A Review on Efficiently Allocating Resources in a Dynamic Cloud Environment

Ananya Aggarwal[1], Rohini Mahajan[2]

Department Of Computer Application ,Chandigarh School Of Business,Chandigarh Groups Of College ,Jhanjeri, Mohali, India

[1]ananya20aggarwal@gmail.com, [2]er.rohinimahajan@gmail.com

Abstract

The need for computing resources becomes more demanding, needs change and workloads fluctuate at lightning speeds in a highly dynamic cloud computing environment. Like juggling servers instead of balls. Throw in several types of cloud players and their conflicting goals and you are right in the middle of the chaos management! Both academia and industry are working on this complicated issue, however preliminary analyses usually omit highly important information such as resources allocation, selection and optimization. This research, in turn, aims to fill this gap by revealing the latest "juggling tricks" that cloud providers are utilizing. By type, we distinguish them and discuss their pros, cons and similarities. Furthermore, we show the world of limitless opportunities, and more.

Keywords:Resource,Allocation,Cloud Computing,Virtualization,Optimization, Utilization.

I. INTRODUCTION

Cloud computing with the characteristics of elasticity and responsiveness is so critical to the processes that if not done prudently, business organizations will likely not realise the full benefits of the cloud. For instance, if an investment bank is managing a business and involves some calculation, like financial modelling. In case the company were to carry out the estimation itself, it will take a significant amount of time under the limited affordable computing resources that are accessible to it such as the network, CPU, and memory [1]. It decreases profits for the company. It also means an additional expense as it pertains to keeping up with its own data centre and running of computing facilities which are both manpower and computer resource intensive. Hence, the computation should be done by the third party like the cloud since it would be more cost-effective. Instability and dynamism of cloud systems (increased loads, altered user requirements, priority switches) leads to both resource management opportunities and obstacles. undefined

Current: which consists of frequently changing pursuits by the customer.

target resources: focusing mainly on resources that are in great request.

optimization: providing an optimal utilization of the available resources.

Scheduling: work on major tasks more carefully to improve performance; and

Power: more effective resource allocation after power reduces[2].

A. *The Development of Cloud Computing*

The strategy, IT infrastructure design and IT services management within an organization has evolved immensely with the development of cloud computing. Indeed, from the beginning of utility computing (UC) to the introduction of infrastructure as a service (IaaS).

PaaS and SaaS models of cloud computing have helped to have their globally utilizable computing resources open to a wider range of organizations and they are then able to create innovation in the digital economy. While companies attempt to address a variety of workloads, heterogeneous environment, and resources, the hence complexity of resource management got even higher while cloud services were emerging. This cloud model covers the enormous cloud entities, considering the five most essential characteristics and the three service models and the four deployment models[3].

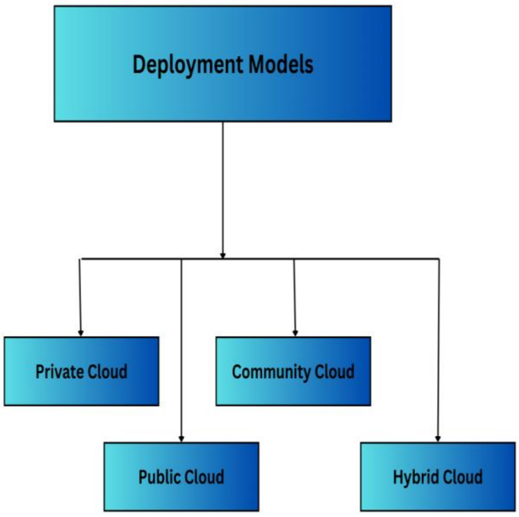


Fig 1.Deployment Models in cloud Computing[10]

B. *The Challenge of Resource Allocation*

It is like a real case to accomplish an optimum resource allocation in cloud being a variable entity. However, advanced calculating has been proved to be considered as an alternative to resource demand prediction, such as CPU and network bandwidth [11]. The unforeseen characteristics and diverse nature of cloud workloads are what most of the organizations experience as a challenge. The factors of fluctuations such as users' behaviour, seasonality and the time of the day also contribute to the organizational instability. Traditional resource allocation methods which were planning centric with static provisioning and over-provisioning will just not suffice for the modern cloud as it is way too resource hungry. These issues lead to the lost resources, rising costs, and less than optimal performance of the agency. In addition, the cloud management in hybrid and multi-cloud environment becomes more tricky because of their goal of complying the advantages of several cloud providers without putting companies outside the reach of vendor lock-in and guaranteeing interoperability.

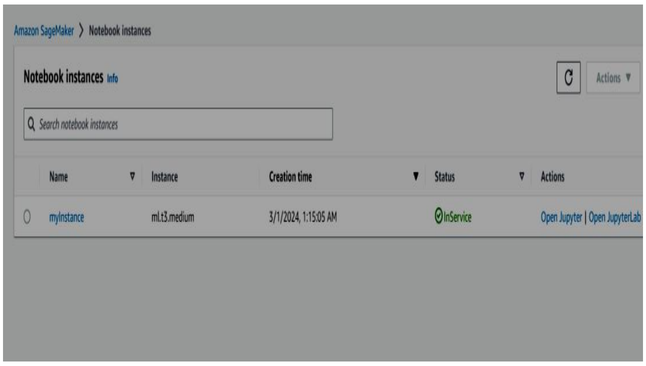
C. *Emerging Trends in Resource Management*

In a cloud model it becomes possible that a very large number of horizontal users can simultaneously interact with resource base in a remote location. [12] Resources of small cloud platforms are provided to users by Round-Robin (RR) [4], Best Fit (BF) and Min-Max [6] scheduling algorithms. They might be simple to use and understandable but they are limited in terms of consumption of resources. Such as for instance, large scale cloud platforms. Though cloud computing is challenged by sharing resources, future is being shaped by new developments aimed at optimal resource management, as well as solutions to the complexities. The appearance of server less computing is one of the trends that shifts the budget for developers from server procurement and management to the abstracting of the underneath infrastructure. Rather server less architectures provide organization with ability to become more agile, scale more easily, and have much less the operational workload, in the meantime they introduce new challenges like in cost control, program bug fixing, and maintenance for the whole system. The increase in orchestration and containerization technologies such as Docker and Kubernetes which allows enterprises to package and distribute apps in lightweight movable containers, apart from automating the management of containerized applications in an instance, is regarded as yet another trend.

Unlike conventional containers, containers feature a higher level of critical components like convenience, adaptability, scalability, and resource efficiency because they separate applications from the foundational infrastructure.

D. *Best Practices for Efficient Resource Allocation*

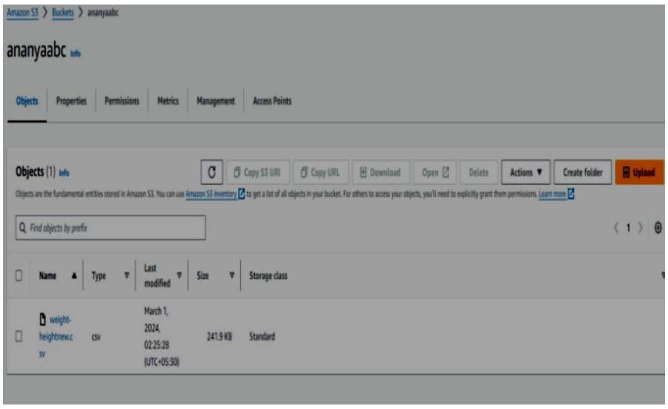
A variety of best practices and methods can be implemented by organizations to help them manage the challenges of resource allocation in a dynamic cloud environment. Using machine learning and predictive analytics to forecast workload demand and dynamically modify resource allocations in real time are a couple of these. Similar patterns of service tenants can be automatically derived based on past resource demands. It is possible to identify service tenants with high and low resource demands by using clustering to examine how similar the data is, and then use ML and DL regression techniques to provide predictions for the high resource demand tenants [7]. To optimise resource utilization, organizations can also adopt policies and automate workflows. Some examples of these include rightsizing instances, putting auto-scaling policies into place, and using spot instances or reserved instances to cut costs. Additionally, as essential elements of their entire cloud strategy, organizations should prioritize cloud cost management and optimization. To optimize return on investment, they should periodically assess consumption patterns, pinpoint areas for improvement, and put cost-cutting measures into place. Ineffective capacity planning greatly increases the possibility of resource inefficiency. Making sure you have the resources to deliver your future pipeline is the goal of capacity planning.


Amazon SageMaker 〉 Notebook instances
Notebook instances Info
Actions ▼
Search notebook instances

	Name	Instance	Creation time	Status	Actions
○	myInstance	ml.t3.medium	3/1/2024, 1:15:05 AM	InService	Open Jupyter | Open JupyterLab

Fig2.Creation of instance in Sage-maker

To guarantee safe access to AWS services, create an IAM user account. Designed and oversaw the creation of an S3 bucket to store project datasets effectively. developed and trained a linear Regression machine learning model using AWS Sage Maker, making use of the Sage Maker instance.



Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

□	Name	Type	Last modified	Size	Storage class
□	weight-heightnew.csv	csv	March 1, 2024, 02:25:28 (UTC+05:30)	241.9 KB	Standard

Fig3.Creation of S3 Bucket

The image above depicts what the configuration of a fully secure and scalable data storage solution on Amazon Web Services that can be utilized for the project datasets would look like. S3 is really a very reliable Storage service that covers virtually any volume of data from any place on the Internet.Thus, from the creation of the S3 bucket, our data set will save its repository in a centralized place, thus storing it safely, becoming accessible and durable. This system enables the storage of data for huge amounts of files, including text and multimedia, as well as managing access to the data that means that only selected people can view, upload, and edit them.Using sage-maker,created the linear regression model.

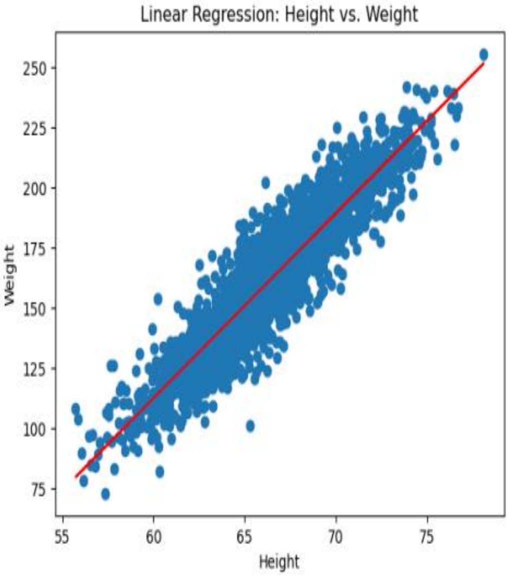


Fig 4.Linear regression:Height vs. Weight

E. Virtualization

In computer systems, virtualization refers to the process of generating a virtual version of anything, including but not limited to an operating system, storage devices, computer network, or hardware programme of a virtual computer[8]. Big IT behemoths with complex resource management include Google, Microsoft, and Amazon. These companies have enormous data centers. These enormous data centre resource management includes servers, virtual machines (VMs), and different management roles, as stated in [9].Care Resource Broker (CRB) is a technique that Thamarai et al. [13] proposed to increase the resources' throughput and reaction time. This method allocates the precise amount of resources to finish the task efficiently. The authors' results show that CRB is resource- and cost-efficient; however, they neglected to take energy into account when developing their technique.

II. LITERATURE REVIEW:

A. The paper titled "Deep Reinforcement Learning for Dynamic Resource Allocation" by Zhang ,Chen and Lei(2021),defines a new methodology for handling resources in cloud computing environments.It uses a methodology called reinforcement learning that inspires how animals learn.Conventionally,it is a complex task for allocating resources because demand Constantly Change.This new methodology aims to adjust resources automatically based on these changes making it more effective and dynamic.

B. The paper titled “Multi-objective Auto-scaling using Machine Learning” by Li, J., Sun, L., & Guo, S. (2020) defines the challenge of automatically scaling of resources in the cloud for performance and cost-efficiency.Conventionally ,scaling of the resources may involve a composition between these elements.

C. The paper titled “Cost and Availability Aware Resource Allocation and Virtual Function Placement for CDNaaS Provision ” by Louiza Yala; Pantelis A. Frangoudis; Giorgio Lucarelli; Adlen Ksentin defines challenges like allocation of resources,Placing virtual functions by considering cost and availability of services. This methodology helps to find stability between making CDN service cost effective and highly attainable for the consumers.

D. The paper titled “Block-chain for Safe and Effective Resource Allocation Article” by Xu, L., Chen, W., Sun, Y., & Tian, Y in 2023 defines how block-chain technology can improve resource allocation where you are using multiple cloud services.

E. The Paper “A Research Paper on Server-less Computing” by Vaishnavi Kulkarni(2022) defines server-less computing by providing back-end services which is based on the pay as you go model.

III. CATEGORIZATION OF RESOURCE ALLOCATION

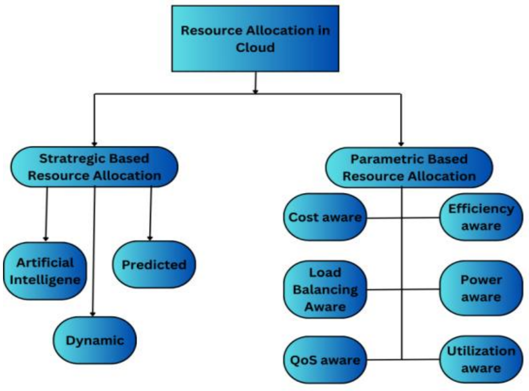


Fig 5.Categorization of resource allocation[14]

A. *Strategic based resource allocation*

Strategic resource allocation refers to the skill of organizing and controlling your cloud resources to get the most long-term benefits for your company. It's about using your applications purposefully to support your specific goals, rather than just providing them with whatever resources they seem to require.Strategic based resource allocation can be further divided into three parts:

1) *Artificial Intelligence Resource Allocation*

In cloud computing the AI based resource allocation techniques behave and function similarly to the human beings In view of artificial intelligence's influence on cloud computing, authors have begun to create AI-based resource allocation methods.

Allocation of cloud resources for AI workloads is necessary since these workloads need large amount of processing power and a specialized hardware.If these requirements are not properly managed then it may lead to higher expenses.Therefore ,in order to minimize expenses and to achieve performance goals,businesses must make effective use of their cloud resources[15].

2) *Dynamic resource allocation*

Dynamic resource allocation is an API that helps for requesting and use of resources like storage space or processing power between the pods and containers. Resource allocation and tracking are handled by third-party resource drivers.Arbitrary parameters are supported by different types of resources for initialization and requirement definition[16].

3) *Predicted resource allocation*

In cloud-based predicted resource allocation , resource demands are predicted by the use of advanced methodologies like machine learning.To improve cloud resource optimization,a number of methods are used including resource scaling based on predicted values and load prediction.These techniques ensure improved prediction quality and cost efficiency in cloud computing settings by adapting to various workload characteristics and variations in resource utilisation.

B. Parametric based resource allocation

In a computer system, making resources like the processing power or storage capacity available in relation to specific criteria is known as parametric -resource allocation. undefined

1) Cost-aware resource allocation

The method of effectively allocating resources in the cloud, and while considering the related cost is termed as cost-aware resource allocation. This strategy, engage in resource provisioning and cost considerations trade-off in order to manage resources in cloud networks as optimally as possible. This strategy balances the conflicting factors of resource allocation and cost constraints to ensure the management of resources in cloud networks is as efficient as it can be.

2) Efficiency aware resource allocation

In cloud computing, efficiency-aware resource allocation relates to distribution of resources in the most efficient way and as cheaply as possible while preserving efficiency. This plan exploits resource scheduling algorithms that are main distributors for jobs in the cloud setup. Efficient algorithm generation is possible by including objectives such as load balancing, power usage, reliability awareness, and security as part of the programming.

3) Load balancing aware resource allocation

Load balancing aware resource allocation refers to the proper distribution of loads among the respective computing resources to improve efficiency of performance, resource utilisation, etc.here are different suggestions to tackle this issue for the example of TARA (Topology Aware Resource Allocation) which uses Genetic Algorithms, RAS (Resource Allocation System) for cloud service, and the skewness algorithm for an eco-friendly computation usage[19].

4) Power aware resource allocation

Power Saving Resource Allocation is an extremely efficient way to provide energy-efficient services & also reduce excess power consumption in data centres by allocating resources to client applications in an intelligent way. This tactic is achieved by engineering and impelling the usage of resource saving allocation heuristics that assists to improve the energy efficiency of data centres by provisioning and balancing resources for applications smartly.

5) QoS aware resource allocation

In cloud computing, QoS-Aware Resource Allocation is a process of giving resources taking into consideration quality of service demands of applications. It consists of ensuring that applications are not under performing and living up to their performance requirements through smart resource management. This method seeks to make sure that the resources are optimise to give users, applications that run seamlessly and hits the mark regarding speed, reliability, and performance.

6) Utilization aware resource allocation

The cloud computing resource planning comprises managing resources in line with the performance of different applications or services so that efficient running of the services is guaranteed and cost-saving is profitable. Such an approach with studying the patterns of resources use can help to direct resources for meeting demand, rather than wasting them, and fully used them on one and the same time as well. Leveraging elasticity, gives the strategy an ability to adjust the resources on demand, which resulted in high level of overall efficiency and performance of cloud computing environments.

IV. STRATEGIES

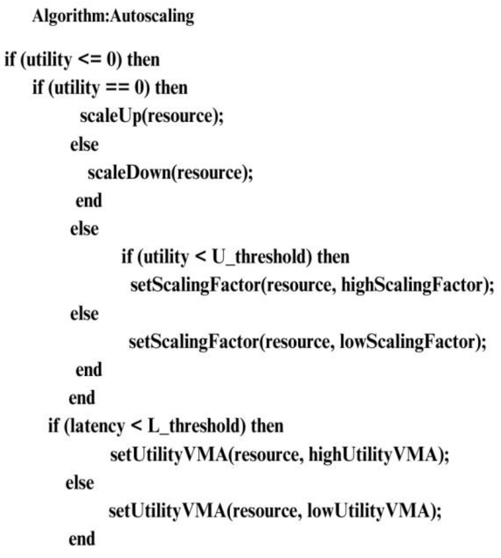
To achieve efficient resource allocation in a dynamic cloud environment, the following strategies can be employed:

A. Auto scaling

Auto scaling constitutes one of the methods of cloud computing that measures the necessary resources of a server farm according to the demand variation. This way it can put up or take down servers running at the back of behind a web app to serve the present load. This is an important factor to consider in the performance optimization, achievable performance, and cost-effective of cloud environment. Hence, organizations can efficiently utilize their resources through adding or sizing down according to their need and being able to supply the exact resources at the best time. Besides autoscaling the investment of human resources in the event of seasonal traffic fluctuations leads to consistent availability of the service and cost effectiveness.

In a traditional cloud data center, virtual machines are categorized based on their CPU utilization levels. Machines with CPU utilization ranging from 0% to 12% are considered free, those with 30% to 67% utilization are in a general status, and those exceeding 68% are classified as busy. Not fully utilizing servers can lead to lower costs, which is a critical issue in information technology[20]. To prevent unnecessary costs, it is recommended to take measures such as adjusting work schedules or triggering the auto-scaling mechanism when CPU utilization is below 20%.

The Algorithm seeks to provide the mechanism for the dynamic optimization of resource allocation relying on the current utility and latency of resources in an attempt to balance efficient resource utilization with performance requirements. Tweaks to thresholds and scaling factors can be done according to the exact specifications of the apps as well as the metrics of their performances.



```
    Algorithm:Autoscaling

if (utility <= 0) then
    if (utility == 0) then
          scaleUp(resource);
        else
          scaleDown(resource);
        end
        else
              if (utility < U_threshold) then
                setScalingFactor(resource, highScalingFactor);
        else
                setScalingFactor(resource, lowScalingFactor);
        end
      end
    if (latency < L_threshold) then
            setUtilityVMA(resource, highUtilityVMA);
      else
            setUtilityVMA(resource, lowUtilityVMA);
      end
```

Fig 6.Auto-scaling strategy[21]

B. *Load Balancing*

Load-balancer is an item of network infrastructure that is designed for balancing a distributed network traffic among multiple servers with the aim of optimizing application performance. It implies flow of the workloads so that they are well-balanced to reduce the latency and preventing specific servers as the overload.

A few researchers have introduced innovative ways like Genetic Algorithms (GA) and Topology Aware Resource Allocation (TARA) models for auto-scaling resources in the Infrastructure as a Service (IaaS) environment achieving the desired performance[22].

Consistently, the technique of Dynamic Resource Allocation for Load Balancing (DRALB) was considered to handle the imbalance of cloud data center networks so as to improve the effectiveness of scheduling and resource usage[23].

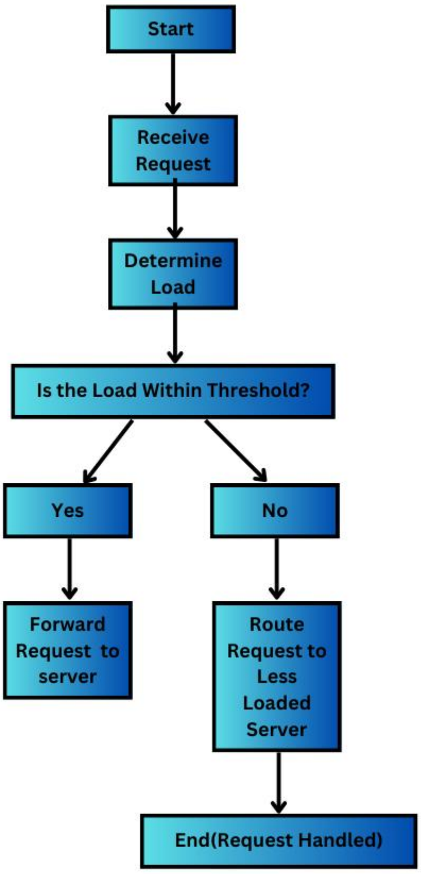

Start
Receive Request
Determine Load
Is the Load Within Threshold?
Yes
No
Forward Request to server
Route Request to Less Loaded Server
End(Request Handled)

Fig 7.Load Balancing

Firstly,client sends request .to load balancer After that the load balancer monitors the capacity of each server in the server pool and verdicts the server which has the smallest amount of work done currently.

Is the Load Within Threshold? Means A balancer of the load will ensure that the threshold of the load of every service is within an proper range.

Yes: once the specified server load is met, the router (load balancer) will redirect the requests to that particular server.

No: When a server is overloaded more than the threshold, Load Balancer cases the request to be directed to a less active server.

Forward Request to Server: Through a load balancer, the request will be forwarded to the target server.

End (Request Handled): At this point the process ends and the processing of the request is now in the hands of the server.

C. Scheduling:

The planning of tasks and resources allocation is a significant topic in cloud computing, currently, the system is using the number of available resources and every process will be running smoothly. Jobs scheduling in cloud environment means the assignment of jobs to suitable resources and looking among all these factors as computational time, financial concerns, and complexity of jobs.

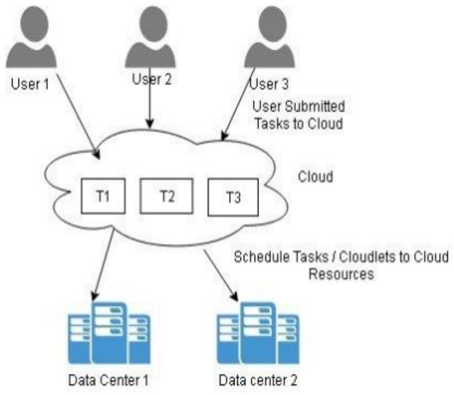

User 1
User 2
User 3
User Submitted Tasks to Cloud
Cloud
T1
T2
T3
Schedule Tasks / Cloudlets to Cloud Resources
Data Center 1
Data center 2

Fig 8.Scheduling in Cloud Computing[25]

In scheduling-based resource allocation, there can be two classification namely mean priority and round robin.

1) *Priority Scheduling*

Operating systems and cloud computing use priority scheduling, which assigns tasks a priority level and schedules their execution accordingly. Higher priority tasks are prioritized over lower priority tasks in this scheduling algorithm, guaranteeing that important tasks are finished on time and effectively. An ideal model for admission control and priority based service scheduling policy for cloud computing environments is put forth by Dakshayini et al. [24]. The goal of the proposed work is to efficiently provision cloud resources in order to maximize the amount of time that requests spend in the queue, meet quality of service requirements, and attain high cloud computing throughput.

2) Round Robin Scheduling:

In cloud computing, a popular algorithm called round-robin scheduling distributes CPU time among tasks in a circular fashion. Each task in a circular queue is given a time slice or quantum, and it can execute for a certain amount of time before moving on to the next one. By using this technique, resources are distributed equitably and the CPU isn't monopolized by one task. When several tasks must be completed simultaneously in cloud computing environments, round-robin scheduling is especially helpful as it promotes equity and effective resource use.

V. Future Scope

The vision of resources in future cloud computing system is going to deal with crucial problems and try untapped solutions for which would be helpful for effective resource management. Some potential areas for future research and development in this field include:Some potential areas for future research and development in this field include:

Dynamic Resource Allocation:
Sophisticated dynamic resource allocation mechanisms can improve resource use efficiency by the corresponding to what the workloads and demands dynamically demand, so they will be allocated in real time.

Machine Learning-Based Resource Management:
Applying machine learning algorithms for resource allocation can boost prediction capabilities and promote automation of routine work. These functions will help improve the efficiency of the system in general.

Energy-Efficient Resource Allocation:
Implementing strategies that lessen energy comprisement through the use of efficient allocation of assets could push companies' costs down, cause environmental benefits, and develop sustainable cloud operations.

Load Balancing:

Effective approaches of the load balancing mechanism distribution of workloads across resources without bottlenecks to increase efficiency of the system in successful response to the workloads

Virtual Machine Placement Optimization:

High quality Virtual Machine Placement (VMP) algorithm producing more effective mapping of virtual machines to physical machines can cause to resource utilization, as well as deprivation in the wastage.

Quality of Service (QoS) Optimization:

By way of appointing optimized Quality of Service measures via allocating appropriate system resources ensuring performance requirements, reliability and user satisfaction enhancement.

VI. CONCLUSION

Properly assigning resources of the cloud is necessary for the purpose of producing an output of improved performance, ensuring maximum resource utilization and guaranteeing the costs being lower. Such strategies as using dynamic resource allocation, energy efficient resource management, auto-scaling issues and continuous monitoring enlarge flexibility of systems, make scalable and trustworthy in cloud computing environments. Utilization of the optimized resource allocation algorithms together with the good scheduling approaches, and the application of (1) priority-based as well as (2) round-robin approaches will further increase system resource utilization and system efficiency. Basically, ensuring the sustainable management of available resources should be the first and foremost task of businesses to help achieve their cloud computing operations that are effective, highly cost-optimal and higher performance.

REFERENCES

[1] Chen, F., Xiang, T., Lei, X., & Chen, J. (2014). *Highly efficient linear regression outsourcing to a cloud.* IEEE transactions on cloud computing, 2(4), 499-508.

[2] Abid, A., Manzoor, M. F., Farooq, M. S., Farooq, U., & Hussain, M. (2020). *Challenges and Issues of Resource Allocation Techniques in Cloud Computing.* KSII Transactions on Internet & Information Systems, 14(7).

[3] Peter Mell, Timothy Grance, *"The NIST definition of Cloud Computing"*, 2011

[4] Pradhan P, Behera PK, Ray NNB (2016) *Modified round Robin algorithm for resource allocation in cloud computing.* Proc Comp Sci 85:878–890

[5] Srivastava S, Dubey R, Shrivastava M (2017) *Best fit based VM allocation for cloud resource allocation.* Int J Comp Appl 158(9):25–27

[6] Katyal M, Mishra A (2014) *Application of selective algorithm for effective resource provisioning in cloud computing environment.* Int J Cloud Computing 4(1):1–10

[7] Khan, T., Tian, W., Zhou, G., Ilager, S., Gong, M., & Buyya, R. (2022). *Machine learning (ML)-centric resource management in cloud computing: A review and future directions. Journal of Network and Computer Applications*, 204, 103405.

[8] Cervone, H. F. (2010). *An overview of virtual and cloud computing. OCLC Systems & Services: International digital library perspectives, 26(3), 162-165.*

[9] Ricardo Bianchini, Marcus Fontoura, Eli Cortez, Anand Bonde, Alexandre Muzio, Ana-Maria Constantin, Thomas Moscibroda, Gabriel Magalhaes, Girish Bablani, and Mark Russinovich. *Toward 39 ml-centric cloud platforms. Communications of the ACM, 63(2):50–59, 2020*

[10] Sharma, S., & Parihar, D. (2014). *A review on resource allocation in cloud computing. Int. J. Adv. Res. Ideas Innovat. Technol, 1, 1-7.*

[11] Rachael Shaw, Enda Howley, and Enda Barrett. *An energy efficient anti-correlated virtual machine placement algorithm using resource usage predictions. Simulation Modelling Practice and Theory,* 93:322– 342, 2019.

[12] Armbrust M, Fox A, Griffith R et al (2009) *Above the clouds: a Berkeley view of cloud computing. University of California, EECS Department, University of California, Berkeley. In: UCB/EECS-2009-28*

[13] T.S. Somasundaram, B.R. Amarnath, R. Kumar, P. Balakrishnan., K. Rajendar. R. Rajiv., G. Kannan.,G.R. Britto, E. Mahendran and B. Madusudhanan, *"CARE Resource Broker: A framework for scheduling and supporting virtual resource management," Future Generation Computer Systems, vol. 26, no. 3, pp. 337-347, 2010. Article (Cross Ref Link)*

[14] Madni, Hamid & Shafie, Abd Latiff & Yahaya, Coulibaly & Abdulhamid, Shafi'i. (2017). *Recent advancements in resource allocation techniques for cloud computing environment: A systematic review. Cluster Computing. 20. 1-45. 10.1007/s10586-016-0684-4.*

[15] https://datafloq.com/read/the-optimizing-cloud-resource-allocation-for-ai-workloads/

[16] https://kubernetes.io/docs/concepts/scheduling-eviction/dynamic-resource-allocation/

[17] Gureya, D. D. (2021). *Resource Allocation for Data-Intensive Services in the Cloud (PhD dissertation, KTH Royal Institute of Technology).* Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-304017

[18] Usman Sana, M., & Li, Z. (2021). *Efficiency aware scheduling techniques in cloud computing: a descriptive literature review.* PeerJ. Computer science, *7*, e509. https://doi.org/10.7717/peerj-cs.509

[19] Afzal, S., Kavitha, G. *Load balancing in cloud computing – A hierarchical taxonomical classification.* J Cloud Comp 8, 22 (2019). https://doi.org/10.1186/s13677-019-0146-7

[20] W. Vogels, "*Beyond Server Consolidation*", ACM Queue, Vol. 6, No. 1, pp. 20-26, 2008.

[21] Liao, W. H., Kuai, S. C., & Leau, Y. R. (2015, December). *Auto-scaling strategy for amazon web services in cloud computing.* In 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity) (pp. 1059-1064). IEEE.

[22] Ashawa, M., Douglas, O., Osamor, J. et al. RETRACTED ARTICLE: *Improving cloud efficiency through optimized resource allocation technique for load balancing using LSTM machine learning algorithm.* J Cloud Comp 11, 87 (2022).

[23] chhabra, S., & Singh, A. K. (2021). *Dynamic resource allocation method for load balance scheduling over cloud data center networks.* Journal of Web Engineering,20(8), 2269-2284.

[24] Dakshayini DM, Guruprasad DH (2011) *An optimal model for priority based service scheduling policy for cloud computing environment.* Int J Comput Appl 32(9):23–29

[25] Ahari, Vinay & Ramachandran, Venkatesan & Latha, D. Ponmary. (2019). *A Survey on Task Scheduling using Intelligent Water Drops Algorithm in Cloud Computing.* 39-45. 10.1109/ICOEI.2019.8862777.