INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN: 2278 – 0211 (Online)

Priority Driven Auction Strategy For Resource Allocation In Cloud Datacenters

Ronak Patel
PG Student of Computer Engineering, LDRP-ITR, Gujarat Technological University, Gujarat, India

Abstract:

Cloud computing is the next generation of technology which unifies everything into one. It 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 services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. A new model known as Priority Driven Auction Strategy (PDAS) for Resource Allocation is proposed in this research paper. Resource allocation takes place based on the priority of resources and auction bids between cloud user and datacenters. Load-balancing policy is used for assigning priorities to the resources. Proposed model is implemented in simulation environment and several experimental tests are performed to analyse its performance.

Key words: *Resource Auction, Priority Driven Resource Allocation, Dynamic Resource Allocation, Cloud Computing, Resource Management, Resource Scheduling, Cloud Datacenter.*

1.Introduction

Cloud computing is the next generation in computation. It attempts to provide affordable and easy access to computational resources over the internet. Possibly people can have everything they need on the cloud. Cloud computing is the next natural step in the evolution of on-demand information technology services and products. 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. Cloud computing proves to be a practical approach to experience direct cost benefits and it has the potential to transform a data centre from a capital-intensive set up to a variable priced environment.

From the cloud user's point of view, cloud computing provides an abstraction of the underlying hardware architecture. This abstraction saves them the costs of design, setup and maintenance of a datacenter to host their Application Environments. From cloud provider's point of view, cloud computing consists of several datacenters having shared pool of resources, applications, data etc. Cloud provider manages these resources dynamically such that they are efficiently shared and utilized amongst all the cloud users. In order to fulfil the efficient resource utilization, virtualization is used along with some resource allocation technique.

Cloud computing technology is rapidly advancing and many IT giants like Microsoft, Amazon, Google, IBM, and Hewlett-Packard provide a platform to developers for deploying their applications across computers hosted by a central organization in cloud infrastructure. These applications can access a large network of computing resources that are deployed and managed by a cloud computing provider. Cloud computing being service-oriented and not application-oriented, it provides generic services to its users. Many organizations are benefited due to cloud computing as it can serve different types of services to different types of users. Even non IT organizations can make use of cloud services to boost their business and reduce the infrastructure, setup and managing costs of in-house datacenter. The main advantage with cloud computing is that the business enterprises can scale up to required capacities instantaneously without cloud provider intervention and they ought to pay only for what they use.

The cloud computing technology makes the resource as a single point of access to the client and is implemented as pay per usage. The order and way in which request for resources are satisfied refers to resource scheduling. The allocation of resources must be made efficiently such that system utilization and overall performance is maximized. The

Quality of Service (QoS) and Service Level Agreement (SLA) aspects are to be taken care of while an attempt for maximizing resource utilization is made. QoS or SLA should not be compromised to achieve high resource utilization in cloud datacenters.

1.1.Resource Allocation

In cloud computing environment, Resource Allocation (RA) is the process of assigning available resources to the required cloud user applications over the internet. Pool of physical resources is shared amongst several applications of cloud users. Virtualization enables the sharing of resources in such a manner that physical level intricacies are abstracted from the user and flexibility is provided.

Services starve if resource allocation is not properly managed. Resources being the core of cloud server, it becomes crucial to manage these resources in a truly efficient manner so that cloud functions smoothly. Therefore, resource allocation is a challenge to be addressed. Resource allocation strategy refers to the technique or methodology used for selecting and assigning the resources to requesting client applications. Efficient resource allocation in the cloud is a challenging task as it needs to satisfy both, the user's requirements and server's performance equally. To address this challenge a new resource allocation scheme is proposed in this paper.

1.2.Cloud Server Hierarchy

The hierarchy of the cloud server is as shown in Figure 1. At the core level is basic cloud infrastructure. A cloud infrastructure may consist of one or more datacenters. Each datacenter may have one or more Hosts, and each host may have one or more Virtual Machines. On each virtual machine multiple cloudlets are executed.

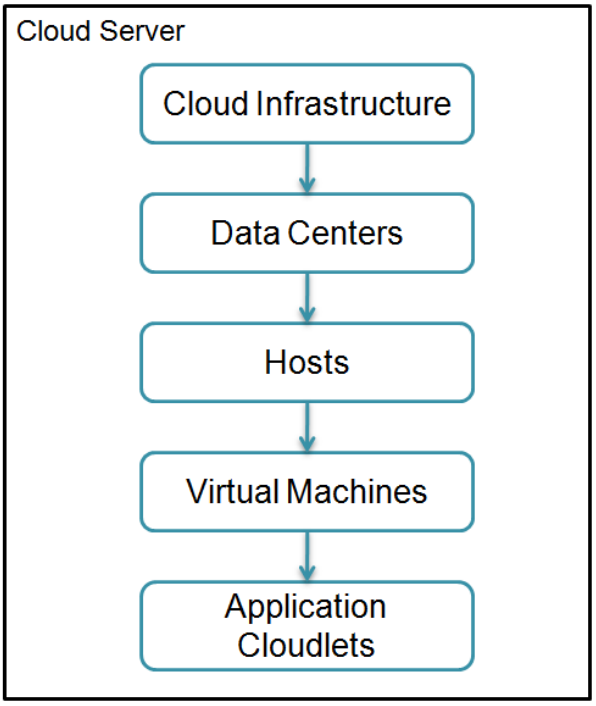

Figure 1: Hierarchy of Cloud Server

1.3.Proposed Resource Allocation Model

Technique for resource allocation in proposed model incorporates two aspects for selection and assigning of resources to cloud users. The two aspects are: priority of resource and auctioning scheme. Resource allocation takes place based on the priority of resources and auction bids between cloud user and datacenters. First, the resources are prioritized and candidate hosts are selected based on priority of the resource and requirements of cloud user. Then, auction is performed between datacenters of selected hosts and cloud client. By using the priority of resources only appropriate hosts will be taken up for auction and so efficiency of the resource allocation scheme increases.

2.Existing Techniques and Related Work

Dynamic resource allocation is one of the most challenging problems in the resource management of cloud datacenters. Many researchers around the world have designed new methodologies and ways for addressing the issue of dynamic resource allocation.

Several techniques and methodologies have slowly evolved for resource allocation in cloud computing environment. The broad and main stream techniques are described below in brief. Survey paper [12] shows a contemporary comparison between resource allocation schemes used in popular cloud architectures like OpenNebula and Eucalyptus.

2.1.SLA-Oriented Resource Allocation

In SLA Oriented resource allocation method, service level agreement (SLA) between the user and service provider is considered while doing allocation of resources. SLA based resource allocation for SaaS environment is depicted in [18].

2.2.Priority Driven Resource Allocation

Priority driven resource allocation for cloud computing can be classified mainly in two types: 1) User priority based and 2) resource priority based.

In User Priority approach to resource allocation, each cloud user is assigned a priority based on some parameters and resources from cloud server will be allocated to the user based on its priority.

In Resource Priority based approach, unlike User Priority approach, Resource Priority Scheme assigns priorities to resources of cloud server instead of users. Based on the priority of resources the resource allocation will be performed to the requesting clients.

Priority based resource allocation models are described in [6] and [13].

2.3.Power Efficient Resource Allocation

In Power efficient resource allocation technique resource allocation will be optimized in such a way so that power consumption by the resources will be minimized.

2.4.Auction Based Resource Allocation

An auctioning scheme is used for allocating resources on the cloud server. Datacenter brokers will accept the requests from clients and performs auction with datacenters connected to it. Best candidate based on its cost will be selected as winner of the auction and application cloudlets will execute on winner hosts or VMs.

Several type of auction methods are described in [1]. Dynamic Auction mechanism for resource allocation is proposed in [9].

2.5.QoS Governed Resource Allocation

Quality of Service to be provided will decide the way in which resources will be allocated. QoS parameters will be decided between user and cloud server before the communication begins. Clients having higher quality requirements will be provisioned with higher number of resources or resources with high efficiency and reliability.

Multidimensional QoS resource scheduling algorithm is presented in [4].

2.6.Application/Network Aware Resource Allocation

In this scheme of resource allocation, the resource allocation algorithm takes in consideration the type of application and type of network structure from which the client is requesting the resources. Based on the types of application requirements the resources will be allocated to it accordingly. Multidimensional resource allocation approach is proposed in [20].

2.7.Genetic Algorithms

Genetic Algorithms derives its name from Biology. Resource Allocation in this scheme will use base of genetic algorithms to find the optimum resource allocation and assign resources to the clients based on outcome of genetic algorithm. Resource Allocation for jobs using neural network is shown in [3].

2.8.Load Balancing Resource Allocation

In load balancing method, resource allocation is done considering the system load. Load calculation is mostly dynamic and so resource allocation also becomes dynamic. No datacenter(s) is/are overloaded in this scheme because workload gets distributed among all the resources.

An efficient VM Load balancing algorithm can be found in [7]. Optimal Resource allocation Technique is proposed in [5]. Dynamic Resource Allocation using Virtual Machines is shown in [19].

I. PRIORITY DRIVEN AUCTION STRATEGY

In priority driven auction strategy, resources of the cloud server are first prioritized into order. Then, selection of hosts based on priority as well as client requirements is made and finally auction is performed between cloud user and selected hosts. Tasks of the cloud users are executed on the datacenter which wins the auction.

As priority is assigned to resources in datacenters of cloud it helps in utilizing resources based on some policy. This priority assigning policy is independent from resource allocation so the policy can be easily changed any time without much complexity. Auction strategy chooses the best match between requirements of user and offerings made by the service provider. It also calculates and negotiates costs dynamically. Therefore combination of priority driven and auction strategy would turn out to be more efficient and powerful then other strategies as cloud users and cloud providers both are benefited through this strategy.

2.9.Entities in Proposed Model

The proposed resource allocation model is based on dynamic priority assignment to resources in cloud datacenters and then performing auction between priority-wise selected hosts and cloud user. Resource allocation algorithm considers requirements of client and priority of resources for assigning the requested resources to the clients. Proposed system consists of three main entities:

- Request Client (Cloud User)
- Resource Manager
- Auctioneer

2.10.Architecture of Priority Driven Auction Strategy

The architecture of Priority Driven Auction Strategy (PDAS) is as shown in Figure [2].

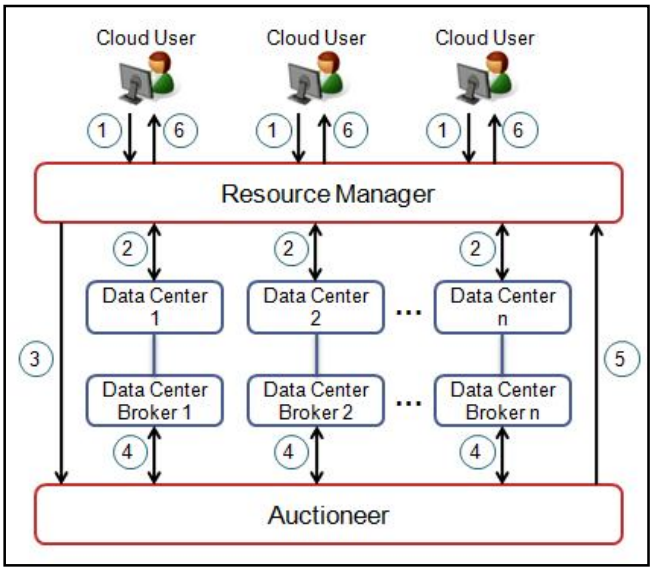

Figure 2: Architecture of Priority Driven Auction Strategy (PDAS)

2.11.Request Client (Cloud User)

Request Client will reside on client side and it will handle the requests made by cloud clients. Primary goals of request client would be 1) to evaluate the request parameters against SLA, 2) Authenticate users 3) Prepare Job Description or Requirements and 4) forward the requests to the cloud server with the provided credentials.

Request client will first authenticate the cloud user. Then, it will check whether the requests made by the user adhere to decided SLA. If SLA is not violated, then the Request client would prepare a job description or requirement list and forward the

request to cloud server through the network. It may also send some parameters which will describe the type of request, service required, type of application, criticality etc.

2.12.Resource Manager

Resource Manager is the middleware entity which will act as a mediator between the actual physical resources and incoming requests from multiple Request Clients. Resource Manager is implemented inside the cloud server. The type of resource manager to use depends on which cloud computing system is used and how it is implemented. The main tasks of resource manager are to manage, maintain and allocate all the resources available in cloud server.

Resource Manager monitors the state of resources and provides statistical data to Auctioneer. Resource allocation/de-allocation is also carried out by the Resource Manager. When the task gets completed, results are sent back to Request Client and resources are released which can be reused for new requests.

2.13.Auctioneer

The Auctioneer is the entity which will govern the entire auction scheme. It gets the list of selected hosts from Resource Manager and performs auction between the selected hosts and Request Client. The datacenter which wins the auction will be chosen and hosts of that datacenter will be assigned to cloud user for execution of tasks or cloudlets. Auctioneer will pass on the results to Resource Manager.

2.14.Control Flow

The control flow and steps in working of Priority Driven Auction Strategy is as explained below. Step numbers correlate to the architecture of PDAS shown in Figure [2].

- Cloud Users will make request for resources and submit tasks to cloud server. Resource Manager will accept the requests of multiple Cloud Users.
- Resource Manager is connected to all the datacenters in cloud so it will evaluate the requirements of Cloud User with all available datacenters and select those datacenters (and Hosts) whose priority is high and which meets the requirements of Cloud User.
- Resource Manager will forward the selected datacenters and Hosts to the Auctioneer.

- Auctioneer is connected to all the datacenter Brokers which act on behalf of respective datacenters. Auctioneer will perform the auction between selected datacenters and the Cloud User. Cloudlets or user tasks are executed in Hosts of datacenter which wins the auction.
- Results of the executed tasks are transferred to Resource Manager.
- Resource Manager will send results back to the Cloud Users.

3.Algorithms And Flowchart

Resource allocation algorithm plays a crucial role in incorporating efficient and effective use of resources. The proposed algorithm is dynamic, which means it will make decisions of resource allocation at runtime based on the statistics of state of resources, load on system, number of clients, type of resource request etc. Resource allocation algorithm will obtain state of resources from the Resource Manager and type of resource request and such parameters from the Request Client.

After fetching the required parameters and statistics, resource allocation algorithm would analyse the client requirements. Then, it will try to compute a resource assignment which will not only fulfil all the requirements but also allocate resources in an efficient manner by using priority of the resources. Auctioneer will perform auction between selected Hosts and resources will be allocated to respective clients through Resource Manager. Consequently, Resource Manager would update its statistical data, to reflect the changes, which would be used for succeeding client requests.

3.1.Select Priority Hosts Algorithm

This algorithm selects the Hosts based on priority which meets the requirements of clients.

3.1.1.Select PriorityHosts

- Inputs: Client Requirements, DataCentersList, HostsList
- Output: Selected Hosts based on Priority

- For each DataCenter in DataCentersList
- For each Host in HostsList of DataCenter
- If (Host meets requirements) then

- Candidates.Add(Host)
- End If
- End For
- End If
- Candidates = SortByPriority(Candidates)
- SelectedHosts = SelectTop(Candidates, 10)
- Return Selected Hosts

3.2.Sort By Priority Algorithm

This algorithm assigns priority to hosts and then sorts them in descending order so that higher priority tasks are at top. There are generally three policies for assigning priorities:

- Packing Policy - Hosts with more number of VMs running are selected first.
- Stripping Policy - Hosts with less number of VMs running are selected first.
- Load-Aware Policy - Hosts with more free CPU are selected first
- In this algorithm Load-Aware policy is used.

3.2.1.SortByPriority

- Input: CandidateHosts
- Output: SortedHosts

- For each Host in CandidateHosts
- Priority[Host] = Host.FreeCPU()
- End For
- Sort Candidates on Priority in Descending
- Return SortedHosts

3.3.Perform Auction Algorithm

This algorithm performs auction between the selected Hosts and cloud user.

3.3.1.Perform Auction

- Inputs: SelectedHosts, Tasks
- Output: Results of Task execution

- For each VM in SelectedHosts
- Perform Auction
- End For
- Execute Cloudlets in VM of Winner Host
- Return Results

3.4.Priority Driven Auction Flowchart

The flowchart of priority driven auction strategy for resource allocation is shown in Figure 3.

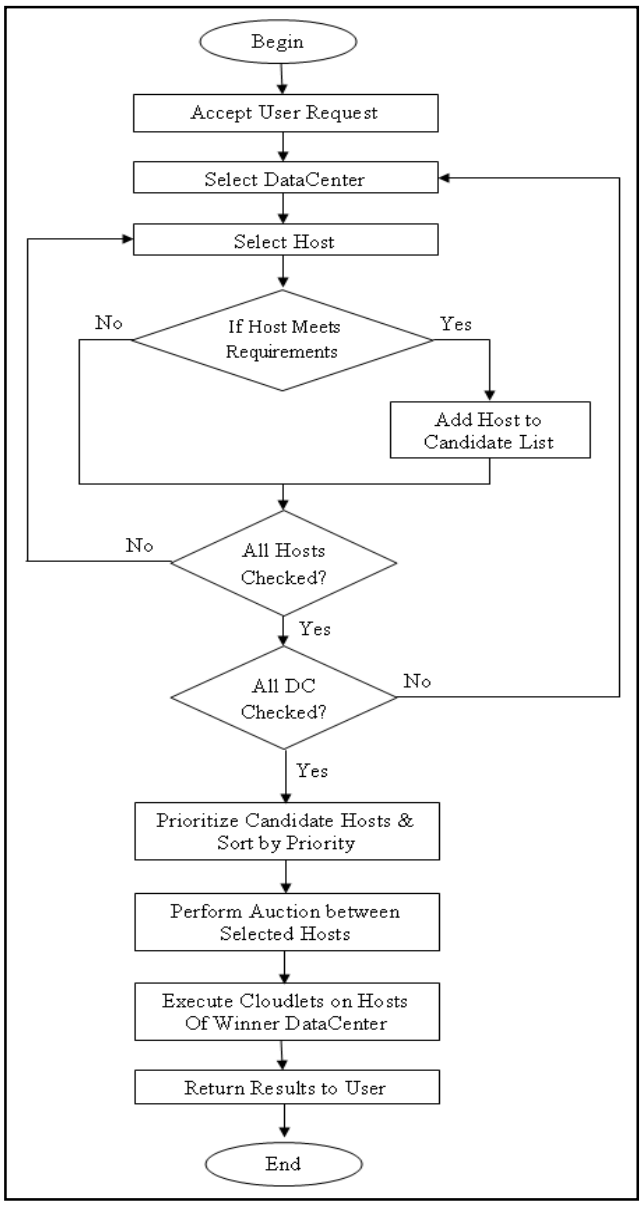

Figure 3: Priority Driven Auction Strategy Flowchart

4.Result Analysis

For the purpose of evaluation and result analysis CloudSim is used as a simulation framework on which proposed research model of Priority Driven Auction Strategy has been implemented. A virtual machine or a host can be considered as a resource unit and performance analysis is generally carried out considering these resource units.

The setup parameters used for evaluation and analysis of the proposed model are represented in the Table I.

Parameters	Value
Datacenters	4
Total Hosts	8
Number of Processing Units	16
Total RAM	38 GB
Storage Capacity	32 TB
Host OS	Linux
Host Architecture	x86
Hypervisor	Xen
Cloud Users	1
Application Cloudlets	2 to 10

Table 1: Experimental Setup Parameters

Proposed model of Priority Driven Auction Strategy is mainly compared with Simple Auction Strategy and normal execution strategy. Performance check on proposed research model was carried out considering parameters like execution time, number of cloudlets, size of cloudlets, datacenters, cost, average resource utilization, response time etc.

4.1.Time Vs. Cloudlets Comparison

Priority Driven Auction Strategy (PDAS) is compared with normal execution and simple auction strategy based on Time Vs. Number of Cloudlets in Figure 4.

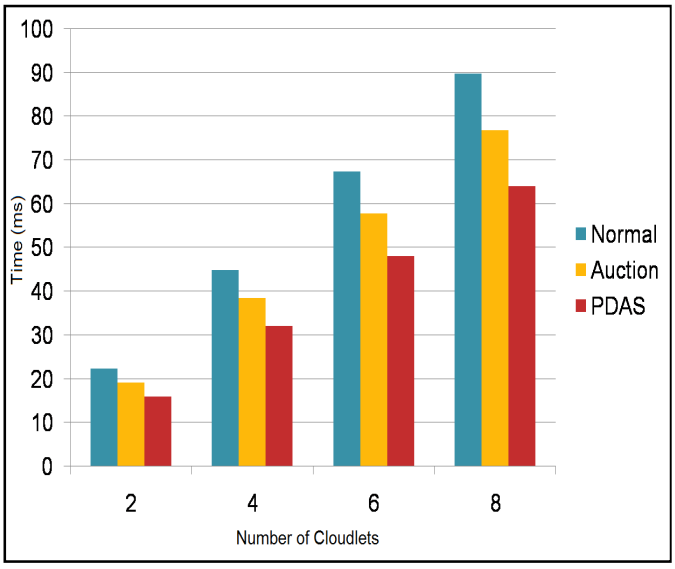

Figure 4: Time Vs. Cloudlets Comparison

4.2.Resource Utilization

Three parameters: CPU, RAM and Bandwidth are considered for evaluation of resource utilization of proposed Priority Driven Auction Strategy model in Figure 5.

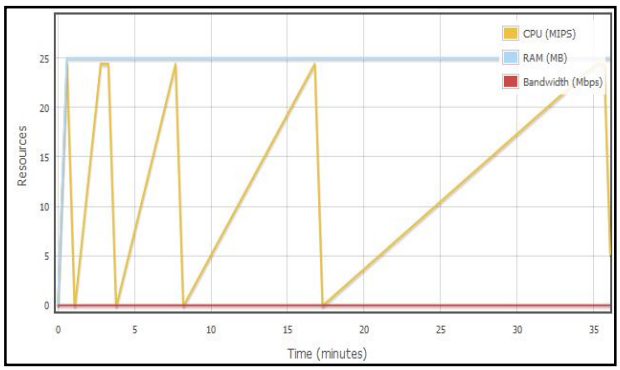

Figure 5: Resource Utilization

4.3.Execution Time

Four parameters: Start time, Finish time, Average start and Average finish times of cloudlet execution are considered for evaluation of execution time in Figure 6.

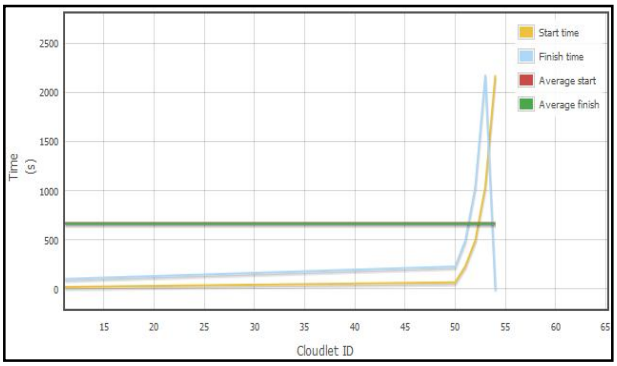

Figure 6: Execution Time

5.Conclusion

Priority Driven Auction Strategy for resource allocation has been proposed in this research paper. Resources being the core of datacenters, managing resources and allocating them in an efficient manner through priority driven auction strategy makes the cloud systems more efficient. In order to effectively meet the varying requirements of cloud users, priority driven auction strategy is designed to be dynamic and scalable. Dynamic priority assignment in PDAS to resources facilitates load-balancing of datacenters. Moreover, through strategy of auction best match is found and cloud users as well as cloud providers, both are benefited. Thus, Priority Driven Auction Strategy for resource allocation is a potential candidate for adaptation in enterprise level cloud softwares. It is intended that this research paper will be helpful to researchers and cloud computing community.

6.Reference

1. Ghalem Belalem, Samah Bouamama, and Larbi Sekhri. An effective economic management of resources in cloud computing. Journal of Computers, 6(3):404-411, 2011.
2. Rajkumar Buyya, James Brobberg, and Andrzej Goscinski. Cloud Computing Principles and Paradigms. John Wiley & Sons, Inc., Publication, New Jersey, 2011.
3. K. Dinesh, G. Poornima, and K. Kiruthika. Efficient resources allocation for different jobs in cloud. International Journal of Computer Applications, 56(10):30-35, 2012.
4. Wuqi Gao and Fengju Kang. Cloud simulation resource scheduling algorithm based on multi-dimensional quality of service. Information Technology Journal 11, 1:94-101, 2012.
5. K. L. Giridas and Shajin Nargunam. Optimal resource allocation technique (orat) for green cloud computing. International Journal of Computer Applications, 55(5):20-26, 2012.
6. K C Gouda, Radhika T V, and Akshatha M. Priority based resource allocation model for cloud computing. International Journal of Science, Engineering and Technology Research (IJSETR), 2(1):215-219, 2013.
7. Jasmin James and Dr. Bhupendra Verma. Efficient vm load balacing algorithm for a cloud computing environment. International Journal on Computer Science and Engineering (IJCSE), 4(9):1658-1663, 2012.
8. Gunho Lee, Niraj Tolia, Parthasarathy Ranganathan, and Randy H. Katz. Topology aware resorce allocation for data-intensive workloads. ACM SIGCOMM Computer Communication Review, 41(1):120-124, 2011.
9. Wei-Yu Lin, Guan-Yu Lin, and Hung-Yu Wei. Dynamic auction mechanism for cloud resource allocation. 2011.
10. P. Mell and T. Grance. The NIST definition of cloud computing (draft). National Institute of Standards and Technology, 53:7, 2010.
11. Thangaraj P, Soundarrajan S, and Mythili A. Resource allocation policy for infrastructure as a service in cloud computing. International Journal of Computer Science and Management Research, 2(2):1645-1649, 2013.

12. Piyush Patel and Arun Kr. Singh. A survey on resource allocation algorithms in cloud computing environment. Golden Research Thoughts, 2(4):1-9, 2012.
13. Chandrashekhar S. Pawar and Rajnikant B. Wagh. Priority based dynamic resource allocation in cloud computing. 2012.
14. Abirami S.P. and Shalini Ramanathan. Linear scheduling strategy for resource allocation in cloud environment. International Journal on Cloud Computing: Services and Architecture (IJCCSA), 2(1):9-17, 2012.
15. Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/cloud _computing
16. V. Vinothina, Dr. R. Shridaran, and Dr. Padmavathi Ganpathi. A survey on resource allocation strategies in cloud computing. International Journal of Advanced Computer Science and Applications, 3(6):97-104, 2012.
17. Daniel Warneke and Odej Kao. Exploiting dynamic resource allocation for efficient parallel data processing in the cloud. IEEE Transactions On Parallel And Distributed Systems, 2011.
18. Linlin Wu, Saurabh Kumar Garg, and Rajkumar Buyya. SLA-based resource allocation for software as a service provider (SaaS) in cloud computing environments. 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, pages 195-204, 2011.
19. Zhen Xiao, Weijia Song, and Qi Chen. Dynamic resource allocation using virtual machines for cloud computing environment. IEEE Transaction On Parallel And Distributed Systems (TPDS), pages 1-11, 2010.
20. Bo Yin, Ying Wang, Luoming Meng, and Xuesong Qiu. A multi-dimensional resource allocation algorithm in cloud computing. Journal of Information and Computational Science, 9(11):3021-3028, 2012.