Journal of Advances in Science and Technology
Vol.22, Issue No. 2 April-2025, ISSN 2230-9659

A review of cloud based applications and efficient resource utilization and allocation

Amita Boral[1]*, Dr. Kishan Kumar[2]

1 Research Scholar, Shri Krishna University, Chhatarpur, M.P.

ouriginal.sku@gmail.com

2 Professor, Shri Krishna University, Chhatarpur, M.P.

Abstract- Cloud-based applications have revolutionized computing by offering scalable, flexible, and cost-effective solutions across diverse industries. These applications rely on efficient resource utilization and allocation to ensure optimal performance, minimize costs, and enhance user satisfaction. This review explores the evolution, architecture, and benefits of cloud-based applications, emphasizing the challenges associated with resource management. Key strategies for improving resource allocation, including virtualization, load balancing, task scheduling, and dynamic resource provisioning, are analyzed. Additionally, emerging technologies such as artificial intelligence, machine learning, and edge computing are examined for their role in optimizing resource allocation. Through a comprehensive assessment of existing studies and methodologies, this review highlights the importance of innovative solutions for addressing inefficiencies, managing workloads, and reducing latency.

Keywords- Cloud Computing, Benefits Machine Learning, Resource Allocation, Virtualization

INTRODUCTION

The modern era of computing is characterized by the widespread use of cloud-based applications and services, providing businesses and individuals with unprecedented scalability, flexibility, and accessibility. The ever-increasing need for these cloud-based applications, however, has presented additional difficulties, such as the optimal allocation of cloud resources & enhancement of application performance. By optimizing cloud application performance & assuring optimal resource use, Machine Learning (ML) approaches have emerged as a potent tool for tackling these issues. Here, we will lay the groundwork for a more in-depth investigation of Machine Learning's contribution to the betterment of cloud-based applications & resource management by outlining the study's core goals, significance, and significant issues.

CLOUD COMPUTING

The term "cloud computing" refers to a model of computing that allows users to share resources (such as servers) and applications over the internet, as well as the underlying platform. The phrase "cloud" is a metaphor that has taken on an international connotation, indicating anything that extends across the whole planet. Cloud computing refers to both a platform & type of application. Servers, in the form of either existing physical computers or virtualized versions of such machines, can be provided, configured, & reconfigured by cloud platform services. Conversely, online applications and web services are hosted by huge data centers and dominating servers in what is known as "Cloud Computing," which explains how applications are expanded to be accessible through the internet.

To maximize cloud performance and resource use, it employs virtualization & load balancing techniques. In addition to this, it employs technology such as web services, distributed computing, networking, etc. on this context, the word "cloud" refers to a globally distributed, configuration-free server on the cloud. People and major corporations alike can take advantage of software and hardware services provided by remote third parties through the cloud paradigm. This includes things like online file storage, social media, email, & business applications.

As a means of negotiating between cloud service providers and their clients, SLAs (Service Level Agreements) include QoS factors like time limitations and ensure that performance statistics are adhered to. By utilizing cloud computing, SMEs can free themselves from the burden of managing, securing, configuring, and selling their IT infrastructure.

APPLICATIONS OF CLOUD COMPUTING

There is a daily growth in the number of companies engaged in cloud-based service trading. The following are some of the many fields in which cloud computing can provide its services:

- **Testing and Development:** The cloud paradigm generates the availability of tools for application building. In order to make application development, testing, and deployment accessible to non-developers, it provides tools that are easy to configure, thus removing the complexity. Online development frameworks reduce the need to spend time and money on resource acquisition.

- **Backup and Disaster Recovery:** In the event of a breakdown, data and services stored in the cloud are replicated for recovery purposes.

- **Large data storage:** Web-enabled interfaces make it easier to store large data sets in cloud datacenters. Skilled teams handle storage management, security, privacy, and consolidation tasks via the cloud, relieving people and companies of the burden of worry about large investments and maintenance needs.

- **Software Library:** The cloud allows organizations to easily get software as needed, whenever they need it, without getting caught up in the legal complexities of pirated versions.

- **E-Commerce:** The advent of cloud computing in the worldwide market has recently boosted the trading of commodities over the internet. Season, festival, and holiday fluctuations affect the e-commerce industry's cloud traffic. The cloud is able to automatically scale up during peak demand and down during off-peak periods because of its proficiency in resource provisioning.

- **Website Hosting:** Developers' focus shifts from development to management when they engage in hosting activities. By delegating maintenance duties to teams who specialize in providing services, developers may rest easy while their apps are hosted on the cloud.

- **Databases in the Cloud:** Developers now have access to complicated and expert-level solutions for tuning their databases on the cloud. By moving their services to the cloud, service providers are increasing the rate at which they generate income. Rackspace is one company that has used this approach.

- **Electronic mail:** E-mail's move to the cloud has increased storage scalability, made client data more secure, and protected privacy. Also, users no longer have to worry about server outages.

SERVICES OF CLOUD COMPUTING

Cloud computing allows for the simultaneous provision of numerous services to a huge user base. The SPI model, which is a component of cloud computing, encapsulates the various services that the cloud provides. As shown in figure 1, the cloud is viewed from the standpoint of the services it offers and the deployment model it supports.

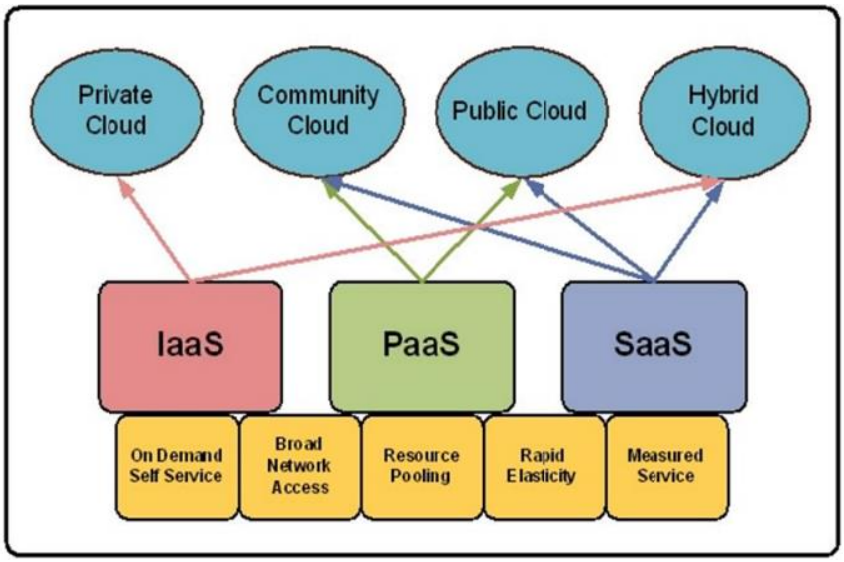

Figure 1: Cloud Computing perspective

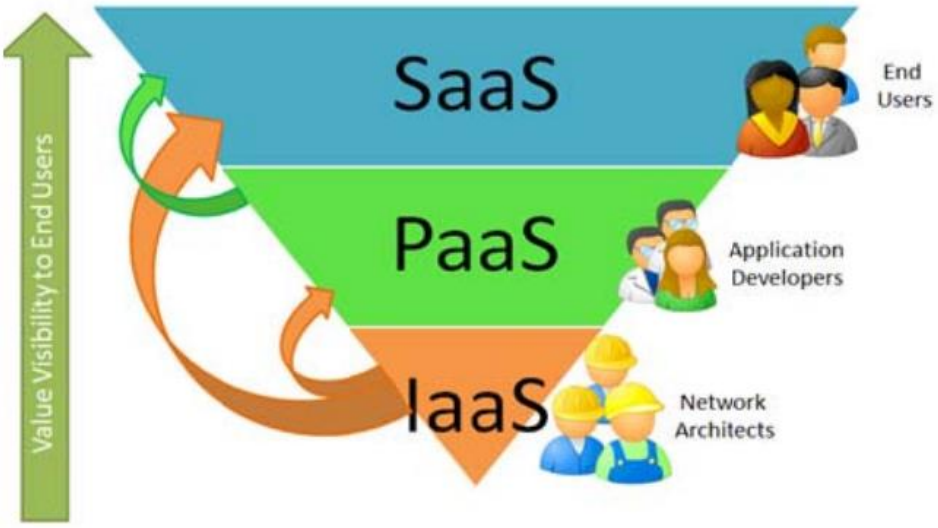

Figure 2: Cloud Computing service models

SaaS, PaaS, and IaaS are the three main subsets of cloud services, as seen in Figure 2. Following is a brief overview of the three models:-

Software as a Service (SaaS): Cloud computing eliminates the need for individuals & businesses to purchase, maintain, and upgrade software by providing on-demand software services through subscription. The provider can control who has access to the SaaS app & how often it needs maintenance because it is hosted and operated on their servers. The program is typically rented out on a per-use basis or purchased as a subscription, with each user receiving their own license. Users in this paradigm are not concerned with the underlying infrastructure or platform; rather, they merely need to access the service as a web application. The SaaS

family of services is enriched by applications like online games, social media sites, and office software.

Platform as a Service (PaaS): Platform as a service, or PaaS, is a comprehensive set of tools that developers may use to build and launch their apps in the cloud, whether it's a public or private one. By doing away with the complexity of managing individual software & hardware components, it allows an organization to take use of critical middleware services. PaaS examples include Engine Yard, Google App Engine, Windows Azure, and Force.com.

Infrastructure as a Service (IaaS): By using the IaaS model, a company can contract out the management of its storage, hardware resources, servers, & networking components, among other things, to third parties. It is the service provider's responsibility to house and maintain the equipment, as he owns it. The customer makes a one-time payment. A safe, standardized, and extensible foundation is the bedrock upon which infrastructure services are constructed. High availability & elasticity of resources necessitate virtualization and some degree of infrastructure redundancy. The foundation of operating services in virtualized environments is server virtualization, most commonly via VMware or XEN. Software automation should be used to easily provide and de-provision these services.

CLOUD COMPUTING ADVANTAGES

Despite the fact that cloud computing resources are distributed across global boundaries, the paradigm nonetheless provides its entities with a wide range of benefits. The following are some of the advantages provided:-

- **Cost Reduction:** Prominent corporations & people have adopted the cloud paradigm due to the cloud's ability to consolidate infrastructure. The user no longer has to worry about making a financial commitment to acquire hardware resources, software licenses, and the necessary infrastructure to integrate them. In addition to avoiding the first setback that occurred when work was about to begin as a result of infrastructure setup procedures, this method has gone a step further.

- **Software Upgrades:** As a result of more varied computing demands, service providers with more advanced capabilities are moving their operations to the cloud, relieving users of the responsibility of keeping their software up-to-date.

- **Application selection flexibility:** The cloud enables users to build and utilize bespoke services from a variety of vendors, each with their own unique set of skills that can handle complex request compositions.

- **Resilient Computing:** Services & data are mirrored over the cloud and made accessible globally to lessen the effect of calamities.

- **Service Focused on Usability:** Metered capabilities take advantage of paying for a service only when it is used. Consequently, the strategy aims to free the user from the responsibility of upkeep. In addition, it prevents spending that may have gone toward equipment purchases.

TYPES OF MACHINE LEARNING

The field of artificial intelligence (AI) recognized as ML is concerned with teaching computers to learn from and make judgments based on data on their own, without being specifically taught to do so. To get better at a task, machine learning algorithms look for patterns in data and draw conclusions about the task based on those patterns.

Different machine learning technique include

- **Supervised Machine Learning-** Supervised ML algorithms are utilized to make predictions about the future by learning from data sets that have been labeled & mapped to certain target values. The primary responsibility of a supervised ML algorithm is to analyze the input data & assign a proper classification to it. Only through training on a large, well labeled dataset with well-defined classes would such a distribution be possible. The two types of ML problems that a supervised ML algorithm is able to address are classification & regression. Classification is a solution method used when the underlying problem has a binary (yes/no) target variable. On the other hand, Regression ML methods are used to handle issues where the target variable is not categorized but continuous.

- **Unsupervised Machine Learning-** The ML model is trained by unsupervised ML techniques using datasets that are neither labeled nor categorized. Unsupervised ML algorithms examine a huge dataset to discover and learn data insights such as patterns, classifications, and categories without human supervision. Two types of unsupervised ML are Clustering and Association. Algorithms based on clustering divide data into

groups based on their shared features. Association-based algorithms, on the other hand, seek out connections between pieces of information that naturally belong together.

- **Semi-Supervised Machine Learning-** It is designed to compensate for the shortcomings of both supervised and unsupervised ML methods. In semi supervised learning, the ML model is trained using both labeled and unlabeled datasets.

- **Reinforcement Machine Learning-** Reinforcement ML makes use of a learning paradigm centered on the use of feedback. The agent is not trained on any supervised datasets and instead is given positive or negative reinforcement for making the right or wrong choices. In Figure 3 we see how various machine learning methods can be categorized.

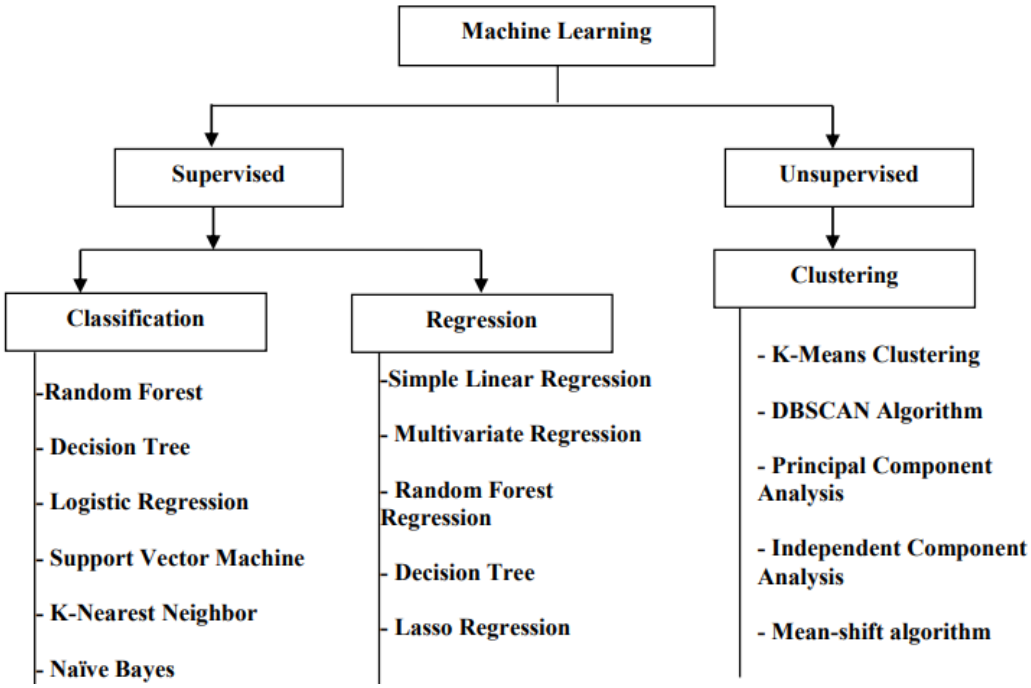

Figure 3: Algorithm ML classification

QUALITY OF SERVICE (QOS) IN A CLOUD COMPUTING

QoS is used to define the procedures and tools used in cloud computing to guarantee that data & programs will function as expected under all conditions. Quality of service, or QoS, is a crucial component of cloud computing because it provides a mechanism for controlling and consistently delivering the standard of service expected by cloud customers. Key Qualitative Service in the Cloud Features:

- **Service Level Agreements (SLAs):** SLAs are the starting point for many QoS arrangements between cloud providers and their clients. SLAs define the expected levels of service, including factors like uptime, response times, and data availability. They set the baseline for QoS.

- **Resource Allocation:** QoS involves allocating resources within the cloud infrastructure to ensure that each service or application gets the necessary computational, storage, and network resources. This includes dynamic resource allocation to handle varying workloads.

- **Performance Monitoring:** Cloud providers use monitoring tools and techniques to track the performance of services and applications. These tools can detect issues such as network congestion, server load, or latency.

- **Load Balancing:** Load balancing is a QoS mechanism that evenly distributes network traffic and requests across multiple servers. This ensures that no single server is overwhelmed and that all users experience consistent performance.

- **Security & Privacy:** QoS in the cloud extends to security & privacy measures. Protecting data and ensuring secure access to services is essential for maintaining a high level of service quality.

- **Fault Tolerance:** QoS mechanisms should include fault tolerance and redundancy. Redundant systems and data backups help ensure that services remain available even if individual components fail.

- **Scalability:** Cloud services should be able to scale up or down as demand changes. QoS measures must ensure that this can happen seamlessly, without impacting service quality.

- **Network Quality:** Network QoS involves managing bandwidth and minimizing latency. Techniques like Quality of Service (QoS) settings and Content Delivery Networks (CDNs) are used to optimize network performance.

- **Prioritization:** Some applications or users may require higher QoS levels than others. Prioritization mechanisms help ensure that critical services receive the resources they need.

- **Dynamic Adaptation:** Cloud systems should adapt to changing conditions. For instance, if a service experiences high demand, it should be able to allocate more resources on the fly to maintain performance.

QoS in cloud computing is essential for businesses and organizations that rely on the cloud for their operations. It ensures that cloud services meet their specific needs and expectations, and it provides a framework for monitoring and continuously improving service quality. Effective QoS measures are a critical part of cloud service management and are essential for building trust and reliability in cloud computing environments.

VIRTUAL MACHINE

In the same way that a physical machine (PM) runs programs, VM does the same. There are two main types of VMs, distinguished by how they are used and how closely they resemble physical machines. Figure.4 shows a virtual machine structure, which is detailed below.

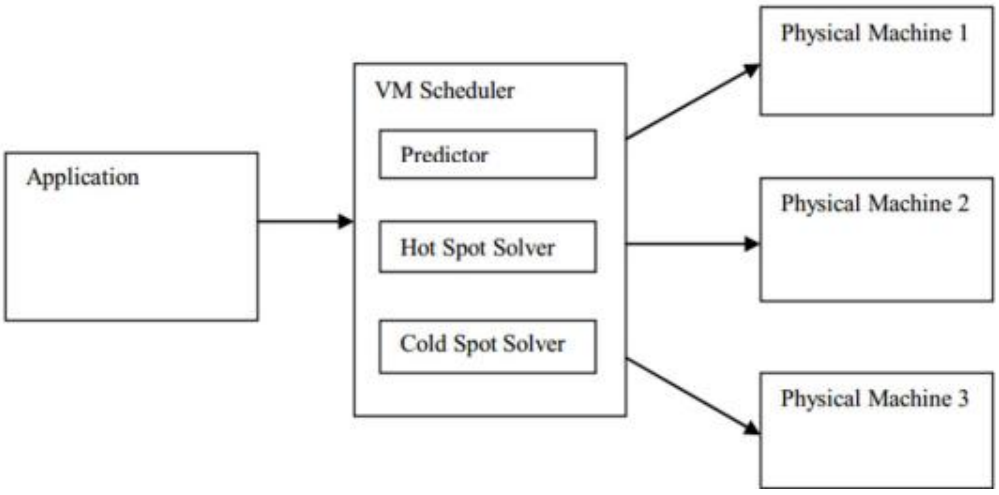

Figure 4: Structure of VM

VM operates as a standalone process because its sole purpose is to run a single application. Such VMs offer system flexibility & ease of use while being well-suited to one or more programming languages.

Virtual Machine Allocation

In a datacenter, VMs are in high demand from cloud users for specific resources like storage, processing power, and memory. Upon receiving this request from the client, the DC will assign the virtual machine to a server and ensure that all critical resources are available on that server.

Depending on the server's capacity & VMs' resource needs, a server can support a large number of VMs.

Since energy consumption has become a major expense and environmental concern for server farms, it is critical to allocate VMs to servers in a way that minimizes energy consumption. The CPU use of the preset assets determines the position of a virtual machine when it is chosen to relocate.

Virtual Machine Migration

Virtual Machine Migration, or VMM, refers to the process of transferring virtual machines from one system to another. Tolerating PM errors, balancing workloads, and reducing DC power usage are all made easier with the VMM. It takes time for every task in the cloud to migrate among virtual machines. The service delivery is not affected during the dynamic migration of VMs in cloud data centers in order to fulfill customer demands. New cloud events, such as workload balancing, online maintenance, server consolidation, and so on, can be supported by the migration mechanism. With VMM, cloud resource allocation is lightning quick, and operational costs are dropping. In addition, the structure of excessive energy consumption is not just the quantity of recording assets & power inefficiency of equipment; it is the direct outcome of inefficient usage of these assets.

Virtual Machine Consolidation

Consolidating virtual machines (VMs) is an important part of creating a dynamic cloud resource management system that uses less energy. By assisting with the migration of virtual machines into physical servers, the VMC allows for more efficient use of cloud resources with less power consumption. However, bad QoS could result from processing multiple VMs onto a single server. To tackle this, VMC algorithms are designed to progressively evaluate the impact on QoS when choosing which virtual machines to transfer. As shown in Figure.5, VMC moves the virtual machines with fewer physical machines (VMs) than previously. At the same time, PMs without VMs can be switched from an active (or "on") state to a less energy-intensive (or "rest") state, like a sleep state, to minimise power consumption.

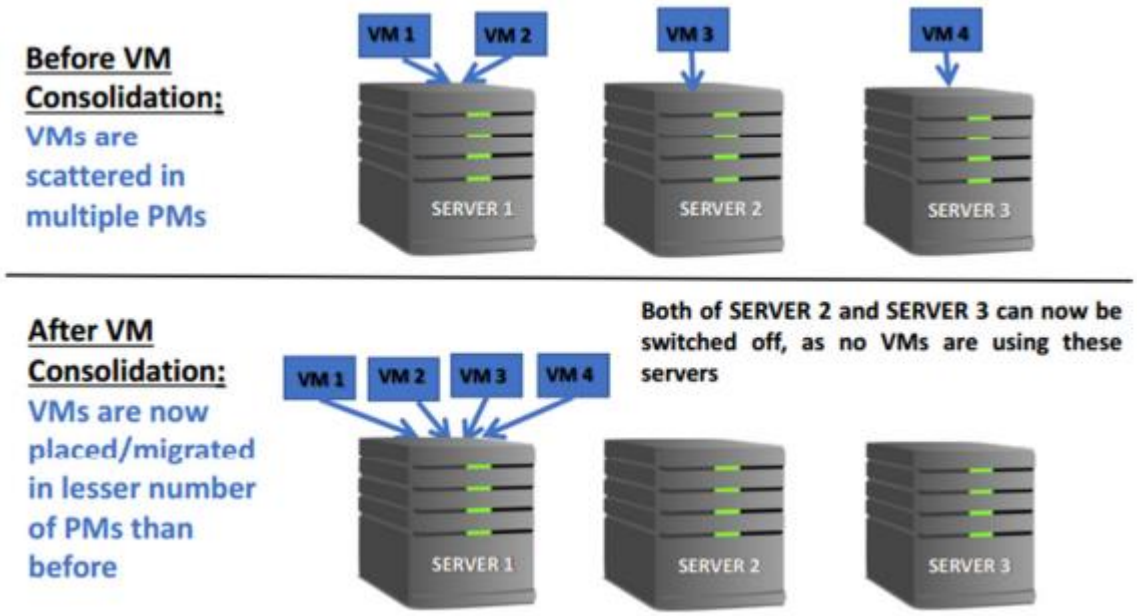

Figure 5: Architecture of VMC

DATACENTERS

Cloud applications are made possible by engineers who balance resources, and their virtual framework is expanded or contracted as needed by customers. Additionally, the coders use this for their personal use. Not only that, they can also put their resources away. Consequently, a secure & robust infrastructure is necessary for a set of cloud computing services. Virtualization & minimal framework are used to process the requests. A DC is necessary for the storage of these requests. At least one organization can profit from these DCs, which are frameworks constructed out of IT segments that prepare storage, limit, and system benefits. Depending on the primary objective of the DC, the current extent of the DC (in terms of number of segments) might range from numerous segments to innumerable pieces. Additionally, DCs use a variety of information technology components, such as switches, stack balancers, storage devices, dedicated capacity systems, and the main portion of any server or DC. Edge computing is a reasonable reaction to cloud computing from a server perspective, since the company has standardized its approach to reduce power consumption, cooling capacity, & physical home usage. While the remaining servers' thin edges are running, the sharp edges are moved to a data center. This problem arises because state-of-the-art servers are measured using specialized computing equipment, and their design and organization typically require honing.

When it comes to process, storage, and system administration constraints in cloud computing,

DC work offers the sought-after capacity to reply to engineer requirements. The connected math multiplexing of engineers' applications allows for higher usage of the hardware's energy in a huge DC operating a virtualization determination.

RESOURCE ALLOCATION

The four main components of resource management are reporting, booking, allocating, and checking. In order to meet the demands of clients, resource revelation identifies the best physical resources to create virtual machines on. Out of all the coordinated physical resources, resource planning selects the best one. In order to set up resources from a cloud foundation, it actually identifies the physical resource that will house the VM [Sriram Kailasam 2013]. The process of assigning resources eliminates the need to select a specific resource for each task or assignment. Actually, it means adjusting the work schedule to fit the selected cloud resource. Once the accommodation of the occupation is complete, the resource is examined.

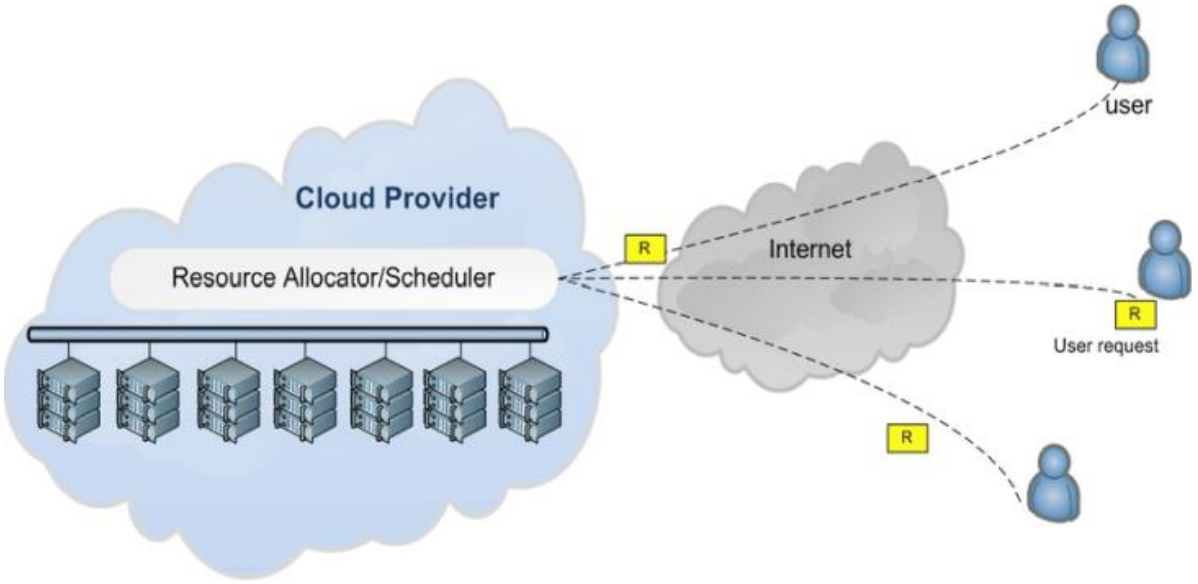

Figure 6: Resource allocations

Cloud-Based Energy Efficient Resource Allocation

In order to meet user requirements while improving cost efficiency & reducing energy usage, etc., resource allocation is primarily responsible for identifying & allocating resources to each incoming user request. Figure.7 shows that schedulers have the option to either ensure the static & underlying asset assignment at request arrival or to distribute both static & dynamic assets in order to constantly oversee assets, optimize, and check previous requests.

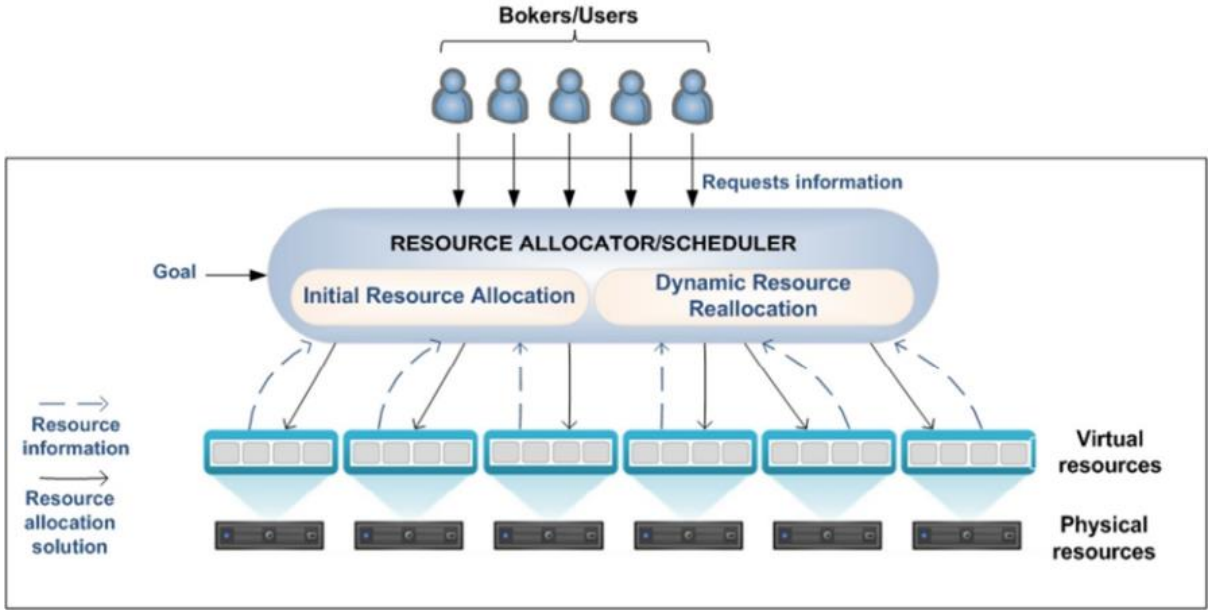

Figure 7: Resource Allocation in Cloud Computing

A wide range of group sizes, from relatively small DCs to massive ones, have resulted from the widespread use of cloud computing & virtualization technologies. The power consumption, DC ownership costs, and carbon footprint are all increased by these DC enhancements. Energy efficiency is thus becoming more and more critical for DCs & Clouds. One of the most difficult problems to solve is how to allocate resources in the cloud in a way that maximizes energy efficiency.

Resource Allocation Technique

A number of scheduling rules, such as the Global scheduling policy, make use of the numerous aspects of the device in order to allocate the work to the multiprocessor. Additionally, these policies manage the performance of the system altogether. The following is a description & listing of a crucial method of resource allocation:

- Static Scheduling Algorithm
- Dynamic Scheduling Algorithm
- Heuristic Scheduling Algorithms
- Opportunistic Load Balancing
- Min-Min technique

- Max-Min technique

CONCLUSION

Efficient resource utilization and allocation are critical for the sustained success of cloud-based applications. This review demonstrates that effective resource management not only improves system performance but also reduces operational costs and ensures reliability. Techniques such as virtualization, task scheduling, and load balancing are foundational, while the integration of AI and edge computing offers promising advancements. However, challenges such as energy efficiency, workload unpredictability, and security remain key concerns. Future research should focus on developing adaptive and intelligent resource management frameworks that leverage real-time analytics and predictive modeling. By addressing these challenges, the cloud computing industry can achieve greater efficiency and scalability, meeting the growing demands of users while fostering innovation and sustainability.

References

1. Al-Tous H. and Barhumi I. (2016) 'Resource allocation for multiple-sources singlerelay cooperative communication OFDMA systems', IEEE Transactions on Mobile Computing, Vo15, No. 4, pp. 964-981.

2. B. Lawson, E. Smirni, Power-aware resource allocation in high-end systems via online simulation, in Proceedings of the19th Annual International Conference on Supercomputing, Cambridge, USA 2005.

3. Chen S. Wu J. and Lu Z. (2012) 'A cloud computing resource scheduling policy based on genetic algorithm with multiple fitness', In Computer and Information Technology (CIT), 2012 IEEE 12th International Conference on, pp. 177-184, IEEE, 2012

4. Dabbagh, M, Hamdaoui, B, Guizani, M & Rayes, A 2015, 'Toward energy-efficient cloud computing: Prediction, consolidation, and overcommitment', IEEE- Network, vol. 29, no. 2, pp. 56-61.

5. Hitesh A. Ravani, Hitesh A. Bheda, Vrunda J. Patel, "Genetic Algorithm Based Resource Scheduling Technique in Cloud Computing", International Journal of Advanced Research in Computer Science and Management Studies, Volume 1, Issue 7, pp. 168-174, 2013.

6. J. E. Haddad, M. Manouvrier, G. Ramirez, and M. Rukoz, QoS-driven selection of web services for transactional composition, InProc. 6th Int'l Conf. Web Services (ICWS'08), 2008, 653–660

7. L. Shao, J. Zhang, Y. Wei, J. Zhao, B. Xie, and H. Mei. Personalized QoS prediction for web services via collaborative filtering, InProc. 5th Int'l Conf. Web Services (ICWS'07), 2007, 439–446.

8. Kapil Bakshi, Cisco Cloud Computing - Data Center Strategy, Architecture, and Solutions, Point of View White Paper for U.S. Public Sector 1st Edition Cisco Systems, Inc., 2009, 1-16

9. Ma Y.B. Jang S.H. and Lee J.S. (2011) 'Ontology-based resource management for cloud computing', In Asian Conference on Intelligent Information and Database Systems, pp. 343-352, Springer, Berlin, Heidelberg, 2011.

10. Sagar M.S. Singh B. and Ahmad W. (2013) 'Study on cloud computing resource allocation strategies', International Journal of Advance Research and Innovation, Vol. 1, No. 3, pp. 107-114.

11. Reddy C. and Suchithra R. (2016) 'Virtual Machine Migration in Cloud Data Centers for Resource management', International Journal of Engineering and Computer Science, Vol. 5, no 09, pp.18029-18034

12. Vignesh V. Sendhil Kumar K.S. and Jaisankar N. (2013) 'Resource management and scheduling in cloud environment', International journal of scientific and research publications Vol. 3, No. 6, pp. 1

13. Wang S.C. Yan K.Q. Liao W.P. and Wang S.S (2010) 'Towards a load balancing in a three-level cloud computing network', In Computer Science and information technology (ICCSIT), 2010 3rd IEEE International Conference on, Vol. 1, pp. 108-113, IEEE, 2010.

14. Xie, R.; Jia, X.; Yang, K.; Zhang, B. Energy saving virtual machine allocation in cloud computing. Distributed Computing Systems Workshops (ICDCSW), 2013 IEEE 33rd International Conference on: IEEE; 2013. p. 132-137.

15. Yuan D. Yang Y. Liu X. and Chen J. (2010) 'A data placement strategy in scientific cloud workflows', Future Generation Computer Systems Vol. 26, No. 8, pp. 1200-1214.

16. Z. Zheng and M. R. Lyu, A distributed replication strategy evaluation and selection framework for fault tolerant web services, InProc. 6th Int'l Conf. Web Services (ICWS'08), 2008, 145–152.