the network in a manner that provides platform independent access to any type of clients.
▪ Cloud Computing offers **on-demand self-service**. The resources can be used without interaction with cloud service provider.

▪ Cloud Computing is highly cost effective because it operates at higher efficiencies with greater utilization. It just requires an Internet connection.

▪ Cloud Computing offers load balancing that makes it more reliable.

One of the main services of cloud environment is resource allocation. Resource allocation is process of assigning the available resources in an economic way and efficient and effective way Resource allocation is the scheduling of the available resources and available activities required by those activities while taking into consideration both the resource availability and the project time.

International Journal of Research

Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

Resource allocation can be done by two methods. One of the methods statically allocates the resources and other dynamically allocates the resources. With static resource allocation the cloud user has to make prior request for the resources. With dynamic resource allocation the cloud resources are requested by the cloud user on the fly or as and when the application needs. This paper provides review of various dynamic resource allocation techniques. This paper provides review of different dynamic resource allocation strategies in cloud environment.

II.   RESOURCE ALLOCATION STRATEGY

Resource Allocation Strategy (RAS) [3] is all about integrating cloud provider activities for utilizing and allocating scarce resources within the limit of cloud environment so as to meet the needs of the cloud application. It requires the type and amount of resources needed by each application in order to complete a user job. The order and time of allocation of resources are also an input for an optimal RAS. An optimal RAS should avoid the following criteria as follows:

a) Resource contention situation arises when two applications try to access the same resource at the same time.
b) Scarcity of resources arises when there are limited resources.
c) Resource fragmentation situation arises when the resources are isolated.
d) Over-provisioning of resources arises when the application gets surplus resources than the demanded one.
e) Under-provisioning of resources occurs when the application is assigned with fewer numbers of resources than the demand.

From the perspective of a cloud provider, predicting the dynamic nature of users, user demands, and application demands are impractical. For the cloud users, the job should be completed on time with minimal cost. Hence due to limited resources, resource heterogeneity, locality restrictions, environmental necessities and dynamic nature of resource demand, we need an efficient resource allocation system that suits cloud environments. Cloud resources consist of physical and virtual resources. The physical resources are shared across multiple compute requests through virtualization and provisioning. The request for virtualized resources is described through a set of parameters detailing the processing, memory and disk needs. Provisioning satisfies the request by mapping virtualized resources to physical ones. The hardware and software resources are allocated to the cloud applications on-demand basis. For scalable computing, Virtual Machines are rented. The complexity of finding an optimum resource allocation is exponential in huge systems like big clusters, data centres or Grids. Since resource demand & supply can be dynamic and uncertain.

III.   LITERATURE REVIEW

Several works related to dynamic resource allocation were done in the past. Some these work are described below.

Valérie Danielle Justafort et. al. [4] discusses about Performance-Aware Virtual Machine Allocation. Cloud Computing has major challenges among which virtual machine (VM) placement is considered as one of the crucial problems. From the cloud provider"s perspective, this process often consists in choosing the best convenient physical hosts with regards to energy efficiency, resource utilization and revenues maximization. However, to secure the loyalty of their clients, cloud providers should also consider the applications performance. In this paper, VM placement approach is proposed tackling the applications performance concern, in an inter cloud environment. VM placement is stated as a mixed integer programming problem which aims to maximize a global utility function considering VM bandwidth requirements and network latencies.

Bo Yin et. al. [5] discusses about multi-dimensional resource allocation. In this paper, study on the resource allocation at the application level is done, instead to map the physical resources to virtual resources for better resource utilization in cloud computing environment. A multi-dimensional resource allocation (MDRA) scheme for cloud computing is proposed that dynamically allocates the virtual resources among the cloud computing applications to reduce cost by using fewer nodes to process applications. In this model, a two-stage algorithm is adopted to solve this multi-constraint integer programming problem. The algorithm can dynamically reconfigure the virtual machines for cloud applications according to the load changes in cloud applications by assigning new applications on the working node instead of opening a new node. Experiment results show that this algorithm can save resources and increase resource utilization as well as centralize working nodes. MDRA provide resource to deal with requirements more steadily. Moreover, the proposed algorithm in the long run can save power efficiently when the demand of user is decreasing.

Amit Nathani et. al. [6] have proposed policy based resource allocation in IaaS cloud. Most of the Infrastructure as a Service Cloud (IaaS) uses simple resource allocation techniques like immediate and best effort. Immediate Resource Allocation Policy allocates the resource if available otherwise the

request is rejected. Best effort policy also resources if currently available otherwise the request is placed in the FIFO queue. So it is not possible for the Service provider to satisfy all the requests as there is finite number of resources at a time. So Haizea is a resource lease manager that addresses these issues by using complex resource allocation strategies. Haizea uses resource leases as resource abstraction and implements these leases by allocating Virtual Machines (VMs). Haizea supports four resource allocation policies: Immediate, best effort, advanced reservation and deadline sensitive. Among the four, deadline sensitive leases by Haizea supports minimal rejection of leases and using this policy, maximum resource utilization is possible. Dynamic Planning based scheduling algorithm is implemented in Haizea which can admit leases and prepare a schedule whenever a new lease can be accommodated.

Sharrukh Zaman et. al. [7] discusses online mechanism for dynamic VM provisioning and allocation. Current cloud provider allocates virtual machine instances via fixed price-based or auction based mechanisms. The limitation in these mechanisms is they are all offline mechanism and they need to collect information and invoked periodically. This limitation is addressed by designing an online mechanism by dynamically provisioning and allocating Virtual machine Instances in Clouds. The algorithm called Mechanism for online VM provisioning and allocation (MOVMPA) is invoked as soon as the user makes a request or some VM instances already become available again. When invoked, the mechanism selects users who would be allocated VM instances they requested for and ensures that those users will continue using for the entire period they requested for.

Mayank Mishra et. al. [8] proposed dynamic resource management using Virtual Machine Migrations. Virtual machine related features such as flexible resource provisioning, and isolation and migration of machine state have improved efficiency of resource usage and dynamic resource provisioning capabilities. Live virtual machine migration transfers "state" of a virtual machine from one physical machine to another, and can mitigate overload conditions and enables uninterrupted maintenance activities. This paper focuses on the details of virtual machine migration techniques and their usage toward dynamic resource management in virtualized environments.

Gihun Jung et. al. [9] discusses about agent-based adaptive resource allocation. Since both cloud users and data centers of a cloud service provider can be distributed geographically, the provider needs to allocate each user request to an appropriate data center among the distributed data centers, so that the users

can satisfy with the service in terms of fast allocation time and execution response time. In this paper, an adaptive resource allocation model is proposed that allocates the cloud user"s job to an appropriate data center. The method to adaptively find a proper data center is based on two evaluations: 1) the geographical distance (network delay) between a cloud user and data centers, and 2) the workload of each data center. The proposed model can successfully allocate cloud users" requests to the data center closest to each consumer. Also, the proposed model shows a better response time for allocation than related resource allocation models.

Xiao-Jun Chen et. al. [10] discusses about resource reconstruction algorithms for on-demand allocation. Resource reconstruction algorithms are used to solve the problem of on-demand resource allocation and can be used to improve the efficiency of resource utilization in virtual computing resource pool. The idea of resource virtualization and on the analysis of resource status transition these algorithms are used. These algorithms are designed to determine the resource reconstruction types and they can achieve the goal of on-demand resource allocation through three methodologies: resource combination, resource split, and resource random adjustment. These algorithms can do the resource adjustment on lower cost and form the logical resources to match the demands of resource users easily.

Ilhem Fajjari et. al. [11] discusses about optimized dynamic resource allocation algorithm for Cloud"s Backbone Network. Because of the unpredictable and varying user"s demands, a flexible and intelligent resource allocation scheme is necessary. This paper tackles the fundamental challenge of efficient resource allocation within cloud"s backbone network. The ultimate goal is to satisfy the cloud"s user requirements while maximizing cloud provider"s revenue. The problem consists in embedding virtual networks within substrate infrastructure. A new dynamic adaptive virtual network resource allocation strategy named Backtracking-VNE is investigated to deal with the complexity of resource provisioning within cloud network. The proposal coordinates virtual nodes and virtual links mapping stages to optimize resources usage.

K C Gouda et. al. [12] proposed a new approach that allocates resource with minimum wastage and provides maximum profit. The developed resource allocation algorithm is based on different parameters like time, cost, No of processor request etc. The developed priority algorithm is used for a better resource allocation of jobs in the cloud environment used for the simulation of different models or jobs in

an efficient way. After the efficient resource allocation of various jobs, an evaluation is being carried out which illustrates the better performance of cloud computing with profit. A performance study of all the algorithms in various systems and case studies are also presented.

N. Krishnaveni et. al. [13] describe that Cloud computing becomes quite popular among cloud users by offering a variety of resources. This is an on demand service because it offers dynamic flexible resource allocation and guaranteed services in pay as-you-use manner to public. In this paper, we present the several dynamic resource allocation techniques and its performance. This paper provides detailed description of the dynamic resource allocation technique in cloud for cloud users and comparative study provides the clear detail about the different techniques.

Sukhpal Singh et. al. [14] emphasis on the development of energy based resource scheduling framework and present an algorithm that consider the synergy between various data center infrastructures (i.e., software, hardware, etc.), and performance. In specific, this paper proposes (a) architectural principles for energy efficient management of Clouds; (b) energy efficient resource allocation strategies and scheduling algorithm considering Quality of Service (QoS) outlooks. The performance of the proposed algorithm has been evaluated with the existing energy based scheduling algorithms. The experimental results demonstrate that this approach is effective in minimizing the cost and energy consumption of Cloud applications thus moving towards the achievement of Green Clouds.

Riddhi Patel et. al. [15] describe Modified best fit decreasing algorithm (MBFD) is discus for reduce the energy consumption and modified the MBFD algorithm as Energy aware Best Fit Decreasing (EABFD) algorithm. Resource allocation is one of the important challenges in cloud computing environment. It depends on how to allocate the resource to the particular task. Resource allocation can be done by two methods. One of the methods statically allocates the resources and other dynamically allocates the resources. The other challenges of resource allocation are meeting customer demands and application requirements. One of the challenges posed by cloud applications is Quality-of-Service (QoS) management, which is the problem of allocating resources to the application to guarantee a service level along dimensions such as performance, availability and reliability.

IV. CONCLUSION

A cloud environment is one of the most shareable environments where multiple clients are connected to the common environment to access the services and the products. A cloud environment can be public or the private cloud. In such environment, all the resources are available on an integrated environment where multiple users can perform the request at same time. In such case, some approach is required to perform the effective scheduling and the resource allocation. This paper provided review of different dynamic resource allocation strategies in cloud environment.

REFERENCES

[1] Anthony T. Velte, Toby J. Velte, Robert Elsenpeter, "Cloud Computing, A Practical approach"

[2] B. Hayes, "Cloud Computing," *Commun. ACM*, vol. 51, no. 7, pp. 9–11, Jul. 2008.

[3] V. Vinothina, Dr. R. Sridaran, Dr. Padmavathi Ganapathi, "A Survey on Resource Allocation Strategies in Cloud Computing", International Journal of Advanced Computer Science and Applications, Vol. 3, No.6, 2012.

[4] Valérie Danielle Justafort, Samuel Pierre, "Performance-Aware Virtual Machine Allocation Approach in an Intercloud Environment", 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), April 29-May 2012, pp 1-4.

[5] Bo Yin, Ying Wang, Luoming Meng, Xuesong Qiu "A Multi-dimensional Resource Allocation Algorithm in Cloud Computing", Journal of Information & Computational Science, 2012.

[6] Amit Nathani, Sanjay Chaudharya, Gaurav Somani, "Policy based resource allocation in IaaS cloud", ACM Future generation Computer Systems, Vol. 28, Issue 1, January 2012, pp 94-103.

[7] Sharrukh Zaman, Daniel Grosu, "An Online Mechanism for Dynamic VM Provisioning and Allocation in Clouds", IEEE Fifth International Conference on Cloud Computing, 24-29 June 2012, pp 253-260.

[8] Mayank Mishra, Anwesha Das, Purushottam Kulkarni, Anirudha Sahoo, "Dynamic Resource Management Using Virtual Machine Migrations", IEEE Communications Magazine, Vol 50, Issue 9, September 2012

[9] Gihun Jung, Kwang Mong Sim ,"Agent-based Adaptive Resource Allocation on the Cloud Computing Environment", 2011 International Conference on Parallel Processing Workshops, pp 345-351.

[10] Xiao-Jun Chen, Jing Zhang, Jun-Huai Li, Xiang Li, "Resource Reconstruction Algorithms for On-demand Allocation in Virtual Computing Resource Pool", Springer, International Journal of Automation and Computing, Vol 9, Issue 2, pp 142-154, DOI: 10.1007/s11633-012-0627-3.

[11] Ilhem Fajjari, Nadjib Aitsaadi, Guy Pujolle, Hubert Zimmermann, "An Optimised Dynamic Resource Allocation Algorithm for Cloud"s Backbone Network", 37th Annual IEEE Conference on Local Computer networks, 22-25 Oct 2012, pp 252-255,

[12] K C Gouda, Radhika T V, Akshatha M, "Priority based resource allocation model for cloud computing", International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 1, January 2013.

[13] N. Krishnaveni, G. Sivakumar, "Survey on Dynamic Resource Allocation Strategy in Cloud Computing Environment", International Journal of Computer Applications Technology and Research Volume 2– Issue 6, 731 - 737, 2013.

[14] Sukhpal Singh and Inderveer Chana, "Energy based Efficient Resource Scheduling: A Step Towards Green Computing", International Journal of Energy, Information and Communications Vol.5, Issue 2 (2014), pp.35-52.

[15] Riddhi Patel, Hitul Patel, Sanjay Patel, "Efficient Resource Allocation in Cloud Computing", International Journal For Technological Research In Engineering Volume 2, Issue 7, March-2015