ISSN 1533-9211
DOI: 10.5281/zenodo.8310627

EFFICIENT RESOURCE UTILIZATION APPROACH IN CLOUD COMPUTING USING OPTIMIZED DIRECT RESOURCE PROVISIONING

THUMMULURU KAVITHA

Research Scholar, Department of Computer Science, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India. Email: kakikavitha83@gmail.com

Dr. THATIMAKULA SUDHA

Professor, Computer Science & Research Supervisor, Department of Computer Science, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India. Email: thatimakula_sudha@yahoo.com

Abstract

In the cloud, users can access large pools of distributed computing resources and storage resources on-demand and anywhere there is an internet connection. In order to manage these many assets efficiently, scheduling methods need to be time-saving. In addition to optimizing efficiency and fairness, the perfect scheduler will improve quality of service (QoS). Large-scale distributed systems require scheduling methods. Now the cloud customers are charged based upon the amount of resources they are consumed or held in reserve. Usage of data in cloud provisioning for different user should be cost effective. Customized usage of storage data with application and allocation to the user is a challenging task. Present work focused on the optimal allocation on storage-based application environment with low memory utilization. An optimized cost-effective based resource provision in cloud (OCRPC) algorithm has been proposed in the present work to utilize the cloud space usage at minimal processing. The technique tested under VM (virtual machine) with different user inter phase and compare with existing techniques. The proposed algorithm given 15-20% better results when compared with existing HMA, DIA and DRP.

Keywords: Cloud, OCRPC, VM, QOS, Cost, Resource Utilization.

1. INTRODUCTION

Cloud computing refers to the internet-based, on-demand delivery of information technology resources (IaaS) on a pay-as-you-go basis. It is a comprehensive ecosystem where users can construct complex, scalable software. Profiting effectively from it requires studying several different areas. Overprovisioning might cost consumers additional expenses, but under provisioning can slow down their applications. [1] Encourage enterprises to transition from stockpiling to the necessary infrastructure. This means that companies can use IT services without breaking the bank by reducing IT infrastructure capital expenditures. As more and more people turn to the cloud as a means of accessing data, concerns about the platform's impact on the environment have arisen. Cloud service providers may boost customer retention and cut down on wait times by managing efficiency in line with QoS client expectations. The cloud environment is becoming more complicated as the number of major apps rises. [2]. Quality of Service (QoS)-based resource allocation optimizes performance by matching demand with supply. Minimum performance time should be considered when allocating and managing resources in the cloud, which is easier said than done. Due to resource demands from

The Seybold REPORT
ISSN 1533-9211
DOI: 10.5281/zenodo.8310627

the cloud and the outcomes of short-term thinking. Therefore, providers must guarantee efficient cloud service delivery in accordance with the requirement of the QoS application, with a minimum of resources and in a short amount of time. Keeping a constant load on cloud services has become more difficult since user demand has fluctuated over time due to heterogeneity. Our literature assessment revealed that previous approaches to resource allocation problems did not adequately deal with cloud uncertainty, power, unpredictability, and complexity. [3]. An efficient resource allocation procedure predicated on QoS is essential for boosting customer retention and revenue. Allocation tactics can have an impact on operating expenses, usage, and energy expenditure. Effective overage is achieved through the use of a number of cloud resource management strategies. Therefore, expected performance and quality of service-based results are established through strategic planning and resource matching. With the Service Level Agreement (SLA) between the service provider and the customer in place, cloud computing is shown as one of the integrated computer resources. This study optimizes the allocation of resources to virtual machines (VMs) running on physical machines by focusing on vertical resource scaling and presenting a method to aid the Triple-N Resource Allocation Framework for vm's and cloud. This infrastructure can handle N-customers, N-agents, and N-virtual machines. The suggested method optimizes a utility function during runtime, which expresses the competing goals of improving application performance and reducing power usage.

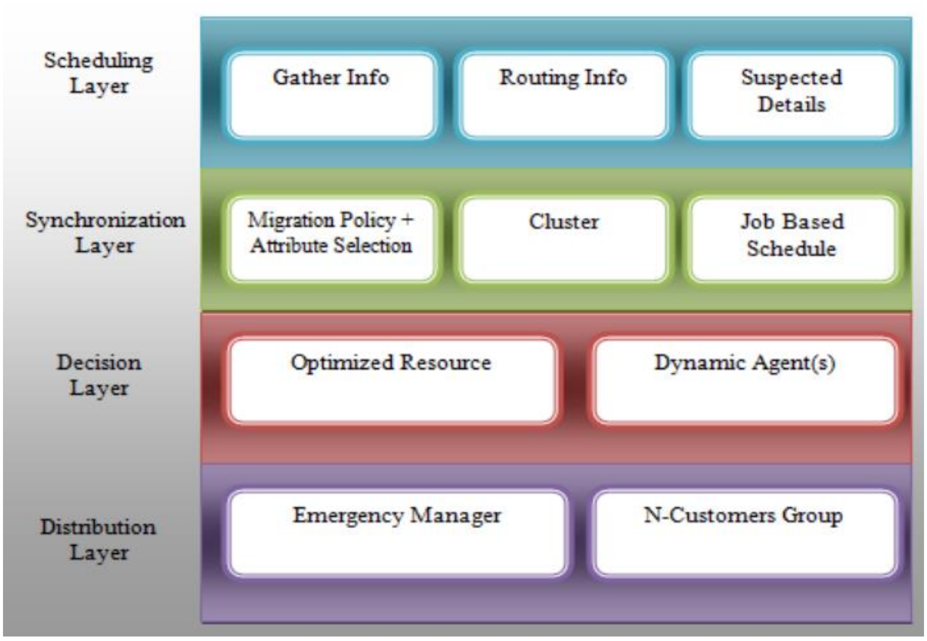


Figure 1: Architecture Diagram of Triple-N Resource Allocation Framework

The suggested framework takes into account the process of managing resources in a virtual machine (VM) or cloud setting at scale. A cloud-ready environment is one that has both the hardware and software controls necessary to set up and operate cloud-based services. These considerations are made from the vantage point of a cloud computing service.

The Seybold REPORT
ISSN 1533-9211
DOI: 10.5281/zenodo.8310627

1.1 Cloud properties that affect scheduling

Homogeneity: A homogeneous cloud is a single company that develops and maintains all its software, from the hypervisor to the intermediate cloud stack to the user portal. Having everything come from the same source makes the management process easy. Since everything is pre-integrated, if there is an issue only one entity is responsible. Users become reliant on one CSP's technical and marketing approach when that CSP has monopoly strength. This type of cloud environment allows consumers to focus solely on services provided by a single CSP. There are benefits and drawbacks to this technique, despite the ease with which administrators can cover for each other. However, the functions are proprietary to the individual providers, so they cannot be accessed technically.

Heterogeneity: The addition of various computing resources with improved memory and storage capacity is being made by CSPs to boost performance and attract more clients. As a result, heterogeneity boosts the efficiency and effectiveness of cloud computing as a whole. High-end, low-priced infrastructure, such fast CPUs, is a common request from users. Current efforts to develop environmentally friendly computer practices center on reducing power usage. Therefore, public CSPs are presently using a variety of architectural solutions for their infrastructure in an effort to enhance power efficiency.

Elasticity: When it comes to cloud computing, elasticity is the degree to which a system can respond to fluctuating workloads by automatically adding and removing resources as needed. This ensures that supply and demand are always balanced. A more adaptable and scalable cloud computing setting is made possible by elastic cloud infrastructure. Amazon's Web Services (AWS) makes it easier to scale up and down web services.

Scalability and auto scaling: When a cloud ecosystem is scalable, it may take on more work without requiring a complete overhaul by either upgrading the existing hardware (scale-up) or adding new nodes (scale-out). To ensure that the system can handle a growing workload, scalability measures are taken in advance, such as the addition of resources. As a result, a CSP may provide services that fulfill the long-term, strategic needs of its clients and live up to the quality standards expected by those customers. Auto-scaling helps reduce user frustration with slow job completion due to shared resources.

Scheduling models in cloud: A scheduling model's goal is to improve service quality through more efficient use of available resources. This study focuses on the approaches that are classified by how resource provisioning, job scheduling, virtual machine placement, and load balancing are carried out, but other categories, such as static and dynamic, are available. Our method can be broadly categorized as shown in Figure.

ISSN 1533-9211
DOI: 10.5281/zenodo.8310627

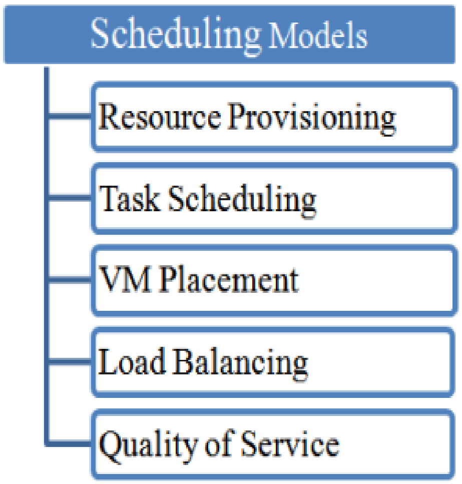


Figure 2: General Classification of Scheduling Model

1.1 Problem statement

Techniques used for intrusion detection are based on the rules that are generated based on which the true and false alarm ratio is predicted. However, the techniques implemented so far are inefficient in terms of detection rate and alarm generation rate. This is because the rules being used to generate the alarms are too broad and often lead to false positives. Additionally, the rules may not capture all types of malicious activity, leading to false negatives. Improving the detection rate and alarm generation rate requires more sophisticated rules and a greater focus on the nuances of malicious activity. If a server never gets the chance to process a virtual machine (VM) request, the cloud service provider loses money. Otherwise, the QoS promises to the users could be broken if too many VM requests are executed on a single node. This will also cut into the revenue of cloud services. We can avoid the aforementioned problem by distributing profits equitably among all computer nodes.

2. LITERATURE REVIEW

The objective of resource management is to reduce the amount of time and money needed to deliver a service. In this section, we detail the various methods available for dealing with available resources. The ideal choice of approach can be affected by a wide range of criteria, including service cost, resource availability, nature of task, and needed number of processors. Other well-known techniques include bin packing and gradient search. Abdul Salam Shah et al **[4]**. To strike a balance between user comfort and energy requirements, so that the user can achieve the desired comfort level with the minimum amount of energy consumption, the requirement of energy optimization and user comfort has gained vital importance in recent years due to the unnecessary wastage of electrical energy in residential buildings. Alambagh, Lucknow **[5].** The concept of virtualization technology in a distributed operating environment is crucial, and cloud computing is now the hottest area of research. Expanded functionality, efficient resource management, and a collaborative execution strategy all contribute to cloud computing's expanded application potential, and virtualization helps to further increase the

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8310627

elasticity of the resources in a distributed and cloud computing setting. Álvaro López García1 et al **[6]** We propose a distributed architecture to supply machine learning practitioners with a suite of tools and cloud services that span the entire machine learning development lifecycle, from model conception and training to testing and sharing. Anu Kaul, et al **[7]** Grid, cloud, utility, and cluster computing are only some of the recent developments in the field of distributed computing. In their research, we addressed the several tiers at which virtualization for distributed computing can be implemented, each of which can help boost distributed computing's efficiency and performance. Avinab Marahattaa et al **[8].** At first, a schedule history is used to categorize the various types of work and virtual machines. Then, jobs of a similar nature are aggregated and planned so as to make the most efficient use of the host's current operational condition. Bhaskar. R et al **[9].** In their research, they suggest using a dynamic approach to allocating resources in the cloud. The term "cloud computing" refers to a method of providing IT services wherein data and programs are accessed over the web instead of being stored locally on a server. Elrasheed Ismail, et al **[10].** In their research, they suggested a method for scheduling tasks that uses the load-balancing capabilities of the Quantum Particle Swarm algorithm (BLQPSO).

The make-span and information-transfer costs are minimized as part of the fitness function. In addition, a comparison of the suggested algorithm's performance with that of the current PSO demonstrates the latter's superiority in load balancing. Geetha Muthusamy **[11]** K-Means clustering is used to propose a cluster-based task scheduling framework (CBTS) that takes VM capacity and task length into account. Here, we group the jobs according to their duration and the virtual machines (VMs) according to their processing power. Georgios P. et al **[12]** to solve a wide variety of process scheduling issues via the creation of optimization-based approaches. Despite the vast literature on the subject, optimization-based scheduling has had only modest adoption in actual, working industries. Gopika Venu et al **[13]** with the jobs' and cloud providers' requirements in mind. Quality of Service (QoS), fairness, and efficiency of jobs all need to be taken into account when a scheduler puts them in order. The latency, cost, priority, etc., that are taken into account during the development and implementation of a scheduling algorithm. I. El Abbassi, et al **[14]** Energy management and optimization are gaining prominence in the industrial industry due to rising energy prices and worries about their impact on the environment.

The end goal is to reduce costs and increase productivity by controlling and optimizing energy consumption. J. Uma, et al **[15]** Load balancing is essential in cloud computing because it ensures that the workload is distributed fairly across all available nodes and that all available resources are utilized effectively. Through load balancing, each processor or network node is assigned roughly the same amount of work at any one time. Throughput, performance, fault tolerance, migration time, response time, etc. are just few of the factors that have been discussed in regards to the method. **Luiz F. Bittencourt et al [16]** in this work, we provide a taxonomy for classifying the scheduling problem in distributed systems, taking into account cutting-edge innovations like cloud computing. We examine the level of interest in each scheduling discipline and review the relevant literature to back up our suggested taxonomy. At last, we point the way forward for scheduling in distributed systems. Mahendra Kumar

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8310627

Gourisaria et al **[17]** to make the most efficient use of available resources while reducing energy use. However, the maximum resource utilization does not imply proper energy utilization. In order to make the most of the time that was not put to use, the idle resources have been given new assignments. Nagashetty et al **[18]** the cloud is a network that allows users to access shared computing resources whenever they need them. In this work, we present an Optimized Task Scheduling Algorithm that takes into account the peculiarities of cloud resources, such as their distribution and scalability, while adapting the benefits of many other current algorithms as necessary. NikhatAzhar et al **[19]** Cloud computing is a well-known technology that provides consistent, safe, and scalable computing services over the internet. By contrasting the new load collating algorithm with the old one, we can ensure its legitimacy and performance. O. M. Elzeki et al **[20]**

The goals of a distributed system are to improve the speed, stability, accuracy, and availability of computing on a broad scale. Distributed environments like the cloud and grid computing are now commonly used. Pinal V Chauhan **[21]** Their research delves into the foundations of cloud computing with the ultimate goal of integrating it into dispersed systems. The authors' goal was to examine the significance of configuration in securing cloud services as a means of addressing the security concerns inherent to cloud computing. Priyanka Sharma et al **[22]** Cloud computing has rapidly become the dominant computer model, as it is well suited to supporting on-demand services. Rakesh Kumar **Phanden et al [23].** There are a lot of obstacles in the way of scheduling studies in cloud manufacturing. Scheduling in the cloud manufacturing environment has unique challenges, therefore it's important to understand the existing state of affairs and anticipate potential difficulties. Reeya Manandhar et al **[24]** In order to maximize system performance while reducing infrastructure costs, the development of a reliable distributed network has assumed more significance. This document provides a summary of the many virtualization methods now in use and the advantages they offer. Reduced infrastructure costs, greater system scalability, and better resource usage are all possible thanks to these methods. Shahbaz Afzal et al **[25]**

Their research provides a comprehensive overview of available load balancing methods. In order to create effective load balancing algorithms in the future, we emphasize the benefits and drawbacks of current approaches and address major difficulties. Sukhpreet Kaur, et al **[26]** In cloud computing, software and services are delivered through the internet rather than on-premises computers. Requests to use services and software are sent from a variety of users. The cloud's capacity is being tested by this unprecedented influx of searches. Supreet Kaur, Amanpreet Singh, et al **[27]** Cloud computing is a relatively new paradigm in information technology that allows users to conveniently and cheaply share a variety of services. High-performance computing is now much more accessible thanks to cloud computing. Umesh Kumar Lilhore et al **[28]** Load balancing between virtual machines and system resources is a significant open problem in cloud computing. The major goal of this study is to improve reaction time, decrease readiness time, maximize source utilization, and shorten activity rejection time all within the context of a scalable cloud load balancing strategy. The suggested MLBL method is heavily reliant on support vector machines and the K-suggest clustering technique.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8310627

2.1 Existing models and draw backs

Cloud intrusion detection and prevention using the Hidden Markov Model is effective in terms of detection rate and cost effectiveness, but it still requires improvements in bandwidth and energy consumption [29]. However, in today's cloud contexts, multicore physical machines offer greater advantages because a single PM CPU core can support numerous multi-CPU virtual machines. The more realistic case of multi-core PMs, which may lead to a more complex scheduling issue, is a worthwhile area to study in the future.

3. PROPOSED MODEL AND SIGNIFICANCE

In cloud computing, resource provisioning refers to the act of allocating a cloud provider's resources and services to a customer. Resource provisioning is the process of selecting, implementing, and administering software (like load balancers and database server management systems) and hardware (such CPUs, storage devices, and networks) to guarantee the optimal performance of applications. In order to make the most use of the available resources while still meeting SLA and QoS benchmarks, it is necessary to either statically provision or dynamically allocate the resources in question. Preventing both over- and under-provisioning of resources is crucial. An additional constraint that must be taken into account is power consumption. Optimal performance depends on minimizing power consumption, heat generation, and the location of virtual machines. Methods should exist to prevent the waste of so much energy. In light of this, cloud customers want economical resource subscriptions. On the other hand, a cloud service provider's goal is to maximize profit through efficient resource allocation.

3.1Resource provisioning:

Provisioning resources can be done in either a static or dynamic setting. Prior to the execution of user tasks, CSPs make resource allocation decisions in a static environment. There is no way to add or acquire resources on the fly, meaning that a workload must make do with the amount of processing power it was originally allotted. Overall cloud performance and efficiency decrease as a result.

3.1.1 Task scheduling: Numerous algorithms for organizing tasks prioritized the completion of certain goals set by the user. User-requested quality-of-service metrics (makes pan, deadline, response time, delay, cost, VM bandwidth, etc.) make up the bulk of these aims. Customer-provider SLAs were the subject of discussion for some of them. Power and energy consumption, along with financial gain, are typical CSP-driven purposes.

3.1.2 Make span: The time it takes to finish a user-submitted task in its entirety. In this overview, makes pan is highlighted as a crucial variable for most of the algorithms discussed. Inconsistent arrival times is a major indicator of service quality. As a result, one of the criteria in this evaluation is the length of time it takes to respond to prospects.

3.1.3 Deadline: Submitting scientific procedures to the cloud typically has a specific time limit. A sufficient number of time-sensitive works were considered in this study.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8310627

Cost: Reduced computational expenses is cloud computing's primary objective. Algorithms attempt to minimize use costs or maximize service efficiency in relation to the money clients pay to employ a service.

3.1.4 Profit: CSPs are attempting to maximize income by acquiring a large number of consumers while providing low-cost services to those customers. Offering a variety of products and services helps achieve this objective, as does supporting high rates of resource application.

3.1.5 Energy: Reducing operational costs is critically dependent on energy use. The cost of electricity is one of the largest expenses in maintaining a cloud server. Power utilization and energy consumption are major considerations in most recently proposed methods.

3.1.6 Multi-Objective: The recent advancements in cloud scheduling methods have given attention to multiple criterions in task scheduling. These criterions are sometimes contradictory, so a tradeoff is needed between different solutions produced by the scheduler

3.2 Optimization methods:

Algorithms can also be classified according to the optimization policies they employ when scheduling. Task scheduling is difficult in the cloud because of inherent fluidity. Finding an optimum solution for task allocation in the cloud's ever-changing environment is challenging since scheduling in the cloud is NP-hard. In addition, the answers are derived from many presumptions about the cloud ecosystem's health. Algorithms that take their cues from nature can use heuristics to generate satisfactory approximations of the most accurate result. Some examples of heuristics can be seen in ants, bees, and bird flocks. Based on the specifics of the application case, sub-optimal algorithms can be further categorized as heuristic, meta-heuristic, or hybrid.

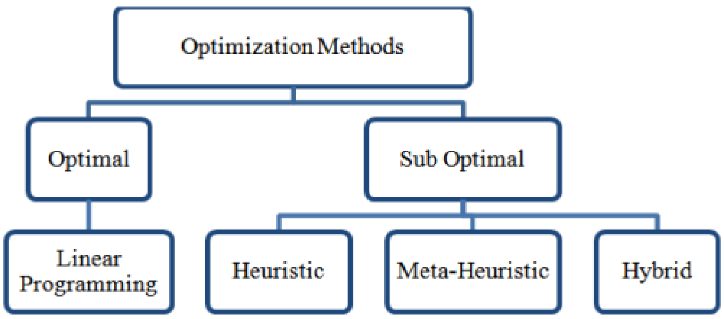


Figure 3: Optimization Methods

3.2.1 Linear programming model:

Linear programming, also known as linear optimization, is a mathematical technique for optimizing a mathematical model based on linear relations between its needs (such as maximum profit or minimum cost). Mathematical optimization is a subset of mathematical programming. Linear scheduling is used in the cloud because it yields the best results for scheduling tasks. A computer system can be considered an agent if it is capable of making independent decisions, acting autonomously, and interacting cooperatively and cooperatively

The Seybold REPORT
ISSN 1533-9211
DOI: 10.5281/zenodo.8310627

with other agents to achieve a state in which all actions complement each other.

3.2.2 Heuristic methods:

A heuristic is an algorithm that returns a solution without a guarantee that it is optimal. These algorithms can find near-optimal solutions rapidly and efficiently. Heuristic algorithms use conventional methods like greediness to streamline and accelerate the process. This is done by downplaying or omitting some of the work to be done. Its goal is to find solutions to complex problems faster and efficiently than more traditional methods. However, it does so at the expense of optimality, accuracy, precision, or completeness. In order to solve NP-complete problems, programmers frequently resort to heuristic approaches. While heuristics are frequently used as a jumping-off point or to generate initial solutions, they are only sometimes the most effective and are thus commonly paired with optimization techniques. The combination of PSO and a resource selection algorithm is one of the most effective in the field of heuristic algorithms. This allows for the generation of optimal energy consumption solutions.

3.2.3Meta-heuristic methods:

Heuristic algorithms perform some tasks because they provide optimal answers in a predetermined time. While more computationally intensive than heuristic algorithms, meta-heuristic algorithms are well-suited to solving various issues. Meta-heuristics are suitable for optimal solutions in elastic computing, where resources are unlimited, and the environment is challenging.

4. MIGRATION POLICY AND FEATURE SELECTION

Virtual machines (VMs) that have been consolidated into one server can be moved live and idle nodes can be switched to sleep mode, which maximizes resource utilization and decreases energy consumption. Due to workload fluctuations experienced by modern applications, online optimization of VM placement is essential. With these policies in place, the system may be programmed to change its behavior automatically based on the workload patterns assessed by the applications. This in turn improves the system's energy efficiency and performance. First, a check for host overload is conducted, followed by a virtual machine selection policy in order to determine which virtual machine should be migrated, and lastly, a virtual machine placement algorithm determines where to submit the migrated virtual machines.

The study looked at the effectiveness of migrating virtual machines based on different single objectives. It verified the effect of multi-objective assessment to find the most appropriate option and implement it across a wide range of circumstances.

4.1 Method

Simulate a cloud environment with dynamic users and datacenters. 2. A packet can be sent to the datacenter as soon as the cloud is established. 3. Initialize the OCRPC with certain parameters at the cloud server or broker level. 4. Cloud users determine the number of states in the model.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8310627

Algorithm: MOVMP (Multi-objective Virtual Machine migration policy)

{

1: migration list L = {}

2: migration process xij = null

3: Calculate targets T-CPU , T-mem

4: Construct matrix X

5: X = MigrationPolicy(X)

6: While (True) 7: Add(L, xij)

8: If (TCPU and Tmem) met

9: break 10: else

11: Update (X,i,j)

12: EndWhile 13: return L 14: Update (X,i,j)

15: remove row(i)

16: update column(j)

}

Virtualization technology enables cloud computing / virtual machine environments to pool and share resources; these pooled resources are represented by "virtual machines," which are subject to certain restrictions, such as processor power, memory size, and system transfer speed. Thousands of directions are obtained for the preparation of work, and each virtual machine has a speed of millions of regulations per second.

Proposed algorithm: OCRPC (optimized cost-effective resource provisioning in cloud)

1: Initialization

2: Input: Task ti from task queue Ti 1, 2, 3n

3: Output: Load-based classification of resources

4: T[i] = Number of task

5: Th = Throughput

6: Wsi = Weight of Server

7: Wsi=n= Weight of all Server

8: HMM = Hidden Markow Model

9: {

10: Get from user(job_info)

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8310627

```
Get from Resource(resource_utilization)
Requirements for specific tasks
/ Calculate the percentage of resources utilized /
Verify the availability of resources
Calculate CPU availability=Total CPU needed by task T[i]
Calculate memory availability = Total amount of memory required By task T[i]
Calculate storage availability = Total storage needed By task T[I]
When Resource Availability exceeds Task Need, then
{
Prepare a list for execution;
}
end if
/ Calculate the throughput based on the given resources /
Prepare a list of tasks with execution times
Prepare a list with the average execution time for each task
No. of Tasks x Average Execution Time
Data center resource usage
HMM for observing usage patterns
Classified resources based on their state
Based on certain usage values defining the current state of the system Resources
State 1 Less utilize
State 2 Average utilize
State 3 Heavy utilize
Apply the scheduling algorithm dynamically to the re-sources
if State 1 = usage below 30% then
{
```

ISSN 1533-9211
DOI: 10.5281/zenodo.8310627

37: Then Resource is under utilized

38: Apply appropriate scheduling algorithm from the scheduling library

39: }

40: else

41: (State 2 = usage above 30% and below 70%)

42: {

43: Then resource is average utilized

44: Apply appropriate scheduling algorithm from the scheduling library

45: }

46: else

47: (State 3 = usage above 70%)

48: {

49: Then resource is heavy utilized

50: Apply appropriate scheduling algorithm from the scheduling library

51: }

52: end if

53: }

At first the collected virtual machine is finished by utilizing correlation-based selection among the n number virtual machines. Initiate the number of populations in the job and the virtual machine, and after that assess the fitness capacity of each virtual machine by utilizing the crow search. The fitness value is computed by the data transfer speed and the limit of the virtual machine and for the resource, with the minimum execution time and deadline are acquired. The best fitness value is picked as the g best virtual machine and after that allocates the resource to the optimal virtual machine

5. RESULTS AND DISCUSSIONS

In this work proposes optimization algorithm (Cost effective based resource provision in cloud) for mapping tasks to virtual machines in order to improve the throughput of the datacenter and reduce the cost without violating the Service Level Agreement for an application in cloud environment. Scheduling refers to the mapping or assigning a task to a specific virtual machine

ISSN 1533-9211

DOI: 10.5281/zenodo.8310627

such that resource utilization increases. The tasks in the group are chose sequentially and proposed to the Virtual Machine. After receiving the list of virtual machines cloud broker assigns quality of service to the virtual machines.

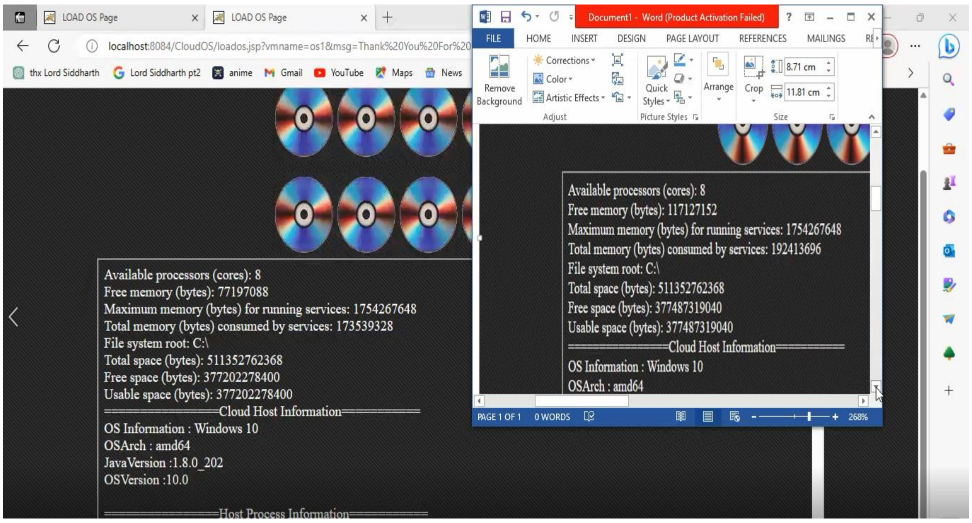


Figure 4: VM utilization in cloud environment for service usage proposed and existing

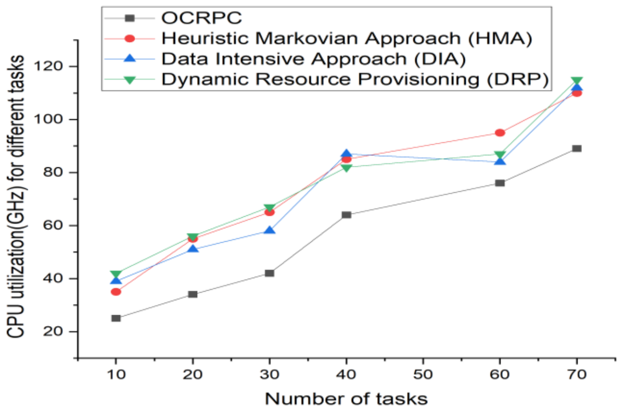
An efficient task scheduling algorithm named as OCRPC improves the overall system performance and helps service provider to provide good quality of services. In Resource selection broker plays an important role. Brokers have the list of virtual machines and its quality of service. A high-performance virtual machine assign with the high quality of services. "Dynamic provisioning" methods allow virtual machines to be moved from one computing node to another within the cloud on-the-fly in situations where application demand changes. Suitable for workloads and demands that are erratic or changing. An example would be a web server that can be configured with a configurable amount of CPU, memory, and storage. Using this approach, the client only pays for what is actually consumed. Clients must ensure that resources are not oversubscribed to avoid skyrocketing fees.

Table 1: CPU utilization for different tasks in cloud

Number of tasks	OCRPC	Heuristic Markovian Approach (HMA)	Data Intensive Approach (DIA)	Dynamic Resource Provisioning (DRP)
10	25	35	39	42
20	34	55	51	56
30	42	65	58	67
40	64	85	87	82
60	76	95	84	87
70	89	110	112	115

ISSN 1533-9211

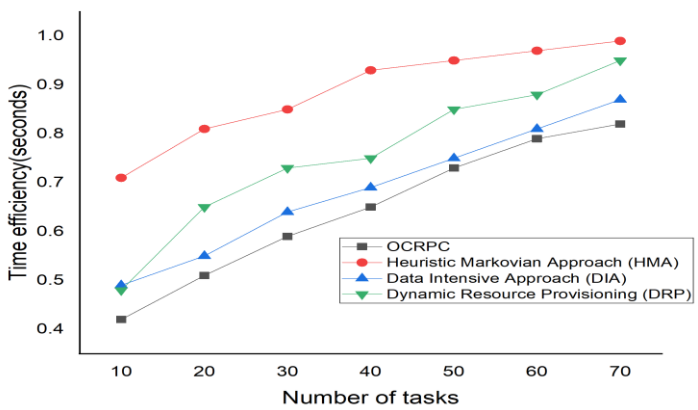
DOI: 10.5281/zenodo.8310627



Graph 1: CPU utilization for different tasks in cloud

Table 2: Time efficiency for different services and tasks in cloud computing

Number of tasks	OCRPC (Seconds)	Heuristic Markovian Approach (HMA) (Seconds)	Data Intensive Approach (DIA) (Seconds)	Dynamic Resource Provisioning (DRP) (Seconds)
10	0.42	0.71	0.49	0.48
20	0.51	0.81	0.55	0.65
30	0.59	0.85	0.64	0.73
40	0.65	0.93	0.69	0.75
50	0.73	0.95	0.75	0.85
60	0.79	0.97	0.81	0.88
70	0.82	0.99	0.87	0.95

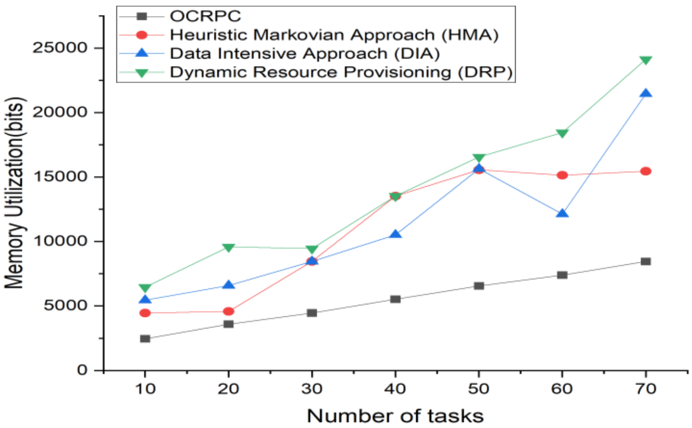


Graph 2: Time efficiency for different services and tasks in cloud computing

DOI: 10.5281/zenodo.8310627

Table 3: Memory utilization for different services in cloud computing

Number of tasks	OCRPC (Bits)	Heuristic Markovian Approach (HMA) (Bits)	Data Intensive Approach (DIA) (Bits)	Dynamic Resource Provisioning (DRP) (Bits)
10	2457	4457	5457	6457
20	3587	4587	6589	9587
30	4458	8458	8462	9458
40	5524	13524	10524	13524
50	6563	15563	15638	16563
60	7395	15145	12132	18452
70	8457	15457	21453	24131



Graph 3: Memory utilization for different services in cloud computing

6. CONCLUSIONS

Cloud Computing resource provisioning entails selecting, deploying, and managing software (such as database servers, load balancers) and hardware resources (such as CPUs, storage, networks, etc.) for achieving guaranteed application performance. These techniques improve response time, performance, save energy, QoS, SLA. From a cloud service provider's perspective, resource provisioning aims to maximize profits and reduce costs from the cloud user's perspective. Instead of allocating VMs through CSPi, our approach allocates VMs directly from server sites using already allocated VM resources. The analysis section clearly indicates that the proposed approach with direct allocation of VMs through server sites gives better optimization of resources, maximum utilization, and better performance.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8310627

References

1) Rudresh Magadum, D.B.Kulkarni, “Power loss reduction by optimal location of DG using fuzzy logic”, IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM-2015), Vel-Tech University, Chennai, T.N., India. May 6 – 8, 2015. pp.338-343.

2) Rudresh B Magadum, D.B.Kulkarni, “Optimal Location and Sizing of Multiple DG for Efficient Operation of Power System”, 4th IEEE International Conference on Electrical Energy Systems (ICEES-2018), SSN College of Engineering, Chennai, T.N., India. February 7-9, 2018.

3) Rudresh. B. Magadum, D.B.Kulkarni, “Optimal Placement of Distributed Transformer with STATCOM to Enhance the Efficiency of the Distribution Networks”, International Journal of Recent Technology and Engineering (IJRTE), Volume-8, Issue-2, pp.3720-3725, July 2019

4) Abdul Salam Shah, Haidawati Nasir, Muhammad Fayaz, Adidah Lajis and Asadullah Shah, A Review on Energy Consumption Optimization Techniques in IoT Based Smart Building Environments, Information 2019, 10, 108; doi:10.3390/info10030108

5) Alambagh, Lucknow, Virtualization for Cloud Computing: A Perspective View, International Journal of Computer Science Trends and Technology (IJCST) – Volume 5 Issue 3, May – Jun 2017

6) Álvaro López García, Jesús Marco De Lucas, Marica Antonacci, Wolfgang Zu Castell, Mario David, Marcus Hardt, Lara Lloret Iglesias1, Germán Moltó7, A Cloud-Based Framework for Machine Learning Workloads and Applications, Doi 10.1109/Access.2020.2964386, IEEE Access.

7) Anu Kaul, meena Gupta, khushwant Kaur, Using Virtualization for Distributed Computing, International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835 Volume-2, Issue-7, July-2015

8) Avinab Marahattaa, Sandeep Pirbhulalc, Fa Zhangb, Reza M. Parizid, Kim-Kwang Raymond Chooe, Zhiyong Liu, Classification-based and Energy-Efficient Dynamic Task Scheduling Scheme for Virtualized Cloud Data Center, 2019, IEEE Transaction on Cloud Comptuing, Vol., No.,

9) Bhaskar. R, Deepu. S.R and Dr. B.S. Shylaja, Dynamic Allocation Method for Efficient Load Balancing In Virtual Machines For Cloud Computing Environment, Advanced Computing: An International Journal (ACIJ), Vol.3, No.5, September 2012

10) Elrasheed Ismail, Noraziah. A, Faisal Alamri, Abdulla. A Optimized Load Balancing based Task Scheduling in Cloud Environment, International Journal of Computer Applications (0975 – 8887) Majan College International Conference (MIC-2014*)*

11) Geetha Muthusamy, Suganthe Ravi Chandran, Cluster-based Task Scheduling Using K-Means Clustering for Load Balancing in Cloud Datacenters, Journal of Internet Technology Volume 22 (2021) No.1

12) Georgios P. Georgiadis 1,2, Apostolos P. Elekidis 1,2 and Michael C. Georgiadis 1,2,*, Optimization-Based Scheduling for the Process Industries: From Theory to Real-Life Industrial Applications, Processes 2019, 7, 438; doi:10.3390/pr7070438

13) Gopika Venu, Vijayanand K S, Task Scheduling in Cloud Computing: A Survey, International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429 Volume 8 Issue V May 2020

14) I.El Abbassi, K. Labadi, S. Hamaci, and A.-M. Darcherif, Energy Management and Optimization in Manufacturing Systems, Proceedings of the Asia Pacific Industrial Engineering & Management Systems Conference 2015

15) J. Uma, V. Ramasamy, P. Vivekanandan, Load Balancing Algorithms in Cloud Computing Environment - A Methodical Comparison International Journal of Engineering Research & Technology (IJERT) Vol. 3 Issue 2, February – 2014 ISSN: 2278-0181

ISSN 1533-9211

DOI: 10.5281/zenodo.8310627

16) Luiz F. Bittencourt, Alfredo Goldman, Edmundo R.M. Madeira, Nelson L.S. da Fonseca a, Rizos Sakellariou, Scheduling in distributed systems: A cloud computing perspective, Computer Science Review 30 (2018) 31–54

17) Mahendra Kumar Gourisaria, S. S. Patra , P. M. Khilar, Minimizing Energy Consumption by Task Consolidation in Cloud Centers with Optimized Resource Utilization, International Journal of Electrical and Computer Engineering (IJECE) Vol. 6, No. 6, December 2016, pp. 3283~3292 ISSN: 2088-8708

18) Nagashetty, Dr. Shiva Murthy G, Ramakrishna Prasad A.L, An Optimized Task Scheduling Algorithm and Maintaining the Load in Cloud Environment, International Journal of Science and Research (IJSR), ISSN (Online): 2319-7064 Index Copernicus, Value (2015): 78.96

19) NikhatAzhar, Mohd. Haroon, Dynamic Load balancing by Round Robin and Warshall Algorithm in Cloud Computing, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-X,Issue-X, January 2021.

20) O. M. Elzeki, M. Z. Rashad, M. A. Elsoud, Overview of Scheduling Tasks in Distributed Computing Systems, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012

21) Pinal V Chauhan, Cloud Computing In Distributed System, International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 10, December- 2012 ISSN: 2278-0181

22) Priyanka Sharma,Dr.Sanjay Shilakari,Uday Chourasia, Priyanka Dixit, Alpana Pandey, A Survey On Various Types Of Task Scheduling Algorithm In Cloud Computing Environment, International Journal Of Scientific & Technology Research Volume 9, Issue 01, January 2020 Issn 2277-8616

23) Rakesh Kumar Phanden, Aditya SV, Pritish Shubham, Prem Narayan Vishwakarma A state-of-the-Art Literature Review on Service and Task Scheduling in Cloud Manufacturing, Proceedings of the International Conference on Industrial Engineering and Operations Management Bangalore, India, August 16-18, 2021

24) Reeya Manandhar and Gajendra Sharma, Virtualization in Distributed System: A Brief Overview, BOHR International Journal of Computer Science 2022, Vol. 1, No. 1, pp. 41–45

25) Shahbaz Afzal and G. Kavitha, Load balancing in cloud computing – A hierarchical taxonomical classification, Journal of Cloud Computing: Advances, Systems and Applications (2019) 8:22

26) Sukhpreet Kaur, Dr. Jyotsna Sengupta, Load Balancing using Improved Genetic Algorithm (IGA) in Cloud Computing, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 6, Issue 8, August 2017, ISSN: 2278 – 1323

27) Supreet Kaur, Amanpreet Singh, Rajeev Kumar, Architecture of Cloud Computing and The Concept of Virtualization, International Journal of Emerging Science and Engineering (IJESE) ISSN: 2319–6378, Volume-4 Issue-11, June 2017

28) Umesh Kumar Lilhore, Sarita Simaiya, Kalpna Guleria, and Devendra Prasad, An Efficient Load Balancing Method by Using Machine Learning-Based VM Distribution and Dynamic Resource Mapping, Journal of Computational and Theoretical Nanoscience Vol. 17, 2545–2551, 2020.6

29) Harsha Banafar et al / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 4808-4817 www.ijcsit.com.

30) Thummuluru Kavitha, Dr.Thatimakula Sudha, Dogo Rangsang Research Journal UGC Care Group I Journal ISSN : 2347-7180 Vol-13, Issue-6, No. 07, June 2023 Page | 62 Copyright @ 2023 "Dynamic And Optimal Rdynamic And Optimal Resource Utilisation And Provisioning Based On Workload In Cloud Environment"

31) Thummuluru Kavitha, Dr.Thatimakula Sudha, Indian Journal Of Research | Volume - 11 | Issue - 07 |July - 2022 | PRINT ISSN No. 2250 - 1991 | DOI : 10.36106/paripex "Load Balancing In Cloud Computing With Genetic Algorithm Based Hybridization"