A Review of Dynamic Resource Allocation Framework for Large Amount of Cloud Enterprises

B Vijaya Laxmi[a], Dr. Harsh Pratap Singh [b] and Dr. Agastyaraju Nagaraja Rao[c],

[a] *Research Scholar, Dept. of Computer Science & Engineering,*
Sri Satya Sai University of Technology & Medical Sciences, Sehore, Bhopal Indore Road, Madhya Pradesh, India
[b] *Research Guide, Dept. of Computer Science & Engineering,*
Sri Satya Sai University of Technology & Medical Sciences, Sehore, Bhopal Indore Road, Madhya Pradesh, India
[c] *Research Co-Guide, Dept. of Computer Science & Engineering,VIT University, Vellore, T.N.*

Article History: Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

Abstract: Cloud computing is an on-demand service because it offers dynamic flexible resource allocation for reliable and guaranteed services in pay as-you-use manner. Because of the consistently increasing demands of the clients for services or resources, it gets hard to allocate resources accurately to the client demands to satisfy their solicitations and also to take care of the Service Level Agreements (SLA) gave by the service suppliers. Dynamic resource allocation problem is one of the most challenging problems in the resource management problems. The dynamic resource allocation in cloud computing has attracted attention of the research network in the last couple of years. Many researchers around the world have thought of new ways of facing this challenge. Ad-hoc parallel data handling has arisen to be one of the executioner applications for Infrastructure-as-a-Service (IaaS) cloud. Number of Cloud supplier companies has started to incorporate frameworks for parallel data handling in their item which making it easy for clients to access these services and to convey their programs. The handling frameworks which are at present utilized have been intended for static and homogeneous bunch arrangements. So the allocated resources may be inadequate for large parts of the submitted tasks and unnecessarily increase preparing cost and time. Again because of opaque nature of cloud, static allocation of resources is conceivable, yet the other way around in dynamic situations. The proposed new generic data handling framework is expected to expressly misuse the dynamic resource allocation in cloud for task scheduling and execution.

Keywords: Cloud Computing, Dynamic Resource Allocation, Resource Management, Resource Scheduling, Map Reduce

1 Introduction

Cloud Computing is an essential element of present day computing frameworks. Computing concepts, innovation and architectures have been created and consolidated in the last decades; many aspects are dependent upon technological evolution and revolution. Cloud Computing is a computing innovation that is rapidly consolidating itself as the following stage in the turn of events and organization of increasing number of circulated application. Presently various companies have to handle large amounts of data in a cost-effective manner. These companies are operators of Internet search motors, similar to Yahoo, Google or Microsoft. The immense amount of data or datasets they have to deal with consistently has made traditional database solutions restrictively costly. So these quantities of developing companies have popularized an architectural paradigm based on an immense number of ware workers. Problems like regenerating a web file or preparing crawled archives are part into several autonomous subtasks, circulated among the available hubs, and figured in parallel. The cloud computing paradigm makes the resource as a solitary purpose of access to the quantity of customers and is executed as pay per use basis. In spite of the fact that there are number of advantages of cloud computing, for example, virtualized environment, furnished with dynamic infrastructure, pay per consume, totally liberated from software and hardware installations, recommended infrastructure and the major concern is the request where the solicitations are satisfied which advances the scheduling of the resources. Allocation of resources has been made effectively that maximizes the framework utilization and overall performance. We proposed another Generic Framework which dynamically allocates resources to the data handling applications. The goal of our proposed approach is to decrease the execution time, migration time for resources and organization latency.

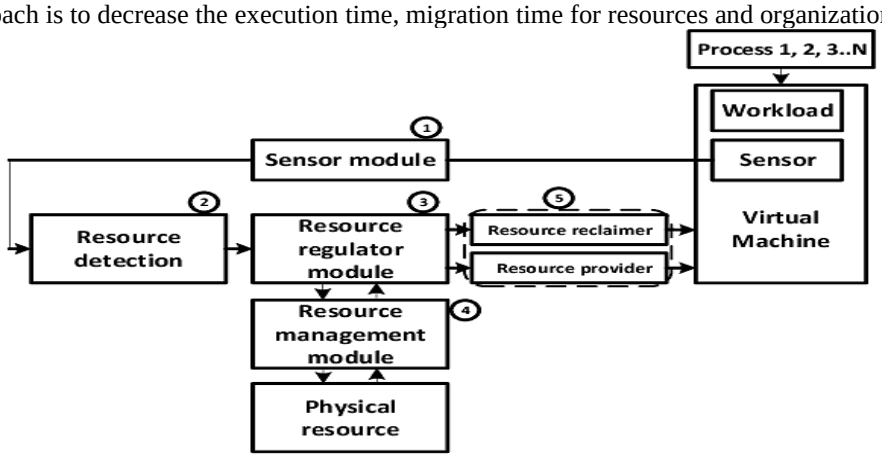

Figure 1.1 Conceptual Frameworks

A Cloud is a kind of parallel and dispersed framework consisting of a collection of interconnected and virtualized PCs that are dynamically provisioned and introduced as at least one bound together computing resources based on service-level agreements established through negotiation between the service supplier and consumers. Cloud computing arises as another computing paradigm which aims to give reliable, modified and QoS (Quality of Service) guaranteed computing dynamic environments for end-clients. Cloud computing is the conveyance of computing as a service rather than an item, whereby shared resources, software and information are given to clients over the organization. Cloud computing suppliers convey application via the Internet, which are accessed from internet browser, while the business software and data are put away on workers at a far off location. Cloud suppliers are able to attain the agreed SLA, by scheduling resources in productive manner and by sending application on legitimate VM according to the SLA objective and at the same time performance of the applications should be upgraded. Compared to past paradigms, cloud computing centres around treating computational resources as measurable and billable utilities. From the customers' perspective, cloud computing gives an abstraction of the hidden hardware architecture. This abstraction saves them the expenses of plan, arrangement and maintenance of a data community to have their Application Environments (AE). Whereas for cloud suppliers, the arrangement returns an occasion to benefit by facilitating many AEs. This economy of scale gives advantages to the two players, however leaves the suppliers in a position where they should have a proficient and practical data community. The main goal is to decrease the overloads of the main cloud and increase the performance of the cloud. Lately ad-hoc parallel data handling has arisen to be one of the executioner applications for Infrastructure-as-a-Service (IaaS) clouds. Major Cloud computing companies have started to integrate frameworks for parallel data handling in their item portfolio, making it easy for clients to access these services and to send their programs. Notwithstanding, the preparing frameworks which are right now utilized have been intended for static, homogeneous bunch arrangements and disregard the particular nature of a cloud.

DYNAMIC RESOURCE ALLOCATION

Resource Allocation Strategy (RAS) is all about integrating cloud supplier activities for using and allocating scarce resources inside the restriction of cloud environment in order to address the issues of the cloud application. It requires the sort and amount of resources required by each application to finish a client work. The request and season of allocation of resources are also a contribution for an optimal RAS.

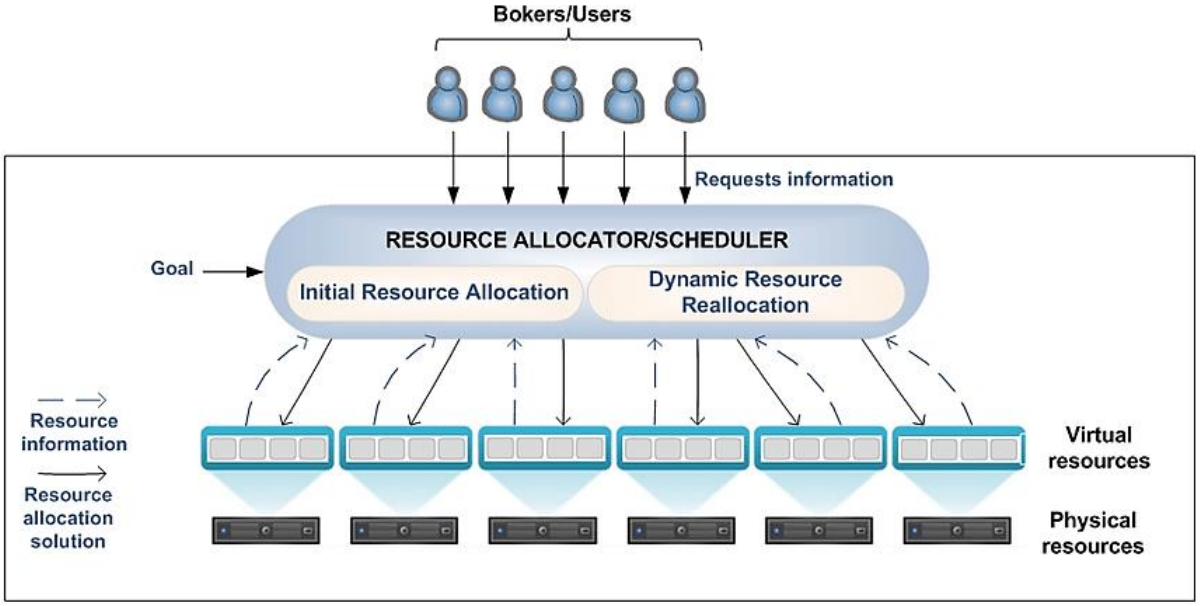

Figure 1.2 Resource Allocation in Cloud Computing

From the viewpoint of a cloud supplier, foreseeing the dynamic nature of clients, client demands, and application demands are impractical. For the cloud clients, the work ought to be finished on time with minimal expense. Subsequently because of restricted resources, resource heterogeneity, locality restrictions, environmental necessities and dynamic nature of resource demand, we need an effective resource allocation framework that suits cloud environments. Cloud resources consist of physical and virtual resources. The physical resources are shared across different register demands through virtualization and provisioning. The solicitation for virtualized resources is portrayed through a bunch of parameters detailing the preparing, memory and plate needs. Provisioning satisfies the solicitation 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 leased. The multifaceted nature of finding an ideal resource allocation is exponential in enormous

frameworks like huge groups, data focuses or Grids. Since resource demand and supply can be dynamic and uncertain.

Literature Review

Amir Mosavi et al (2017),Using dynamic resource allocation for load balancing is considered as an important optimization measure in cloud computing. To achieve maximum resource productivity and scalability in a rapid manner this cycle is concerned with various destinations for a viable distribution of loads among virtual machines. In this realm, investigating new algorithms, as well as improvement of novel algorithms, is profoundly wanted for technological advancement and continued advancement in resource allocation application in cloud computing. Accordingly, this paper investigates the application of two relatively new optimization algorithms and further proposes a cross breed algorithm for load balancing which can contribute well in maximizing the throughput of the cloud supplier's organization. The proposed algorithm is a crossover of teaching-learning-based optimization algorithm (TLBO) and gray wolves optimization algorithm (GW). The half and half algorithm performs more proficiently than using each and every one of these algorithms. Moreover, it well balances the needs and successfully considers load balancing based on schedule, cost, and avoidance of local ideal traps, which consequently leads to minimal amount of waiting time. To evaluate the adequacy of the proposed algorithm, a comparison with the TLBO and GW algorithms is conducted and the experimental outcomes are introduced.

Sabbi Vamshi Krishna, et al (2020),Conventional occupation management frameworks (JMS) consider physical resources alone as computational resources. Computational resources are relegated an assignment as calculation of hubs in HPC (High Performance Computing) bunch frameworks and the work strategies are executed legitimately on the allotted computing hubs. It is seen that the decrease in organization latency and better bandwidth usage in SDN is accomplished exactly when it has dynamic allocation of resources and occupation scheduling standards. In this way in this overview it is illustrated about various effective dynamic resource allocation plans alongside better occupation scheduling procedures. On a similar time the slacking in each strategy are talked about and from that the answer for improving the resource allocation in HPC found. In this way this audit may give an approach to characterize better and capable resource allocation and occupation scheduling plans to improve the performance of HPC with SDN.

Pratik P. Pandya et al (2014), Dynamic resource allocation is a lot of popular research area in cloud environment because of its live application in data place. Because of dynamic and heterogeneous nature of cloud, allocation of virtual machine is affected by various parameters like QOS, time consumption, cost, carbon impact and so on Gathering of virtual machine which is communicate to each other to execute one large solicitation is comes into affinity gathering. Here we will consider details of allocation strategy, affinity of virtual machine and how it will give great performance over non-affinity gathering and give some idea about new method to improve performance which we will actualize in future.

P. Prathap Nayudu et al (2018),Cloud computing environment provisions the stockpile of computing resources on the basis of demand, as and when required. It expands upon the advances of virtualization and conveyed computing to help cost effective usage of computing resources, emphasizing on resource scalability and on-demand services. It allows business results to scale here and there their resources based on requirements. Managing the client demand creates the challenges of on demand resource allocation. Further, they can make utilization of companywide access to applications, based on a pay-as-you-go model. Henceforth there is no requirement for getting licenses for individual items. Virtual Machine (VM) innovation has been utilized for resource provisioning. It is normal that utilizing virtualized environment will decrease the average occupation response time as well as executes the task according to the availability of resources. Successful and dynamic utilization of the resources in cloud can assist with balancing the load and avoid situations like moderate run of frameworks. In this paper, various resource allocation strategies and their challenges are examined in detail. It is accepted that this paper would profit both cloud clients and researchers in defeating the challenges faced.

Ali Belgacem et al (2020),the dynamic resource allocation is a decent feature of the cloud computing environment. In any case, it faces significant problems as far as service quality, fault tolerance, and energy consumption. It was necessary, at that point, to locate a compelling technique that can viably address these important issues and increase cloud performance. This paper presents a dynamic resource allocation model that can fulfill client need for resources with improved and faster responsiveness. It also proposes a multi-target search algorithm called Spacing Multi-Objective Antlion algorithm (S-MOAL) to limit both the make span and the expense of utilizing virtual machines. In addition, its impact on fault tolerance and energy consumption was contemplated. The simulation revealed that our strategy performed in a way that is better than the PBACO, DCLCA, DSOS and MOGA algorithms, especially as far as make span.

Chandrasekhar S. Pawar et al (2013),Today Cloud computing is on demand as it offers dynamic flexible resource allocation, for reliable and guaranteed services in pay-as-you-use manner, to Cloud service clients. So there should be a provision that all resources are made available to mentioning clients in proficient manner to satisfy their requirements. This resource provision is done by considering the Service Level Agreements (SLA)

and with the assistance of parallel handling. On-going work considers various strategies with single SLA parameter. Henceforth by considering different SLA parameter and resource allocation by pre-emption mechanism for high need task execution can improve the resource utilization in Cloud. In this paper we propose an algorithm which considered Pre-emptible task execution and different SLA parameters, for example, memory, network bandwidth, and required CPU time. An obtained experimental outcomes show that in a situation where resource contention is furious our algorithm gives better utilization of resources.

Mohit Kumar et al (2017),The most challenging problem for a cloud service supplier is maintaining the quality of service parameters like reliability, elasticity, keeping the deadline and limiting the makes pan time as also the task rejection ratio. Along these lines, the cloud service supplier needs a dynamic task scheduling algorithm that lessens the makes pan time while increasing the utilization ratio of cloud resources and meeting the client characterized QoS parameters. In this paper, we have built up a dynamic scheduling algorithm that balances the workload among all the virtual machines with elastic resource provisioning and deprovisioning based on the last optimal k-interval. Further, the algorithm has been tried on variable number of tasks (10 to 30) to achieve better scalability.

Yong Lu et al (2017),With the expanding of its scale and the energy cost factors being overlooked in green cloud computing, the problem of high energy cost and low effectiveness is uncovered. Based on the concepts and standards of load balancing, a novel energy-effective load balancing global optimization algorithm, called resource-aware load balancing clonal algorithm for task scheduling, is proposed to deal with the problem of energy consumption in green cloud computing. Initially, the problem is formulated as a combinatorial optimization problem that aims to improve both energy consumption and load balancing. At that point, the resource-aware scheduling algorithm is proposed based on load balancing strategy and clonal selection guideline. Finally, simulation examines show that the proposed algorithm can adequately lessen energy consumption in green cloud computing, and its exploration and exploitation abilities can be enhanced and even.

Liyun Zuo et al (2015),for task-scheduling problems in cloud computing, a multi-target optimization strategy is proposed here. In the first place, with an aim toward the biodiversity of resources and tasks in cloud computing, we propose a resource cost model that characterizes the demand of tasks on resources with more details. This model mirrors the relationship between the client's resource costs and the spending costs. A multi-target optimization scheduling technique has been proposed based on this resource cost model. This technique considers the make span and the client's spending costs as constraints of the optimization problem, achieving multiobjective optimization of both performance and cost. An improved ant colony algorithm has been proposed to take care of this problem. Two constraint functions were utilized to evaluate and give feedback regarding the performance and spending cost.

Zhen Xiao et al (2013),Cloud computing allows business clients to scale here and there their resource usage based on requirements. Many of the promoted gains in the cloud model come from resource multiplexing through virtualization innovation. In this paper, we present a framework that utilizes virtualization innovation to allocate data focus resources dynamically based on application demands and backing green computing by upgrading the quantity of workers being used. We present the concept of "skewness" to measure the lop-sidedness in the multi-dimensional resource utilization of a worker. By limiting skewness, we can consolidate various sorts of workloads pleasantly and improve the overall utilization of worker resources. We build up a bunch of heuristics that forestall overload in the framework adequately while saving energy utilized. Trace driven simulation and investigation results demonstrate that our algorithm achieves great performance.

Conclusion

Cloud computing technology is increasingly being utilized in enterprises and business markets. In cloud paradigm, a compelling resource allocation strategy is needed for achieving client satisfaction and maximizing the benefit for cloud service suppliers. This paper summarizes the classification of RAS and its impacts in cloud framework. A portion of the strategies talked about above mainly centre around CPU, memory resources, yet are lacking in certain factors. Subsequently this review paper will ideally motivate future researchers to think of smarter and made sure about optimal resource allocation algorithms and framework to fortify the cloud computing paradigm. This paper zeroed in on Hadoop MapReduce resource allocation management methods for multicluster environments. It proposes a novel dynamic space allocation strategy to improve the performance of yarn scheduler and eliminates the load balancing problem. This work demonstrates that the dynamic opening allocation performs in a way that is better than the yarn framework. In future, it is prescribed to concentrate on CPU bandwidth and handling time.

References

1. Seyedmajid Mousavi, Amir Mosavi, Anna-Maria R. VarkonyiKoczy, Gabor Fazekas, “Dynamic Resource Allocation in Cloud Computing” (2017)
2. Sabbi Vamshi Krishna, Dr. Azad Shrivastava, Dr. Sunil J. Wagh, Dr. T.V. Prasad, “Dynamic Resource Allocation and Job Scheduling To Enhance the Performance of HPC with SDN - A Review” (2020)

3. Pratik P. Pandya, Hitesh A. Bheda, “Dynamic Resource Allocation Techniques in Cloud Computing” (2014)
4. P. Prathap Naidu, K. Raja Sekar, “Cloud Environment: A Review on Dynamic Resource Allocation Schemes” (2018)
5. Ali Belgacem1, Kadda Beghdad-Bey, Hassina Nacer, Sofiane Bouznad, “Efficient dynamic resource allocation method for cloud computing environment” (2020)
6. Chandrasekhar S. Pawar, Rajnikant B. Wagh, “Priority Based Dynamic Resource Allocation in Cloud Computing with Modified Waiting Queue” (2013)
7. Mohit Kumar, S.C. Sharma, “Deadline constrained based dynamic load balancing algorithm with elasticity in cloud environment” (2017)
8. Yong Lu, Na Sun, “An effective task scheduling algorithm based on dynamic energy management and efficient resource utilization in green cloud computing environment” (2017)
9. Zhen Xiao, Weijia Song, and Qi Chen, “Dynamic Resource Allocation using Virtual Machines for Cloud Computing Environment” (2013)
10. Kong, Zhen, Cheng-ZhongXu, and MinyiGuo. "Mechanism design for stochastic virtual resource allocation in non- cooperative cloud systems." Cloud Computing (CLOUD), 2011 IEEE International Conference on. IEEE, 2011
11. Endo, Patricia Takako, et al. "Resource allocation for distributed cloud: concepts and research challenges." IEEE network25.4 (2011)
12. Minarolli, Dorian, and Bernd Freisleben. "Utility-based resource allocation for virtual machines in cloud computing. “Computers and Communications (ISCC), 2011 IEEE Symposium on. IEEE, 2011