IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Volume 6, PP 40-42

Review on Resource Allocation Strategies In Cloud Computing

Varsha Salunkhe[1], Komal Gothwal[2] , Ashmita Shetty[3], Dimple Bafna[4]
[1](Asst.Professor, Information Technology , Atharva College, Mumbai University, India)
[2](Asst.Professor, Information Technology , Atharva College, Mumbai University, India)
[3](Asst.Professor, Information Technology , Atharva College, Mumbai University, India)
[4](Asst.Professor, Information Technology , Atharva College, Mumbai University, India)

Abstract: *Cloud Computing has a great impact on IT industries. Central storage is being the need of current generation and this can be achieved using cloud computing. Cloud uses huge number of resources to provide its services. Managing the resources effectively has become a new challenge in today's world. The workload in cloud computing changes as per the demand and thereby needs dynamic resource provisioning and allocation. In cloud computing, resources are allocated effectively with high resource utilization and low cost. This work presents a comprehensive In this paper, we have discussed various optimal resource allocation schemes reviewing the resource allocation strategies.*

Keywords: *Cloud Computing, Cloud Services, Resource Allocation.*

I. Introduction

Cloud computing is an emerging trend used for delivering services via on a pay-as-you-go basis internet in IT industry[1]. "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources e.g., networks, servers, storage, applications, and any other available resources that can be rapidly provisioned with minimal management effort or extra provider interaction."[1] Cloud computing offers service models such as Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) [2]. Cloud allows customer to select resources based on requirements and pay as per the usage. In a cloud computing environment, the clients wish to minimize the cost and maximize the performance.

Cloud computing is a approach which provides a delivery of computing resources over the Internet. Cloud computing is one of the latest technology in the world of Information Technology, which describes the meaning of the cloud is that clouded cover computing is a set of web based computing resources that deliver on demand information services to users from anywhere in the world. The main aim of using cloud is that cloud is a huge group of interconnected computers. These computers can be any kind of personal computers or network servers or they can be public or private. This cloud of computers extends beyond a single company organization. Cloud storage is a model in where data can be maintained, managed and backed up remotely and made available to users over a Internet. The applications and data served by the cloud are available to huge group of users who requests for services on cloud connected through internet. Any authoritative user can use these documents, different services and applications from any computer over any Internet as it is managed in The cloud is web storage where your all data remain store. Any user can access that data from the cloud according to user requirements. The clouds flexibility is most useful for growing business as it provides ease of access, particularly with data sharing. Documents and files can be stored in the cloud and accessed via an Internet connection, which is ideal solution for business. Also in terms of data backup, the cloud provides peace of mind with securing useful data. The main aim of cloud computing is to keep data secure used in supercomputing, or high-performance computing, normally used in military and research areas to perform number of computations per second. The data is also secured in user-oriented applications such as financial applications to provide customized information and to provide data storage on a large scale. Cloud computing has been changing how mostly users use the web and how they store their files of data on web. There are example services like Google Docs and Gmail etc. which allows user to use cloud to securely store the data. Cloud computing technology uses huge group of servers in a network which are low-cost user PC technology with specialized connections to increase data processing tasks over the network. The resource is generally depend on computer mainframe and can also be classified into 3 layers from top to bottom: First, in the application layer– the operation status of network server, database server and other related servers, Second, Network layer – routing between hosts, data transmission bandwidth, communication data layer, Third, System Layer – CPU Information, Memory size. In other hand, cloud computing resources can also comprise of operation system, disk space, network communication capabilities, data resources, equipment and other related resources.

In cloud computing, resource allocation is one of the crucial thing used where available resources are allocated to the ongoing processes or any needed cloud application users, cloud resources can be provisioned on-demand in a fine-grained, multiplexer manner. In the cloud the resource allocation is based on the software,

platform and infrastructure respectively known as software as a Service(SaaS), Platform as a Service(PaaS) and Infrastructure as a service (IaaS). In cloud platforms, resource allocation takes place at two levels as follows: In first level, when cloud users uploads any data over cloud, then the load balancer assigns the required resources and space as per the demand of the cloud user to balance the computational load of multiple applications across available physical computers. In second level of resource allocation, multiple requests for the use of cloud are served by assigning available resources to specific instance of application to balance the computational load. Resource Allocation is a method to allocate the available resources as per the demand in efficient and effective way. This process helps to prevent the over provisioning and under provisioning of available resources. A Resource Allocation Strategy in cloud computing is a technique that aims to fulfill the requirements of cloud users by using physical and/or virtual resources with proper utilization. Resource allocation means allocating optimal resources to the jobs requested by the user so that these jobs are completed efficiently. In terms of cloud environment, it means allocating a virtual machine satisfying the configurations mentioned by the user [3].Another way to do this is by having predefined virtual machines with in-built environment. Users must submit their jobs or workloads which may have their own time constraints. The effective way in which these workloads can be allocated to the virtual machines and processed is another form of resource allocation possible method in the cloud. There must be an effective Service Level Agreement (SLA) between the cloud vendors and the client. This SLA should contain the resource requirements of the client like CPU and memory [3].

II. Resource Allocation Strategies

1. Linear Scheduling Methods: The FIFO or LIFO scheduling methods plays vital role in linear scheduling. Linear Scheduling for Tasks and Resources (LSTR) scheduling algorithm was proposed by Abirami and Ramanathan [4] which applies scheduling for processing in tasks and resources respectively. It combines different services to a server node to form the Infrastructure a service cloud environment. KVM/ Xen Virtualization method is used along with Linear scheduling method to allocate resources. The dynamic allocation could be carried out by the scheduler dynamically on requests for additional resources with uninterrupted assessment of the threshold value. The requests for resources are collected and are sorted in different queues based on a threshold value. Then, the requests are satisfied by the VM's.
2. Virtual Machine (VM): A system which can automatically scale its infrastructure resources is designed in [5]. The system is made up of a network of virtual machines that are able for live migration across multi-domain physical infrastructure. By using dynamic availability of infrastructure resources and dynamic application demand, a virtual computation environment is able to automatically relocate itself across the infrastructure and scale its resources. But the above work considers only the non- preemptable scheduling policy. VM based resource allocation executes the skewness algorithm [6]. VM resources are allocated in IaaS based on the load, type, cost, and speed. It executes if the user adds or deletes one or more occurrence of resources on the basis of the VM load and circumstances. The virtual computation environment rearranges the resources and balances its resources due to the dynamic availability of infrastructure resources and application demand. The resource allocation and administration approaches imposed the VMs to deliver truthful types and efficient virtual resource allocation [7].
3. Nature Inspired Optimization Methods: Biologically inspired methods are based on modeling animals' natural behavior to reach a solution for optimization problems. It includes ant colony, bee colony, and firefly and eagle strategy. A study based on ant colony optimization has been proposed [8]. Authors present a task classification based on QoS with network bandwidth, service completion time, the system reliability and costs as a QoS parameters. In this experiment they set the number of task from 20 to 100, the number of node calculation of 8, In order to show distinction, they designed the QoS attribute of node set up large gap, mainly including the CPU, memory and network bandwidth. Application of ant colony optimization and random distribution algorithm respectively carry out 10 times they realized with the increase of the quantity task, through the ant colony optimization algorithm performs all the tasks, it takes the time less than general algorithm
4. Auction based allocation: Auction based allocation uses sealed-bid. Sealed-bid is a process where service provider gathers users' bids and fixes the price. The truth telling property minimizes the resource allocation problem into an ordering problem and reduces the difficulty of cloud service provider decision rule and clear-cut allocation rule. The constraints of the truth-telling property do not ensure profit maximization. In a cloud computing environment [9], the dynamic auction strategy solves the resource allocation problem. A second-price auction mechanism determines the price of resource allocation and capacity of the computation allocation using truth-telling mechanism. The cloud provider uses an auction-based market to achieve maximum profit over time for each VM [10]. The demands of each type of VMs vary over a period, and hence it becomes important to make the capacity of each VM type adaptive to increase the profit while reducing the energy cost. Model Predictive Control achieves a balance between customer demand and customer satisfaction. Several auction-based resource allocations solve the resource allocation as a dynamic

capacity problem without considering its profit. The work in resource allocation in spot markets in cloud computing[10] maximizes the returns using Model Predictive Control and reduces the trade-offs between the customer's demand and satisfaction. The auction based resource allocation issues on provider's revenue, energy minimization, customer satisfaction, and response time degrades the overall performance of the resource allocation.

5. QoS based resource allocation: QoS based resource allocation focuses on customer satisfaction, cost-efficient and effective utilization of the resource. Resource allocation in cloud computing varies from the traditional distributed computing environment due to the presence of various QoS metrics such as CPU memory, speed, and stability. Resource allocation technique mainly considers the QoS parameters on resource providers, including price and load [11]. The resource allocation strategy focuses SLA driven consumer based QoS metrics to improve the profit of SaaS providers [12]. The profit model evaluates the performance of web-facing application and considers the consequences of queries returned in a short-response time as well as queries returned in long response time. It reveals the significance of the response time in computing the trade-offs between QoS and resource utilization in IaaS clouds.
6. Service-level Agreement: A service-level agreement (SLA) is a agreement between a service provider and its customer describing what services the provider will furnish. SLAs started with network service provider, but are now used by telecommunication service providers and cloud computing service providers on a large scale [12]. This resource allocation algorithm ensure that Service providers are able to manage the dynamic change of users, handling user requests to infrastructure level parameters and handling heterogeneity of Virtual Machines minimizing infrastructure cost and SLA violation. CloudSim [13] is used to simulate the cloud computing environment that utilizes the proposed algorithms for resource allocation. Performance is measured from both customers and SaaS providers' point of view. Considering customers' perspective, how many SLAs are violated has been observed and considering SaaS providers' perspective, how much cost reduced and how many VMs are initiated has been observed.

III. Conclusion

Resource allocation is the process of allocating resources based on customer requirements over the internet using cloud. This review shows different types of Research Allocation Strategies , their advantages and their role in the cloud. Selection of suitable Research Allocation Strategy will result in efficient throughput, resource utilization, less response time and latency of resources in the cloud. Research Allocation Strategy is used to reduce response time, to increase the performance and avoid over provisioning and under provisioning of resources in cloud.

References

[1] Rajkumar Buyya et.al.,2009, Cloud Computing and Emerging IT Platforms Vision, Hype, and Reality for Delivering IT Services as the Utility, *Future Generation Computer Systems, vol. 25, no. 6, pp. 599-616.*

[2] C. N. Hofer,G. Karagiannis.,2011, Cloud Computing Services:Taxonomy and Comparison, *Journal of Internet Services and Applications, vol. 2, Issue 2, pp. 81-94.*

[3] S.ThamaraiSelvi, C. Valliyammai and V. Neelaya Dhatchayani,Resource allocation issues and challenges in cloud computing,, *International Conference on Recent Trends in Information Technology, 2014 IEEE*

[4] Abirami S.P. and Shalini Ramanathan, Linear Scheduling Strategy for Resource Allocation in Cloud Environment, *International Journal on Cloud Computing: Services and Architecture, February 2012.*

[5] P.Ruth,J.Rhee, D.Xu, R.Kennell and S.Goasguen, Autonomic Adaptation of virtual computational environments in a multi-domain infrastructure, IEEE *International conference on Autonomic Computing, 2006,pp.5-14.*

[6] *Zi Chen., 2013*, Dynamic Resource Allocation using Virtual Machines for Cloud Computing Environment, *IEEhen Xiao, Weijia Song, and QE Transactions on Parallel and Distributed systems, vol. 24, pp. 1107-1117.*

[7] *Brendan Jennings and Rolf Stadler., 2014, Resource Management in Clouds: Survey and Research Challenges, Journal of Network and System Management.*

[8] Linan Zhu, Qingshui Li, and Lingna He, Study on Cloud Computing Resource Scheduling Strategy Based on the Ant Colony optimization Algorithm, *International Journal of Computer Science, September 2012.*

[9] Wei-Yu Lin, Guan-Yu Lin, Hung-Yu Wei., 2010, Dynamic Auction Mechanism for Cloud Resource Allocation, *10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, pp. 591-592.*

[10] Qi Zhang, Quanyan Zhu, Raouf Boutaba., 2011, Dynamic Resource Allocation for Spot Markets in Cloud Computing Environment, *IEEE 4 th International Conference on Utility and Cloud Computing, pp.178-185.*

[11] I. Popovici et al.,Profitable Services in an Uncertain-World, *ACM/IEEE Conference on Supercomputing, vol. 36,2005.*

[12] Linlin Wu,Saurabh Kumar Garg and Raj kumarBuyya., 2011, SLA based Resource Allocation for SaaS Provides in Cloud Computing Environments, *IEEE, pp.195-204.*

[13] Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov1 and Cesar A. F. De Rose, CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms" *24 August 2010 in Wiley Online Library.*