

PalArch's Journal of Archaeology
of Egypt / Egyptology

An Extensive Survey On Resource Allocation Mechanisms In Cloud Computing

Jansi Sophia Mary C¹

¹Associate Professor,

Department of Computer Science and Engineering
Idhaya Engineering College for Women

Dr. Mahalakshmi K²

²Professor,

Department of Computer Science and Engineering
Kalaigarkarananidhi Institute of Technolo

Jansi Sophia Mary C¹, Dr. Mahalakshmi K², An Extensive Survey On Resource Allocation Mechanisms In Cloud Computing - PalArch's Journal of Archaeology of Egypt/Egyptology 17(9). ISSN 1567-214x, Keywords: Cloud Computing, Resource Allocation, Consumer Context, Virtualization, Infrastructure as a Service Cloud, Optimization

Abstract-

At present scenario, Cloud Computing is the emerging trend technology in order to store, process, visualize and analyze data in an efficient way. It deals with on demand dynamic resource allocation for giving assured services to the consumers. Pay-As-You-Use model is used here to the public. Resource Allocation is carried out beside with intention of reducing the price associated with it. In recent days, the major functions of Resource Allocation are meeting the needs of customers and application requirements. Different resource allocation techniques had been inspected in this paper.

1. Introduction:

In an internet, Cloud Computing (CC) is explicated as New- Fashioned technique of assessing, storing the Flexible and Virtual assets dynamically. CC is an empowering framework on demand to get entry for distributed group of sources such as servers, storage units which yields the responsibilities that may additionally be saved and discarded with tons much less control work [1]. CC suggests cutting edge fashion in Information generation that shifts the computing facts into the massive statistics facilities from the desktop. It is elucidated as app delivered as services throughout the web. The computing effort through all cloud environments is furnished by the way of gaggle of knowledge centers, in diverse location are interrelated through excessive velocity network. Cloud may be bunch of distributed systems in cloud computing. It supplies required statically asset to digital clients at some stage in the internet. The method of controlling the useful resource aids to synchronize data assets in retort to control works finished via each the Consumer and givers of cloud. It is called as Resource Allocation (RA) to resource's customers from the resource's providers. User can use the reachable ability greater effectively due to the fact the aid administration lets in to reallocating the assets dynamically. In CC, Resource Allocation is the approach of allotting obtainable resources to the required cloud apps throughout the web. Because of predefined aid allocation process, the Infrastructure as a Service (IaaS) assigns assets to contesting the demands. When the assigning is now not managed accurately, RA craves utilities. The issue is deduced by way

of permitting the utility vendor to regulate the assets for each and every single frame. RA is a section of aid administration and used to allocate the obtainable assets in cost effective manner.

1.1 Cloud Architecture

The cloud computing structure can be consisting of 3 distinct layers. a) Resource b) Platform and c) Application. The first layer i.e., Resource layer (infrastructure layer) which similarly comprised of two aspects i) Physical and Virtualized computing ii) Storage and Networking Resources [5] For e.g., in storage, the purchasers solely can pay only for the section past neither purchasing any disk nor even understanding nothing with respect to the area of record they interrelate with. Occasionally, IaaS is acknowledged as Hardware as a service (HaaS)[6]. Likewise, Platform layer also additionally elucidated as Platform as a Service (PaaS) usually extract the infrastructures and hold a crew of software application confederate to the programming clouds. It is the midway platform allying the physical devices and the software. Google App Engine and Microsoft’s Azure Services Platforms are the frequent examples of PaaS.

The Application layer (also recognized as Software as a Service) restores the functions which are processing on the system. Don’t want to deploy and run the vast software program on your PC if you are the use of SaaS. Pursuing Pay Per Use sample can minimize the whole price instead than shopping for the excessive fee software program[5].

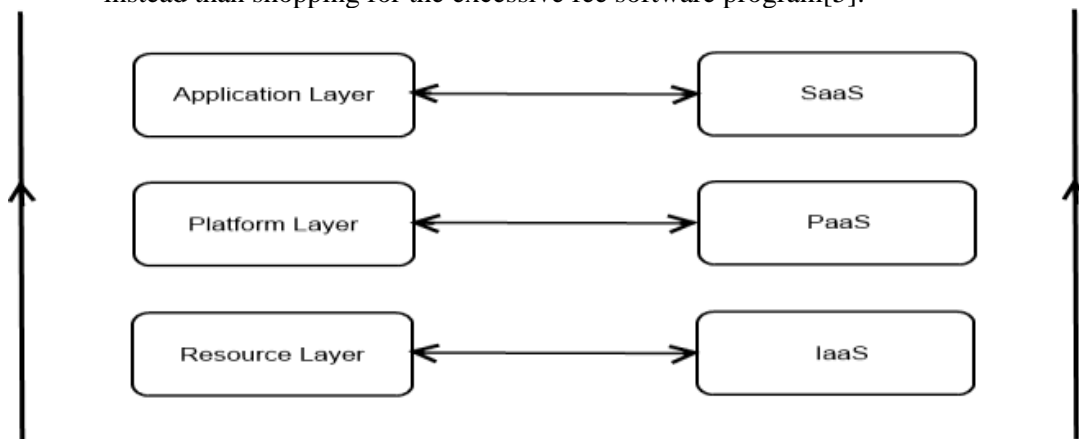


Fig 1: Architecture for cloud computing

The cloud mannequin includes 4 deployment models.

- 1) Private Cloud: The cloud infrastructure is carried out infrequently for organizations. It would possibly be managed either by business or by third party. It can exist either on premise or off premise.
- 2) Community Cloud: In this cloud, the infrastructure is shared amongst the exclusive corporations and helps a unique crew that had shared concerns. For e.g., mission, safety requirements, policy, and compliance concerns etc. It might be managed either by agencies or by a third party. It may also exist either on premises or off premise.
- 3) Public Cloud: owning the cloud with the aid of a company promoting cloud services. It is on hand to all customary public or to crew of massive industry.
- 4)Hybrid Cloud: Hybrid cloud is the charter of each public and private cloud. It may stay distinctive entities however are tightened collectively via regulated techniques which approves records and software transportability.[4]

1.2 Virtualization

For CC, Virtualization methodology gives the technical factors. Normally, virtualization act towards the virtual resource creation like operating systems, storage devices or servers.

Several types of virtualization: System Virtualization boosts up a physical device extraction on above the gadgets, which is recognized as Hypervisor or Virtual Machine Monitor. In a non-privileged environment, Virtual Machines don't have any direct get admission to the devices and the Virtual Machines are processed through the hypervisor.

Numerous Virtual Machines which may deploy on different OS, can be deployed on individual hardware using the System Virtualization. System Virtualization is the most essential methodology which is used to furnish IaaS, PaaS and SaaS resources. Virtual Machine Monitor (VMM) is also known as Hypervisor, is actually software which safely divides the system resources into one or more virtual machines. Instead of running directly on hardware, a guest operating system process under the control of virtual machine monitor. The Virtual Machine Monitor runs in the kernel mode(1), when a guest operating system deploys in the user mode(0). Various Hypervisors supports several criteria of the CC. Hypervisors are in extraordinary types:

- 1) Traditional Hypervisors which perform on devices environment directly are the very likely used to yield prominent results for the single vendors
- 2) The second type is that embedded hypervisor which are combined into processor on a specific chip. The service providers obtain best results by using this type of hypervisor
- 3) The third type is thy hosted hypervisor which runs as a special software layer above the both OS and hardware. These hypervisors are useful in both private and public cloud to get the better outputs.

2. Related Works

To give consultation related allocation, the recent reservation algorithms in Haizea had been enlarged by Janki Akhani,[8]. To expand the odds of assent, the proposed calculation draw out the counter which provides the considering limitations. According to their needs, Consumers may get proper arranged offers using ranking algorithm. These innovations will not grasp the situations while model has many needs for an individual user. An approach, which in consideration to the pre-emption is presented by Amit Nathani [9]. It has two significant ideas. They are: i) Swapping ii)Backfilling. It needs the rearranging of lesser leases when match up to the enduring Haizea technique. When scheduling a timeframe sensitive lease, the proposed algorithm discovers many slots in inclusion to noticing a one slot. An evolved algorithm can expand the reaction season of best work lease. The Backfilling type model is simply expanded only, it is neither tried nor actualized and this can be advised as future enhancement. N.R.RamMohan,E.Baburaj,[10] calculates different web assets assignment policies and their apps in Climate. Depends on distinctively adapted changing quantity, detail explanation for network assets designation in CC had been completed.

In this article, the writers find and categorizes the disputes to the asset's allotment cycle of CC normally as far as different sorts of assets designation methods. A global and experimental execution prize review of restoring and allocating strategies in IaaS clouds exposed by David Villegas[11] Initially, taxonomy for both type of policies is introduced based on kind of data used in destiny process. Next, with the improved design, they examine the work execution and prize of these approaches. A cost-effective leasing algorithm and combined with the Haizea elaborated by SkyMark, Hemant Kumar, Mehata, Eshan Gupta [12] This cost savvy algorithm being answerable for both the client and the supplier. It utilizes the streaming technique concepts to advances the prizes. The motive for consumers is lesser prize of performance of its rent on the fit hub when compared to already existing non economy oriented algorithms. The algorithms enhance the quantity of acknowledged rent, if proper resource is not found by using contraction on cost and assets request. The proposed algorithm arranges the lease depending on its total cost and amount assets together along with the capabilities.

Ahmed Shawish and Maria Salama[6] offered outlook on the cloud's structure, definition, credits, demerits, framework and driven innovation. To offer a full view of cloud service

vendors, they categorize the cloud's deployment and service models clearly. They also noticed the consumer related criteria for e.g., service cost, Service Level Agreement and security issues. In conclusion, it covers explained correlations betwixt the Cloud Computing framework and other exiting works inconsideration of critical difficulties.

A versatile assets allotment structure that allocated the customer's duty to the correct the data centres built by Gihun Jung[13]. The approach to adaptively notice a proper data centre depends upon two developments. i) the geographical partition allying a customer and (ii) the outstanding task at hand of each data centre. The proposed technique can be enhanced to not only for CPU-bond job with non-dynamic workload, but also various of memory region, dynamic workloads and live relocation position. Moreover, the proposed technique can be developed in real VMM platforms like Xen Hypervisor.

A new resource planning technique by discovering faith-oriented fidelity mechanism by Yanbing Liu[14] into CC. The paper explains that CC is a shared figuring model which is present on above the lower cost and uncertain hardware. The treacherous of hardware need the upper level softwares. The framework gives judgment regarding the respect to the continuous condition of the framework and afterwards relegates the assets depends on the evaluation. This type of dynamic observation methods assures the cohesion of the method and constancy of the work viably.

Unique method to meet the cloud buyer asks for and orchestrate the assets within CC environments offered by Chunlin Li and Layuan Li,[15]. CRA procedure is developed and deals with two levels. i) Interaction allying SaaS supplier and SaaS client at the application layer ii) interaction allying SaaS supplier and cloud asset supplier at the resource layer. This paper delineates SaaS arrangement as a numerous best issue. The creator examines limitations like SaaS provider profit, SaaS user budget and deadline, cloud resource provider benefits as well as noticing the SaaS supplier as both supplier and vendor. Moreover, SaaS gives targets at diminishing the installment of utilizing virtual machine from providers, and needs to build the advantages picked up by means of serving SaaS user's solicitation.

Chunlin Li, Layuan Li[16] developed the constitution of various layers in cloud such as IaaS and SaaS and its joint advancement for best asset distribution. The streamlined asset designation issue is divided into sub problems. The created cloud asset assigning algorithm is cultivated by means of iterative algorithm. A versatile asset assigning algorithm for the cloud model with advance emptiable work presented by Jiayin Li, Meikang Qiu[17], algorithms modify the asset designation dependent on redesigned of unique work performance. Adaptive List Scheduling (ALS) and a versatile scheduling Min –Min booking calculations are used for assignment orchestrating which consists unaltered work booking for static asset assigning is incited disconnected. The online versatile advancement is use for re-building up lingering static asset allocation frequently with advanced frequency. In each, re-evaluation process, the schedulers are recompiling the complete time of their specific finished assignment task, not the errands that are allocated to that cloud.

Asset Allotment approach named as SLB developed by Zhenzheong zhang [18], utilized to remunerate the issue of burden unevenness in cloud environment. SLB comprised of two section i) online virtual machine work information statistical examines and asset require estimate ii) A procedure for the motive of burden unevenness which selects an appropriate host in asset bunch dependent on asset needs prediction of virtual machine and the historical load data of hosts. The aim of economical asset to memory utilization and disk I/o throughput, and additionally look into the unchanged burden unevenness when virtual machine running can be extended.

3. Resource Allocation (RA) in clouds

As we already said that, RA is the technique of assigning the assessable assets to required cloud apps in the cloud computing over the internet. RA is the one of the requiring problems in

CC, whereas the scarce resources are shared. From the viewpoint of consumer, RA associated how the services are shared among the consumers. The merits of RA are that the user has no need to install any hardware and software in order to access, develop and hold the application all through the web. There is no restriction of geographical places and carrier. Consumer can arrive the data and application anyplace, on any framework on the planet. During resource scarcity, the cloud providers distributed their resources in the internet. Coming up next are 4 different methods utilizing the figuring limits from cloud supplier have been thought of.

- i. Advance Reservation (AR): Here, the assets are reserved ahead of time and must be accessible in constantly.
- ii. Best effort: Requests are placed in queue and assets are restored as quick as could be expected under the circumstances.
- iii. Immediate: Depends on available resource, either the client request is rejected or the resources are provisioned immediately as per client request.
- iv. Deadline sensitive: if the planning calculation of Haize can guarantee that it very well be accomplished before its time limit, then it is just pre-emptible. It is thought to be pre-emptible but there is restriction to their preempt ability.

3.1 Agent Oriented Resource Allocation:

Recently, there have been some researches are undergoing widely based on cloud resource optimization. The author in this paper[19] had proposed a system for resource fully perfect technique uses map reduce for the cloud resource optimization. PSO is about the ability of group of birds flying towards a destiny through the communication as well as independent searching. Each particle gains and communicated with each other and they can update their newer place. Every particle has personal good and apart from that, there is one universal top. The particle moves with some velocity. The cloud resource optimization framework has been divided into three aspects. i)time ii)revenue iii)user satisfaction. The time and model have furthermore divided. Queuing theory is used to calculate the medium response time with respect to response time model. The proposed algorithm is used to determine the work and to assign the resource in the cloud. Zhengqiu Yangl,[20]has used the ant colony algorithm to assign assets to the consumer. The approach is based on ant's foraging characteristics. The algorithm induces the food searching characteristics of natural ants and embraces the techniques to solve the NP issues. The algorithm is also used to minimize the length of paths. The work agent gathers work from the consumer and ACO algorithm processed on that. The heaviness of pheromones is found which is used for communication. Nodes are chosen by the Ant colony algorithm. This policy is used for cloud resource allocation. Xiao-long Zhen t.al.,[21]had proposed Pareto based fruit fly optimization algorithm(PFOA) to assign resources and task arranging in the cloud. An analytical is developed to trigger the population. Search based on the smell and vision is processed using the algorithm. The searching process extends until the enhanced smell value is not available. The author in this paper [22] developed a system for resource optimization allocation by minimizing the cost. In this article, they had introduced a optimization algorithm to minimize the implementation prize while meeting the needs of consumers. The algorithm is intended for web applications implementation in cloud data center. They had taken the input aspects from the cloud provider and web applications. The algorithm is planned for selecting the most appropriate consolidation of cloud assets. Fan-Hsun Tsengetal.,[23] had used genetic algorithm to forecast the resources which are present in cloud data centers. Further resources are forecasted based on the traditional data in the already existing time slots. The forecasting is completed using GA algorithm.VM are assigned by using the forecasting outputs.

3.2 Priority based Resource Allocation:

In this paper[24], the author represented a method to achieve high priority duty. Moreover, a notion to recycle the virtual machine is presented. New virtual machines are not created in order to run the newly arrived job. The proposed technique has capability to execute a high priority work or duty by hanging the low priority task. The hanged job is reopened in a virtual machine if that virtual machine has done single duty completely. Dilip Kumar.M et.al.,[25] had proposed a system for allocating the resources. They are assumed as peer to peer cloud approach. The demand for larger amount of CPU is taken as higher priority job. K-means algorithm is used to distinguish the jobs into categories such as high, medium, and low priority sets and the job is arranged in the job list based on their priority.

In the proposed method, the cost is estimated according to the current requirements for a resource and it's all times constantly. In high controversy across the web, assets are introduced from the peer clouds. A system to assign resources with Fault Tolerance exposed by Shubhan kankshi Goutamt et.al.,[26] presented. Pre-emption begins from the low priority task and moving towards the high priority task. An advanced reservation model is proposed here. The technique creates a work list based on the priorities and priority-based work arranging is carried out. Advanced reservation provides to job having high priorities. Additionally, this algorithm is also applicable to fault tolerance in cloud applications. Table 1 and Table 2 shows various dynamic resource allocation techniques and QoS based RA techniques.

Table1: Dynamic Resource Allocation Techniques

Techniques	Advantages	Disadvantages	Metrics Improved
A Threshold-based dynamic resource allocation technique[27]	Continuous monitoring and prediction of resource requirements, Reduced usage cost, Increased resource utilization, less resource wastage	The system has longer run time, if VM exceed to threshold value, lags in performance and load balancing	Resource utilization and overhead
Nephele’s structure [28]	Provides efficient parallel processing	It requires user annotations	Average instance utilization
Distributed Multiple Criteria Decision Analysis (MCDA) with PROMETHEE method [29]	Achieves scalability and feasibility	Occurrence of migration overhead	CPU utilization, number of migration
Sandpiper architecture of black-box and gray-box approach [30]	Eliminates hotspots and improves responsiveness	VM swap incurs more overhead	CPU utilization, completion time and migration
MeasureForecast-Remap (MFR) algorithm [31]	Reduces SLA violations and resourced consumption	High computational cost	CPU utilization, relative frequency and number of used PMs
Evaluation of a dynamic resource assigning approach [32]	Users add or remove their resources dynamically according to demands	Unexpected demand introduces allocation failure	Response time and average load

Table 2: QoS Based Resource Allocation Technique

Techniques	Advantages	Disadvantages	Metrics Improved
Profit based scheduling and admission control	Higher profit-rate variance focuses on	Does not focus on the user side	Profit rate, profit variance and average

algorithm [33]	resource provider's side		job size
ProfminVmMinAvaiSpace [34]	Maximizes SaaS provider's profit, and minimizes SLA violations	Residual information of VM creates security issues	Arrival rate, SLA violations, penalty cost and total cost
Openstack version of generic SLA manager [35]	Achieves QoS during migration	Introduces less penalty	SLA violations and availability
Adaptive, iterative, soft resource allocation algorithm [36]	Maximizes utilization and throughput while guaranteeing QoS	Does not discuss optimal allocation	Profit, Good put and throughput
Genetic Algorithm (GA)-based automatic price adjusting algorithm [37]	Improves resource utilization, and achieves equilibrium state of demand	Resource wastage due to uncertainty demand	Price, demand and benefit
Deadline-driven provisioning mechanism [38]	Minimizes application execution time	Does not applicable to HPC-data intensive applications	Deadline, execution time and usage cost
Customer-driven SLA based resource provisioning algorithm (BFResvResource and BFReschedReq) [39]	Minimizes cost and SLA violations, and improves customer satisfaction level	Does not applicable for compute-intensive and multi-tier applications	Total cost, SLA violation and service quality improvement
Greedy Particle Swarm Optimization algorithm [40]	Minimizes total cost of workload, and enhances VM configuration	Solves only short-term problems	Estimated cost and cost improvement

4. Adaptive Management of Resources:

The obscurity method has presented the trust in controlling apps and resources in the Cloud Computing system. Consequently, this paper analysis the evaluation and application of obscurity methods in the Cloud Computing section. Brownout is an independent adaptive framework that qualify or disqualify the arbitrary sectors such as application components or services in the model to control the uncertain workload.[41] The motive behind this brownout method is to be adaptive. Arbitrary sections may be disenabled provisionally so that the important activity of the model is assured and apps evade saturation. Deactivating particular arbitrary activities can leads to maximize the rate of request acceptance by employing those resources for essential activities. Additionally, brownout permits to enhance the resource promotion when keeping the apps responsive to evade the workloads.

4.1 Brownout Approach-Evolution:

The brownout application in Cloud Computing had proposed by Klein et.al.,[41] who discovered brownout as an independent versatile application framework in the year 2014. They also presented the brownout which can maximize the system reliability when uncertain demands are received by the system. In this paper [42] and [43] two application prototypes were proposed which had been used in follow up search widely. This system presented the better performance of the brownout method, even though the executions were only compensated on an individual machine. In this paper [44], the author had presented the burden adjustment plans for application copies depends on brownout and entry handle to enhance the application execution and enlarge the system flexible. Nevertheless, the restriction is that framework application is not systematic. To discover faults before and evade the restrictions of time interval, event driven observing is presented. Two load balancing algorithms for

brownout-aware assignments introduced by Klein et.al., The outputs presented that the proposed method has efficient performance in fault tolerance when compared existing works.

Maggio et.al.,[46] presented and examined a various kinds of control policies for brownout aware apps to evade the work overloads. Forecast for arriving load and services time are also applied in the presented method. To evade Service Level Objective (SLO) infraction and insensitive, an environment to control the resources adaptively for flexible clouds according to SLOs is innovated by Nikolov et.al., [47] It can vanquish short time request issues. Consequently, fault handling is not seen this work. Tomas et.al., connected overbooking and brownout together to enhance resource employment without the application humiliation. The automatic algorithms depend on brownout got more response in the year 2015.Desmerurus et.al., [49] proposed an event driven brownout method to examine the tradeoff value allying employment and response time for web applications. Additionally, different automatic strategies on machine learning and admission control were introduced in this work .This work leads a way to propose better energy efficient cloud data centers, but the output was not evolved under a real time environment. One more method presented by Dupont et.al.,[50] to control cloud flexibility in both hardware and software. The proposed technique analyses and takes merits of dynamic selection of various policies.

4.2 Dynamic Load Balancing:

The dynamic load balancing is the process of making the load balancing distribution decision as per the present status of the system. Additionally, they needn't bother with any before information about the framework assets[51]. This methodology is appropriate for heterogeneous system. Load balancing decision isn't static as rule, which means load balancing decision can be at run time[52]. This method offers excellent enhancement in performance rather than the static approach. Ant colony, throttled, GA, Honey Bee algorithm etc., are not many instances of dynamic algorithm. Throttled Load Balancing Algorithm [53][54] is a dynamic algorithm for sharing the workload that actualize on virtual machines completely. Here the virtual machines are bunched based on request they handle. If the client sends the request, the load balancer gets alert immediately and analyze for the unit which can handle easily and assigns that requests. yet, the issue is that load balancer needs to investigate for the appropriate virtual machine gathering, which prompts activity delay.

Conclusion

Cloud Computing is remarkable technology in field of industries and venture markets. The height of Cloud Computing is trading the standpoint of data innovation emphatically. In the end it turns the advantage processing into real-time. In framework of cloud, an asset designation method is needed for achieving client prerequisites and improving the advantage for cloud specialist. A few approaches for asset distribution had been introduced and sent so far with regards to Cloud Computing. Resource Allocation is being one of the most logical components which widespread exhibition of the cloud. Exclusively, it has been relying upon the specialized perspectives tended to up until now. But CC being a rapidly broadening the undertakings in the IT companies must also necessitate the use of components interrelated to user's resource assignment terms. This method could give the best outcomes in overall performance and also fulfillment of consumer for the equivalent.

REFERENCES

- 1) Munich, Gerald kaefer "Cloud Computing Architecture" IEEE Spectrum, February 2009
- 2) Dikaiakos, M.D.; Katsaros, D.; Mehra, P.; Pallis, G.; Vakali, A., "Cloud Computing: Distributed Internet Computing for IT and Scientific Research," Internet Computing, IEEE , Sept.-Oct. 2009

- 4) Talib, A.M.; Atan, R.; Abdullah, R.; Azrifah, M., "CloudZone: Towards an integrity layer of cloud data storage based on multi agent system architecture," Open Systems (ICOS) IEEE Conference on , 25-28 Sept. 2011.
- 5) Dan C. Marinescu, "Cloud Computing Theory and Practice", Elsevier, 2013
- 6) Seyyed Mohsen Hashemi, Amid KhatibiBardsiri, "Cloud Computing Vs. Grid Computing", ARPN Journal of Systems and Software ,May 2012.
- 7) Ahmed Shawish and Maria Salama,"Cloud Computing: Paradigms and Technologies"Springer, 2014
- 8) Miss. Rudra Koteswaramma, "Client-Side Load Balancing and Resource Monitoring in Cloud", International Journal of Engineering Research and Applications, November- December 2012
- 9) JankiAkhani, Sanjay Chuadhary and Gaurav Somani, "Negotiation for Resource Allocation in IaaS Cloud", Bangalore Compute Conf., pp 15 ACM, 2011
- 10) Amit Nathani, Sanjay Chaudhary, and Gaurav Somani, Policy based resource allocation in IaaS cloud. Future Gener. Comput. Syst. , 2012
- 11) Mohan, N.R.R.; Raj, E.B., "Resource Allocation Techniques in Cloud Computing -- Research Challenges for Applications, " Computational Intelligence and Communication Networks(CICN), 2012 Fourth International Conference on , 3-5 Nov. 2012
- 12) Villegas, D.; Antoniou, A.; Sadjadi, S.M.; Iosup, A., "An Analysis of Provisioning and Allocation Policies for Infrastructure-as-a-Service Clouds," Cluster, Cloud and Grid Computing (CCGrid), 12th IEEE/ACM International Symposium on , 13-16 May 2012
- 13) Hemant Kumar Mehta and Eshan Gupta. Economy Based Resource Allocation in IaaS Cloud. Int. J. Cloud Appl. Comput. , April 2013
- 14) Gihun Jung; Kwang-Mong Sim, "Agent-Based Adaptive Resource Allocation on the Cloud Computing Environment," Parallel Processing Workshops (ICPPW), 40th International Conference on , 13-16 Sept. 2011
- 15) Yanbing Liu,Shasha Yang, Qingguo Lin and Gyoung-Bae Kim "Loyalty-Based Resource Allocation Mechanism in Cloud Computing",Recent Advances in CSIE, Springer, 2012
- 16) Chunlin Li and Layuan Li, "Resource allocation in cloud computing: model and Algorithm",Int. J.Web and Grid Services, 2013
- 17) Chunlin Li • Layuan Li, "Efficient resource allocation for optimizing objectives of cloud users, IaaS provider and SaaS provider in cloud environment",Springer, 2013
- 18) Jiayin Li, MeikangQiu, Jian-Wei Niu, Yu Chen, Zhong Ming, "Adaptive Resource Allocation for Pre-empt able Jobs in Cloud Systems," in 10th International Conference on Intelligent System Design and Application, Jan. 2011.
- 19) Zhenzhong Zhang; Haiyan Wang; Limin Xiao; Li Ruan, "A statistical based resource allocation scheme in cloud," Cloud and Service Computing (CSC), International Conference on ,12-14 Dec. 20
- 20) Zhengqiu Yang1,"study on cloud resource allocation strategy based on particle swarm ant colony optimization algorithm"/2012 IEEE
- 21) [20] Xiao-long Zheng,"A Pareto based Fruit Fly Optimization Algorithm for Task Scheduling and Resource Allocation in Cloud Computing Environment"/2016 IEEE
- 22) Fan-Hsun Tseng,"Dynamic Resource Prediction and Allocation for Cloud Data Center Using the Multiobjective Genetic Algorithm"/2017 IEEE
- 23) Seyedehmehrnaz Mireslami, "Simultaneous Cost and QoS Optimization for Cloud Resource Allocation"IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, VOL. 14, NO. 3, SEPTEMBER 2017

- 24) Tseng, Fan-Hsun & Wang, Xiaofei & Chou, Li-Der. (2017). Dynamic Resource Prediction and Allocation for Cloud Data Center Using the Multiobjective Genetic Algorithm. *IEEE Systems Journal*. 12. 1-12. 10.1109/JSYST.2017.2722476.
- 25) Saraswathi AT a, Kalaashri.Y.RA b, Dr.S.Padmavathi c1, "Dynamic Resource Allocation Scheme in Cloud Computing"
- 26) Dilip Kumar S. M, Naidila Sadashiv, R. S. Goudar, "Priority Based Resource Allocation and Demand Based Pricing Model in Peer-to-Peer Clouds"/2014 IEEE
- 27) Shubhakankshi Goutam, Arun Kumar Yadav, "Preemptable Priority Based Dynamic Resource Allocation in Cloud Computing with Fault Tolerance"/2015 IEEE
- 28) Weiwei Lin, James Z. Wang, Chen Liang, Deyu Qi., 2011, "A Threshold-based Dynamic Resource Allocation Scheme for Cloud Computing" *Elsevier Transaction on Procedia Engineering*, vol. 23, pp. 695-703
- 29) Daniel Warneke and Odej Kao., 2011, "Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud", *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, pp. 985-997.
- 30) Yağız Onat Yazır, Chris Matthews, Roozbeh Farahbod, Stephen Neville, Adel Guitouni, Sudhakar Ganti and Yvonne Coady., 2010, "Dynamic Resource Allocation in Computing Clouds using Distributed Multiple Criteria Decision Analysis" *IEEE 3rd International Conference on Cloud Computing*, pp. 91-98.
- 31) T. Wood, P. Shenoy, A. Venkataramani, and M. Yousif., 2007, "Black-box and Gray-box Strategies for Virtual Machine Migration," in *Proc. of the 4th USENIX Symposium on Networked Systems Design and Implementation*, pp. 229-242.
- 32) N. Bobroff, A. Kochut, and K. Beaty., 2007, "Dynamic Placement of Virtual Machines for Managing SLA Violations," in *Proceedings of the IFIP/IEEE International Symposium on Integrated Network Management*, pp. 119-128.
- 33) Atsuo Inomata, Taiki Morikawa, Minoru Ikebe, Yoshihiro Okamoto, Satoru Noguchi, Kazutoshi Fujikawa, Hideki Sunahara, Sk. Md. Mizanur Rahman., 2011, "Proposal and Evaluation of a Dynamic Resource Allocation Method based on the Load of VMs on IaaS", *IEEE 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, pp. 1-6.
- 34) I. Popovici et al., 2005 "Profitable Services in an Uncertain World", In *proceedings of the ACM/IEEE Conference on Supercomputing*, vol. 36.
- 35) Linlin Wu, Saurabh Kumar Garg and Raj kumar Buyya., 2011, "SLA based Resource Allocation for SaaS Provides in Cloud Computing Environments", *IEEE*, pp.195-204.
- 36) Kuan Lua, Ramin Yahyapoura, Philipp Wiedera, Constantinos Kotsokalisb, Edwin Yaquba, Ali Imran Jehangiria., 2013, "QoS-based Resource Allocation Framework for Multidomain SLA Management in Clouds", *International Journal of Cloud Computing*, vol. 1, no.1.
- 37) Jack Li, Qingyang Wang, Deepal Jayasinghe, Simon Malkowski, Pengcheng Xiong, Calton Pu, Yasuhiko Kanemasa, Motoyuki Kawaba., 2012, "Profit-Based Experimental Analysis of IaaS Cloud Performance: Impact of Software Resource Allocation", *IEEE Ninth International Conference on Services Computing (SSC)*, pp. 344-351.
- 38) Xindong You, Jian Wan, Xianghua Xu, Congfeng Jiang, Wei Zhang, Jilin Zhang., 2011, "ARAS M: Automatic Resource Allocation Strategy based on Market Mechanism in Cloud Computing", *JOURNAL OF COMPUTERS*, vol. 6, no. 7, pp.1287-1296.
- 39) Christian Vecchiola, Rodrigo N. Calheiros, Dileban Karunamoorthy, Rajkumar Buyya., 2012, "Deadline-Driven Provisioning of Resources for Scientific Applications

- in Hybrid Clouds with Aneka”, Elsevier Transaction on Future Generation Computer Systems, vol. 28, pp. 58–65.
- 40) Linlin Wu , Saurabh Kumar Garg , Steve Versteeg, and Rajkumar Buyya., 2014, “SLA-Based Resource Provisioning for Hosted Software-as-a-Service Applications in Cloud Computing Environments”, IEEE Transactions on Services Computing, vol. 7, no.3, pp. 465-485.
 - 41) Fatma A. Omara, Sherif M. Khattab, and Radhya Sahal., 2014, “Optimum Resource Allocation of Database in Cloud Computing”, Egyptian Informatics Journal, vol.15, pp.1–12
 - 42) Cristian Klein, Martina Maggio, Karl-Erik Årzén, and Francisco Hernández-Rodriguez. 2014a. Brownout: Building more robust cloud applications. In Proceedings of the 36th International Conference on Software Engineering. 700–711.
 - 43) RUBiS. 2009. RUBiS. RUBiS: Rice University bidding system. Retrieved March 27, 2018 from <http://rubis.ow2.org/>.
 - 44) RUBBoS. 2005. RUBBoS: Bulletin Board Benchmark. Retrieved March 27, 2018 from <http://jmob.ow2.org/rubbos.html>.
 - 45) Jonas Dürango, Manfred Dellkrantz, Martina Maggio, Cristian Klein, Alessandro Vittorio Papadopoulos, Francisco Hernández-Rodriguez, Erik Elmroth, and Karl-Erik Årzén. 2014. Control-theoretical load-balancing for cloud applications with brownout. In Proceedings of the 2014 IEEE 53rd Annual Conference on Decision and Control. 5320–5327.
 - 46) Cristian Klein, Alessandro Vittorio Papadopoulos, Manfred Dellkrantz, Jonas Dürango, Martina Maggio, Karl-Erik Årzén, Francisco Hernández-Rodriguez, and Erik Elmroth. 2014b. Improving cloud service resilience using brownout-aware load-balancing. In Proceedings of the 2014 IEEE 33rd International Symposium on Reliable Distributed Systems. 31–40.
 - 47) Martina Maggio, Cristian Klein, and Karl-Erik Årzén. 2014. Control strategies for predictable brownouts in cloud computing. IFAC Proceedings Volumes 47, 3 (2014), 689–694
 - 48) Vladimir Nikolov, Steffen Kächele, Franz J. Hauck, and Dieter Rautenbach. 2014. Cloudfarm: An elastic cloud platform with flexible and adaptive resource management. In Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing. 547–553.
 - 49) Luis Tomás, Cristian Klein, Johan Tordsson, and Francisco Hernández-Rodríguez. 2014. The straw that broke the camel’s back: Safe cloud overbooking with application brownout. In Proceedings of the 2014 IEEE International Conference on Cloud and Autonomic Computing. 151–160.
 - 50) David Desmeurs, Cristian Klein, Alessandro Vittorio Papadopoulos, and Johan Tordsson. 2015. Event-driven application brownout: Reconciling high utilization and low tail response times. In Proceedings of the 2015 International Conference on Cloud and Autonomic Computing. 1–12.
 - 51) Simon Dupont, Jonathan Lejeune, Frederico Alvares, and Thomas Ledoux. 2015. Experimental analysis on autonomic strategies for cloud elasticity. In Proceedings of the 2015 IEEE International Conference on Cloud and Autonomic Computing. 81–92.
 - 52) Desai, Tushar, and J. Prajapati. "A survey of various load balancing techniques and challenges in cloud computing." International Journal of Scientific & Technology Research 2.11 (2013): 158-161
 - 53) [52] Raghava, N. S., and Deepti Singh. "Comparative study on load balancing techniques in cloud computing." Open Journal of Mobile Computing and Cloud Computing 1.1 (2014).

- 54) [53]S. Mohapatra, K. Smruti Rekha, S. Mohanty, "A comparison of Four Popular Heuristics for Load Balancing of Virtual Machines in Cloud Computing," IJCA Journal, vol. 68(6), pp. 34-38, April 2013.