Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

CRA-DP-GA for Efficient Utilization of Resource through Virtual Machine and Efficient VM in Cloud Data Centre

[*1]S. Boopalan,[2]Dr. Puneet Goswami

[*1]Research Scholar, SRM University, Delhi NCR, Sonepat, Haryana

[*2]Professor, Department of Computer Science and Engineering, SRM University, Delhi-NCR, Sonepat, Haryana

[*1]sribalulohith@gmail.com

Article Info

Page Number: 233 - 251

Publication Issue:

Vol 72 No. 1 (2023)

Abstract

Cloud computing is linked to cost reduction and efficient resource utilisation. Existing systems have a substantially higher cost of resources. Various resource utilisation, energy efficiency, and resource problems exist in cloud computing systems. To address these difficulties, integrated technologies such as task scheduling and virtual machines (VMs) are deployed. The literature on job scheduling is voluminous. For many parameters and objectives, this problem has been investigated. These data centres limit energy usage without sacrificing performance in order to be environmentally friendly data centres. Because processor energy usage accounts for 60% of overall power consumption, it is a key indication of server energy conservation.Identification of costs and renewable energy using a cluster selection approach and a virtual machine (FFD) algorithm, a dynamic PUE genetic algorithm (CRA-DP-GA) is investigated for deploying virtual machines (VMs). The host selection algorithm is one of two algorithms that enable DVFS. The suggested algorithm's major goal is to keep the server load balanced while changing the cooling load dynamically in response to the load.The suggested solution supports the VM clustering process, which installs and assigns virtual machines (VMs) based on the size of work required by the bandwidth level in order to increase efficiency and availability. According to the migration, the recommended clustering procedure is split into two parts: pre-clustering and post-clustering. The suggested virtual machine cluster's major goal is to map jobs to appropriate virtual machines using bandwidth for high availability and dependability. Task execution and assignment time are lowered when compared to previous techniques.

Article History

Article Received: *15 October 2022*

Revised: *24 November 2022*

Accepted: *18 December 2022*

Keywords: Cloud computing, virtual machines (VMs), First-fit Decreasing (FFD),Cost and Renewable Energy-Aware Dynamic PUE Genetic Algorithm (CRA-DP-GA), pre-clustering, post-clustering.

Introduction:

To keep the data on your computer, you must first have access to it. It's a lot more difficult to get data from all across the world. We used connected data centres to store data in the 1980s, but they were expensive to construct. It is need to be able to connect to the grid in order to get resources from a fixed wired server. This is a time-consuming job; however, the cloud can be utilised for more than just storage; it may also be used for computation and networking. A computer network, like any other network, includes complex components that affect a set of machines. The clouds in the illustration indicate the fact that the system's intricacies are concealed in the clouds since they have no bearing on the target.The user does not need to know about the computer system that provides the service. Because the specifics of what transpired are irrelevant, the system will appear in the cloud.Users have begun to purchase or install their own software and data centres in order to avoid this. It ended up costing a lot of money. If a resource isn't being used after work is completed, it's being used inefficiently [1]. As a result, resource use is inefficient and scalable. The cloud is also utilised for processing and networking, in addition to storage. To solve these issues, the notion of cloud computing was born from the usage of resources for lease or lease-based processes. These facilities can connect to resources from anywhere on the earth, and the user has no notion where the data is kept.

Every year, a significant amount of data is produced. As a result, a lot of processing power and storage is needed.Many scientific domains, including astronomy, bioinformatics, meteorology, environmental science, and geology, use it to handle large-scale data.The processing of massive amounts of data generated by numerous scientific fields may have a significant impact on cloud performance [2]. Ensure efficient task scheduling in cloud computing to improve cloud computing performance can be a difficult undertaking to accomplish. Scheduling can be done at the IaaS, PaaS, and SaaS tiers in a cloud computing system [3]. The term "load balancing" refers to the process of distributing the load among available resources. Requests are accepted and distributed across available resources or virtual machines via a VM (load balancing method). A load balancer's job is to determine how much load is on the available resources and split it accordingly. When resource are used in load balancing algorithms incorrectly, the quality of service (QoS) decreases.

Forecasting resource use has gotten a lot of attention, and there's a lot of information out there. Model accuracy, time and memory complexity, and multi-resource processing are all factors that

influence the resource prediction model. How accurate is the resource utilisation prediction, as measured by several metrics (CPU and memory usage, disc I/O, network throughput, etc), the relationship between CPU and memory usage, the disc I/O and memory relationship, and so on.Relationships across resource types are difficult to identify and foresee, and managers must be able to handle several metrics at the same time.

A data centre is a critical piece of infrastructure that brings together large-scale computer and storage capabilities to enable on-demand computing. Grid computing platforms that use virtualization technologies make it easier for data centre clients to get computing resources as a service [4]. Energy demands for data centres are increasing. Energy consumption is growing at a pace of 10% to 12% each year [5]. To deliver the quality of service (QoS) required by hosting applications while enabling energy-efficient data centres, synchronised power and resource management is critical [6]. Maximizing servers usage while lowering energy consumption is a cost-effective strategy [7]. To save energy, use virtual machine (VM) integration and autoscaling [8]. To achieve the Service Level Agreement (SLA), which accounts for more than 80% of the IT expenditure, the maximum number of production servers is necessary. The idle server consumes two-thirds of the server's energy while the total load is 100% [9].The differing power models of physical servers have an impact on idle and dynamic power consumption. The energy savings gained by virtualizing existing resources can limit resource availability and jeopardise provider reliability. Putting a server under a lot of stress might raise its temperature and reduce its longevity.

Resource usage should be optimised according to the computational power of the server to reduce power consumption of idle and active servers [10]. With these considerations in mind, apply the approaches recommended to choose the server CPU (central processing unit) that uses the least amount of energy.

Literature Survey:

The following is a summary of what is known about binning mechanisms at the moment.

Iwendi C*et al.*,2021 propose an empty-package-based virtual machine deployment (VMP) strategy for cloud data centres that targets the heuristic, energy-aware, bandwidth-aware, and QoS-aware aspects employed in [11]. He also talked about the VMP methods for resource identification, power identification, network identification, and cost identification. When it comes to saving energy, the bin packing heuristic is extremely effective. To manage the initial assignment of virtual machines to

physical servers, RM, S.P*et al.*,2020 [12] devised an approach that uses VMware Hypervisor ESXi 5.5. In comparison to batch manual and indiscriminate allocation methods, the approach is based on the bin-packing first-fit reduction algorithm.

Zhang, X*et al.*,2019 propose the "MinTotal DBP" problem as a variation of dynamic binpacking to reduce server costs in [13]. For this task, we looked at the competitiveness rates of all packing algorithms (first fit, best fit, and arbitrary fit). Hybrid Priority, a new competitive algorithm, is also proposed in this study.

Rashida, S.Y*et al.*, (2019) looked into the possible connection between present resource allocation architecture and cloud computing, which is predicted to be at the heart of the Internet in the future. The authors also emphasise the importance of continuing to improve network awareness and resource allocation strategies, as well as identifying concerns that the study team should look into further [14]. In cloud technology, we also present a proprietary orbit-based network resource allocation strategy.Zhao, H*et al.*, (2018) devised a strategy for addressing these flaws. This concept is based on the use of analytical processing modelling methodologies combined with stochastic assessment metrics, as well as simulation studies to illustrate the strategy's efficiency, which takes into consideration both online and batch requirements [15].

Within the parameters of proposed service scaling strategies, Khoobkar, M.H*et al.*, (2022) provide a cloud data centre design problem. When executing tasks, the proposed programming method can efficiently increase resource utilisation while lowering energy usage. Context switching, processing time, and processing for the suggested estimate problem.He investigates and presents specific quality evaluation measures for translation rules, such as time and response time [16].

We looked into the energy efficiency of cloud data centres in depth. [17] proposed energy-saving technologies for cloud data centres in order to cut down on energy usage and increase resource utilisation. The approach is based on task and virtual machine classification, which cuts down on scheduling time. Based on prior task schedule data, categorise user tasks and choose VM types. The algorithm's goals include resource use, energy consumption, and fault tolerance. Merging comparable job types reduces total energy consumption while shortening average response times.

In order to plan workflows, Babazadeh Nanehkaran, A.*et al.*,2021 [18] created the BULLET algorithm. The technique, which is based on PSO, employs QoS measures to schedule cloud computing workloads based on user requirements. Set weights for each service's QoS needs metrics, such as energy consumption, execution time, and execution cost. When consumers request QoS, services are assessed using QoS measurements and weighted according to demand. The Workload Analyzer examines the load of various workloads to see whether they are cloud-portable. It will be sent to the workload administrator if the workload is runnable.

Karuppiah, S.V.*et al.*,2021 [19] suggested an energy computing framework for scheduling physical machine resources. Existing boxing approaches have been compared to the framework. Rodrigo N. Calheiros*et al.*,2011 [20] are a group of researchers who have come up with a novel approach to solving. This paper describes a binary programming technique to developing diverse strategies by enhancing heuristics. Resolved an issue with virtual machine placement considering start and finish times. The proposed heuristic can tackle big problems while keeping the optimality gap to a minimum. Based on the optimal reduction empty packing technique for impossible chromosomes, Gao, Y*et al.*,2013 introduced a hybrid genetic algorithm called "Hybrid Genetic Optimal Fitting (HBF BP)" in [21]. Virtual placement technology has been shown to be effective.People should be harassed. The first fit packing algorithm is used to present a new approach to the MinUsageTime problem and to set a new upper bound in [22]. M. Keshavarznejad et al., look into an asymptotic optimization method in [23], which has been proved to gradually reduce the number of servers.

[24] presented a dynamic method to take advantage of long-term activity bookings' lower prices and boost multiplexing to reduce costs. The main issue for cloud customers is limiting expenses by selecting from a variety of on-demand evaluation choices. Another benefit of cloud finance is that it keeps a significant burst from cloud providers while still giving clients value rebates. Long-term bookings and multiple pickups would ideally be abused by users.

Chauhan, S*et al.*,2021 [25] primarily created algorithms based on programming techniques, ideal cloud service providers, and cloud metaphysics programming algorithms to identify customer wants. Provides a user-friendly broker execution management system in a cloud computing environment, complete with superior prescription measurement programming approaches and a load balancing strategy that lets you to assign resources to many clients for various activities.

The cloud computing paradigm allows users to access some services, such as processing and storage, through the Internet. Cloud computing services are becoming more accessible and faster via the internet. Cloud-based efficient resource management can boost resource utilisation, increase application performance, and cut expenses. Cloud resource utilisation prediction approaches have a large body of literature.The procedure is described in full in this section. The researchers presented a regression ensemble method in [26] to forecast intelligent resource utilisation. To increase utilisation and performance, the suggested strategy combines resource usage with feature selection. As a result, we can see that the suggested model outperforms previous models in terms of both accuracy and execution time. This method decreases errors and allows for fault-tolerant scheduling in addition to better prediction. For cloud consumers, the scalability of virtualization technology means more time or fewer resources [27].

The rest of the document is organised in the following way: The first section looks at some of the most prevalent data centre power consumption figures. Section 2 delves into some of the task's closely connected research methodologies. The system model and phrasing of the research topic are discussed in detail in Chapters 3 and 4. The strategies for solving the formalised random issue provided in this study are described in Chapter 5. The experimental setup is discussed in Chapter 6. The importance of appropriate frequency and dynamic cooling loads are discussed in Chapter 7, as well as load balancing and simulation findings. The study's findings are finally presented in Chapter 8.

Contribution:

Operating frequency, cooling unit power consumption, and CPU power consumption are three triangle-related parameters examined for dynamic PUE. By distributing the load evenly across all servers, the Carbon Aware Power Efficiency Optimal Frequency (C-PEF) and Carbon Aware C-FFF (First-Fit Optimal Frequency) algorithms proposed in this white paper can be realised. The following variables are taken into account by the placement algorithm:

- To lower the data center's overall carbon impact, PUE and CO2 emissions are used to select data centres and clusters.
- Load balancing refers to a server operating at the lowest frequency possible for the current workload and quality of service in order to identify and eliminate CPU hot spots that have a direct impact on hardware life and performance.
- Examine how static and dynamic power efficiency (PUE) affects cooling load power and location decisions.

Problem Statement:

The issue is that virtual machines are supplied in isolation, and each disc is booted many times for a set amount of time. A cluster management solution based on docker containers in various configurations is used to solve this challenge. Existing docker container and virtual machine deployment is done separately and accomplished utilising the container VM-PM approach. The Internet of Things (IoT) is critical for processing real-time data from hardware devices that create large amounts of data. As part of big data analysis, these files are housed in massive data centres. A huge number of servers will be utilised to store the received data if the data volume is large. It is confronted with a costly issue that can be remedied using ProCon, a cabinet-based tool.Physical servers can now run on a wide number of computers with diverse resources thanks to virtualization technologies. Virtual storage minimises read/write latency by maximising IO efficiency. Synchronization allows user data to be processed and saved in virtual discs attached to the VM. End users can use Amazon cloud providers to get a variety of services, all of which are backed up by dependable and secure computing power. AWS offers multiple versions of VMs and resources for Elastic Compute Cloud (EC2). EC2 instances are used in the suggested technique as a starting point for additional analysis.

Methods:

The full mechanism is seen in Figure 1. All components have the following descriptions and characteristics:

- Data centre resource management system: This system maintains data such as cluster lists, PUE, CFR, total utility power, current IT load, and other metadata.
- Management Node (MN): On the management node, the daemon executes the resource allocation management (RAM) algorithm. Update the data centre cluster list, host list, power usage efficiency, carbon footprint percentage, and other information. With the VM-to-PM mapping plan and resource release method enabled, perform resource recovery and update the target virtual machine queue (TargetVMQ) with the VM-to-PM mapping data.
- CM (Cluster Manager): The cluster manager is the head node of the cluster.Cluster managers keep track of total cluster utilisation, number of systems turned on and off, maximum and minimum utilisation, number of virtual machines running in the cluster, and power consumption.

- PMM (Physical Machine Manager): In Headnode CM, PMM provides PM details for maintenance and updating, such as available memory and CPU capacity, current operating frequency, power consumption, CPU utilisation, number of active VMs, and other PM-related information.
- Virtual Machine Manager (VMM): VMM is a daemon that runs on all PMs. Upkeep of PM VMs is your responsibility. VMM keeps track of VM resource usage, % CPU usage, transfer time, deployment time, active and inactive status, remaining runtime, power consumption, and other VM details.

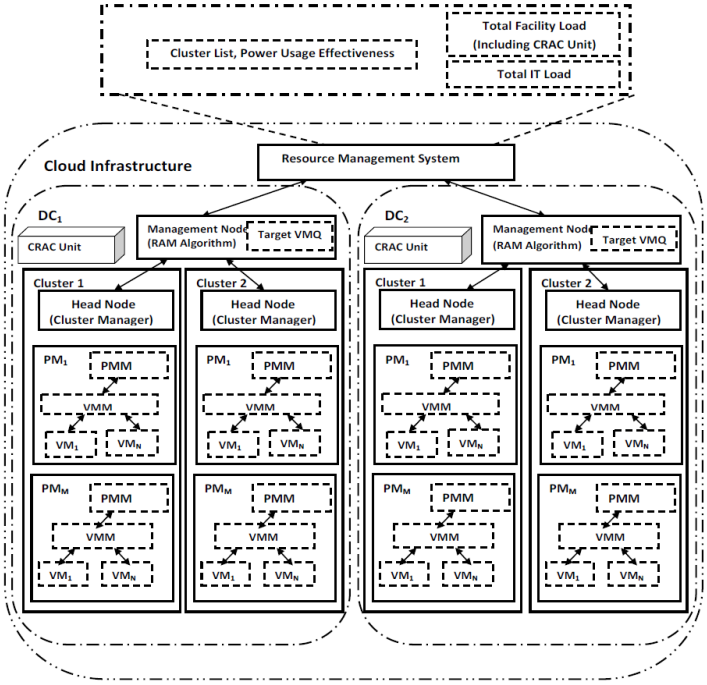

Fig 1: System Model

The suggested system is depicted in Figure 2 as a flowchart. This demonstrates the relationship between events. An activity diagram depicts how a process starts and finishes, as well as the many states and activities that occur during those states. To access the system's login module, you must first authenticate the user. Your user name and password will be emailed to you. They also provide you access to the cloud home page, which lists all of the important functions. Only authorised users are allowed to access the system through this module. It only allows authenticated users to access and create virtual machines, as well as see existing machines, isolate tasks, and examine utilisation reports.The first module option guides the user through the process of creating a module by simply

inputting settings for the new VM. Instead of typing instructions at the terminal, you can use this module to construct a virtual machine by simply typing values.

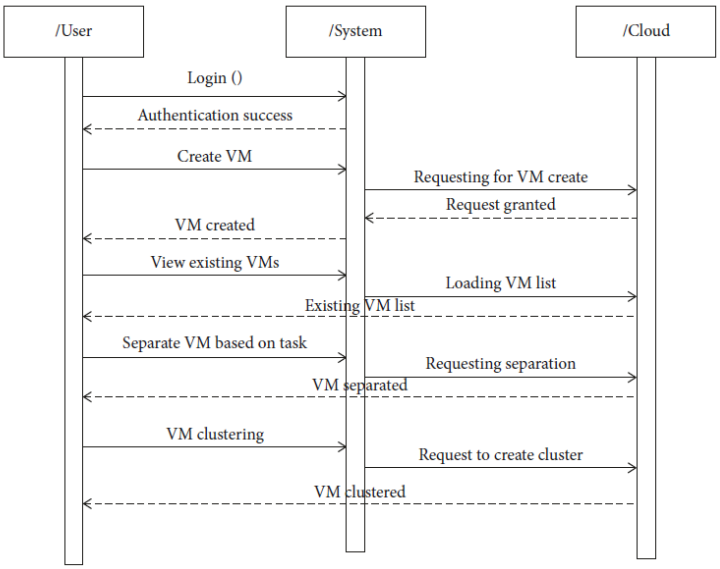

Fig 2: Phase of the proposed VM clustering process

Problem Formulation:

The reserved frequency is f ϵ F, the resource demand is r ϵ R, and the execution interval is e ϵ I in VM requests. M distinct frequencies (f 0, f 1, f 2, f 3, f 4,::: fk) and uses (U0, U1, U2, U3 ::::, Uk) Take a look at a server that serves as a model. U0 = 0% (idle), Uk = 100% (fixed dynamic power consumption) (P0, P1, P2 P3 P4:::, Pk). The power consumption is P0, and U0 is assumed to be in an idle state. For each Sj, let S = S1, S2, S3::.SM stand for M servers. where j"[1,M] is the utilisation rate Pj,3,Pj,4:::,Pj,k) in server Sj's power consumption (CUj,CPj,Cj), and *Cj* is the total processing capacity of *Sj*.

The R relationship between the jth PM and the ith VM decides whether VMi belongs in PMj, as shown in the diagram below:

$$R_{j,i} = \left\{ \begin{array}{ll} 1 & VM_i \text{ is allocated to } PM_j \\ 0 & \text{otherwise} \end{array} \right\}$$

The service level agreement is determined by the RVA (ratio of virtual machine acceptance) (SLA).

$$RVA(V) = T(R)/N$$

T(R) is the total number of VM requests that have been approved and mapped to accessible PMs, where N is the total number of VM requests. It is calculated in the following way:

$$T(R) = \sum_{j=1}^{M} \sum_{i=1}^{N} R_{j,i}$$

First-fit Decreasing (FFD) algorithm:

The FFD (First Fit Descending) technique is one of the most basic heuristics for solving the Bin packing problem, and it yields a quick and optimal result [35]. The FFD pseudocode is shown in Algorithm 1 The greedy allocation strategy is used again in Algorithm 2. Assume that the cloud proxy is inundated with requests for virtual machines iv. A buffer is used to keep track of this data. First, use multiple parameters to sort the buffer list in descending order. The number of CPU cycles required for each VM request is then sorted in descending order. The processor speed selected will be used if this option is set to the same value. The request will be sorted by memory if all of the above criteria are equal.Finally, the hard disc space necessary to perform the VM request will be used as the sorting parameter if all of the aforementioned request parameters are the same. Search for the first datacenter and choose the right host for the VM request after sorting the list. If the present data centre doesn't have any acceptable hosts, the next data centre is searched, and the best data centre and internal hosts are chosen.

Algorithm 1- FFD

Input: $DatacenterList, VmRequestList$

Output: $destinations$

$destinations = empty$

$sortedVmRequestList = sortVmRequestListbasedoncpu, Ram, Storage$

$\boldsymbol{for} i in sortedVmRequestList.size$:

$vmi = sortedVmRequestList.get(i)$

$destination = greedyAllocation(vmi, DatacenterList)$

$\boldsymbol{if}$ $(destination != null)$

$destinations.put\ (vmi, destination)$

else

reject vmi

return $destinations$

Algorithm 2- $greedy\_Allocation$

Input: $vm, datacenterList$

Output: $destination$

$destinations = empty$

$sortedVmRequestList = sortVmRequestListbasedoncpu, Ram, Storage$

for $jinDatacenterList.size$:

$di = DatacenterList.get(i)$

$hostsdi = di.getHosts( )$

for $kinhostsdi.size$:

$hk = hostsdi.get(k)$

if $(isHostFeasibleForVm(hk, vmi) == true)$

$destinasions.put(vmi, di, hk)$

return di, hk

else

continue

endoffor

return *null*# there is no feasible host for this VM at this time and we have to reject this VM request

endoffor

CRA-DP-GA Algorithm:

It is seen as a crucial remedy to children's problems. CRA-DP stands for "Cost and Renewable Energy Identification Dynamic PUE." Then, to overcome the problem's NP-hardness and time complexity, we present two meta-heuristics for solving combinatorial cost and scheduling optimization problems.

Algorithm 3-$GeneticAlgorith(CRA\_DP\_GA)$

Input:$acenterList, HostList, VmRequestList, PopulationSize, Pcrossover, Pmutation$

Output:$testChromosomep0 = initialPo(DatacenterList, HostList, VmRequestList, PopulationSize)$;

$current = p0$

$\boldsymbol{while}(convergencecriterionnotreach)$

$current = fitnessFunction(current)$;

$current = generateNextGeneration(current, Pcrossover, Pmutation, PopulationSize)$;

$\boldsymbol{end}$

$\boldsymbol{return} current.\ getFittestChromosome(\ )$;

Results and Discussion:

EAFT and C-PEF data are compared in this section. Because C-PEF employs the same simulator as C-PEF and is one of the most extensively used energy-efficient resource scheduling strategies, it was chosen for comparison. Under different settings, there exist results comparing load balancing, energy, and fault consequences. This section goes over the power consumption and energy efficiency of RAM, CPEF, FFT, and CRA-DP-GA scheduling algorithms. The switch power consumption of tiers and servers are compared in Table 1. Network equipment and computer servers are the two major sources of power consumption in data centres. The core, concentrators, and access switches are all placed in a highly redundant manner in a three-tier data centre network configuration.

Energy Consumption:

This number represents the total energy consumed in the data centre by all physical machines (PMs). The PM's energy consumption is estimated using a linear cubic energy model. With the CPU utilisation of this power mode, the physical host's power consumption grows linearly. The table 2 shows the energy consumption comparison.

Take a look at the power model's parameters:

(i) Pmaxk: maximum power used when the host k is fully loaded

(ii) Pidlek is the host's idle power value

(iii) uk: host k's current CPU usage

(iv) T: the data center's total number of hosts

The following is a formula for calculating the power consumption of the host Pk:

$$P_k = P_k^{idle} + \left(P_k^{max} + P_k^{idle}\right) * U_k^3$$

Table 1: Data Center Power Consumption in kWh

Element	RAM Scheduler	CPEF Scheduler	FFT Scheduler	CRA-DP-GA Scheduler
Core Switches	5.97	5.97	5.97	11.94
Aggregate Switches	5.97	5.97	8.96	23.89
Access Switches	1.67	1.83	2.16	5.33
Network Switches	13.6 (33%)	13.8 (34%)	17.1 (42%)	41.2
Servers	136.51 (37%)	155.2 (42%)	169.3 (46%)	368.6
Data Center	150.1 (37%)	168.9 (41%)	186.4 (45%)	409.8

Table 2: Comparison of Energy Consumption

No. of VMs	RAM Scheduler	CPEF Scheduler	FFT Scheduler	CRA-DP-GA Scheduler
100	100	80	75	65
200	200	185	165	120
300	310	290	260	185
400	400	380	345	265
500	520	460	410	315
600	605	575	525	385

Execution Time:

A cloud provider's ability to fulfil all user requests in a reasonable length of time is crucial. As a result, while analysing algorithms, one of the most significant factors to examine is the execution time. The table 3 shows the comparison of execution time.

Let *T ={T1,T2,. . ..TN}* be the set of tasks. |*T*| representsthe total number of tasks in the set. *VM={VM1,VM2,..,VMM}* denotes a set of virtual machines. TE(*i, j*)represents the execution time of task Ti on virtual machine *j*.It is expressed as

$$T_E(i,j) = \frac{t_l(i,j)}{VM_c}$$

Table 3: Comparison of Execution Time

No. of VMs	RAM Scheduler	CPEF Scheduler	FFT Scheduler	CRA-DP-GA Scheduler
100	30	28	25	18
200	50	45	40	35
300	75	70	65	50
400	130	110	100	75
500	145	135	125	90
600	175	160	150	115

Resource Utilization:

Cloud data centres create several sorts of virtual machines based on resource requirements in order to process user requests. The goal of VMP technology, which strives to increase resource use, is to arrange virtual machines on appropriate real equipment. (CPU is being considered as a resource.) Consider the case below: There are N PMs and M VMs in your system).Let *PM={PM1, PM2,. . .., PMN}* denote the set of PMs, where*PMi ∈PM*. *PMCi* represents the CPU capacity of *PMi*. Let*VM={VM1, VM2,..,VMM}* denote a set of virtual machines,where *VMj ∈VM*. *VMCj* represents the CPU requirement of

VMj.

Let *Pij* denote whether *VMj* is placed on *PMi*. If *VMj* isplaced on *PMi*, then *P*ij = 1 or else if *VMj* is not placed on*PMi*, then *P*ij = 0. The requirements of all virtual machines installed on a physical machine must not exceed the real machine's resource limit, as shown in equation (9). The table 4 shows the comparison of CPU utilization.

$$\sum_{j=1}^{M} VMC_j \cdot P_{ij} \leq PMC_i \forall PM_i \epsilon PM$$

Table 4: Comparison of CPU Utilization

No. of VMs	RAM Scheduler	CPEF Scheduler	FFT Scheduler	CRA-DP-GA Scheduler
100	17	18	19	20
200	19	20	21	23
300	22	23	24	26
400	23	24	25	27
500	25	26	27	29
600	27	28	29	32

Average Start Time and Finish Time:

Cloud companies must now focus on delivering exceptional performance to their customers. User request/task start and end times can be considered essential in this regard. The average start and end timings for various algorithms are shown. As a result, GA-RF completes user requests/jobs faster than competing algorithms on the market. The table 5&6 shows comparison of average start and end time.

Table 5: Comparison of Average start Time

No. of VMs	RAM Scheduler	CPEF Scheduler	FFT Scheduler	CRA-DP-GA Scheduler
100	20	19	18	15
200	40	35	20	18
300	78	74	58	40
400	100	96	66	48
500	130	105	90	72
600	135	115	100	80

Table 6: Comparison of Average End Time

No. of VMs	RAM Scheduler	CPEF Scheduler	FFT Scheduler	CRA-DP-GA Scheduler
100	25	23	20	15
200	60	50	45	25
300	90	80	75	70
400	110	105	90	80
500	115	110	100	85
600	120	115	105	90

Conclusions:

The proposed system's main goal is to cluster virtual machines (VMs) depending on several performance parameters like workload type and bandwidth. CPU-based, storage-based, and IO-based tasks make up the majority of your requests. The desired job must first be confirmed, after which it is assigned to the job classification process. The job classification method divides jobs into categories based on their characteristics. A clustering technique is used to arrange these activities into categories. Pre-clustering and post-clustering are the two sorts of clustering methods. The worker cluster's bandwidth is clustered according to their jobs. The suggested technique keeps two types of clusters, but they are bandwidth clusters that perform the same functions as functioning clusters.Virtual machines are classified and assigned to tasks that can be completed.The suggested technique's major goal is to match the requested task with the right virtual machine. Service processing latency is reduced more effectively.

List of abbreviations

VM	Virtual Machines
FFD	First-Fit Decreasing
CRA-DP-GA	Cost And Renewable Energy-Aware Dynamic PUE Genetic Algorithm
QoS	The Quality of Service
CPU	Central Processing Unit

Declarations

Availability of data and materials

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding received by any government or private concern.

Author's contribution:

S.B contributed to technical and conceptual content, architectural design. P.G contributed to guidance on the writing of the paper. All authors have read and approved the manuscript.

Acknowledgements

Not applicable.

References:

[1] Khosravi, A., Andrew, L.L. and Buyya, R., 2017. Dynamic VM placement method for minimizing energy and carbon cost in geographically distributed cloud data centers. *IEEE Transactions on Sustainable Computing*, *2*(2), pp.183-196.

[2] Ahvar, E., Orgerie, A.C. and Lebre, A., 2019. Estimating energy consumption of cloud, fog and edge computing infrastructures. *IEEE Transactions on Sustainable Computing*.

[3] Thiam, C. and Thiam, F., 2019. Optimizing Electrical energy Consumption in Cloud Data Center.

[4] Xu, M. and Buyya, R., 2020. Managing renewable energy and carbon footprint in multi-cloud computing environments. *Journal of Parallel and Distributed Computing*, *135*, pp.191-202.

[5] López, J., Kushik, N. and Zeghlache, D., 2019. Virtual machine placement quality estimation in cloud infrastructures using integer linear programming. *Software Quality Journal*, *27*(2), pp.731-755.

[6] Parvizi, E. and Rezvani, M.H., 2020. Utilization-aware energy-efficient virtual machine placement in cloud networks using NSGA-III meta-heuristic approach. *Cluster Computing*, pp.1-23.

[7] Aboutorabi, S.J.S. and Rezvani, M.H., 2020. An Optimized Meta-heuristic Bees Algorithm for Players' Frame Rate Allocation Problem in Cloud Gaming Environments. *The Computer Games Journal*, *9*(3), pp.281-304.

[8] Tavakoli-Someh, S. and Rezvani, M.H., 2019. Multi-objective virtual network function placement using NSGA-II meta heuristic approach. *The Journal of Supercomputing*, *75*(10), pp.6451-6487.

[9] Mohammadi, A. and Rezvani, M.H., 2019. A novel optimized approach for resource reservation in cloud computing using producer–consumer theory of microeconomics. *The Journal of Supercomputing*, *75*(11), pp.7391-7425.

[10] Laganà, D., Mastroianni, C., Meo, M. and Renga, D., 2018. Reducing the operational cost of cloud data centers through renewable energy. *Algorithms*, *11*(10), p.145.

[11] Iwendi, C., Maddikunta, P.K.R., Gadekallu, T.R., Lakshmanna, K., Bashir, A.K. and Piran, M.J., 2021. A metaheuristic optimization approach for energy efficiency in the IoT networks. *Software: Practice andExperience*, *51*(12), pp.2558-2571.

[12] RM, S.P., Bhattacharya, S., Maddikunta, P.K.R., Somayaji, S.R.K., Lakshmanna, K., Kaluri, R., Hussien, A. And Gadekallu, T.R., 2020. Load balancing of energy cloud using wind driven and firefly algorithms in internet of everything. *Journal of parallel and distributed computing*, *142*, pp.16-26.

[13] Zhang, X., Wu, T., Chen, M., Wei, T., Zhou, J., Hu, S. and Buyya, R., 2019. Energy-aware virtual machine allocation for cloud with resource reservation. *Journal of Systems and Software*, *147*, pp.147-161.

[14] Rashida, S.Y., Sabaei, M., Ebadzadeh, M.M. and Rahmani, A.M., 2019. A memetic grouping genetic algorithm for cost efficient VM placement in multi-cloud environment. *Cluster Computing*, pp.1-40.

[15] Zhao, H., Wang, J., Liu, F., Wang, Q., Zhang, W. and Zheng, Q., 2018. Power-aware and performance-guaranteed virtual machine placement in the cloud. *IEEE Transactions on Parallel and Distributed Systems*, *29*(6), pp.1385-1400.

[16] Khoobkar, M.H., Dehghan Takht Fooladi, M., Rezvani, M.H., Gilanian Sadeghi,M.M.,, 2022, BLMDP: Partial Offloading with Stable Equilibrium in Fog-cloud Environments using Replicator Dynamics of Evolutionary Game Theory, *Cluster Computing*, to be appear.

[17] Esfandiari, S. and Rezvani, M.H., 2020. An optimized content delivery approach based on demand–supply theory in disruption-tolerant networks. *Telecommunication Systems*, pp.1-25.

[18] Babazadeh Nanehkaran, A. and Rezvani, M.H., 2021. An Incentive-Compatible Routing Protocol for Delay- Tolerant Networks Using Second-Price Sealed-Bid Auction Mechanism. *Wireless Personal Communications*, *121*(3), pp.1547-1576.

[19] Karuppiah, S.V. and Gurunathan, G., 2021. Secured storage and disease prediction of E-health data in cloud. *Journal of Ambient Intelligence and Humanized Computing*, *12*(6), pp.6295-6306.

[20] Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov, César A. F. De Rose, Rajkumar Buyya, "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms", software practice and experience journal, Volume : 41, Issue1, pp.23 – 50, 2011.

[21] Gao, Y., Guan, H., Qi, Z., Hou, Y. and Liu, L., 2013. A multi-objective ant colony system algorithm for virtual machine placement in cloud computing. *Journal of computer and system sciences*, *79*(8), pp.1230-1242.

[22] Misra, S.K. and Kuila, P., 2022. Energy-Efficient Task Scheduling Using Quantum-Inspired Genetic Algorithm for Cloud Data Center. In *Advanced Computational Paradigms and Hybrid Intelligent Computing* (pp. 467-477). Springer, Singapore.

[23] Keshavarznejad, M. Rezvani, M.H., Adabi, S. 2020. delay-aware optimization of energy consumption for task offloading in fog environments using metaheuristic algorithms. *Cluster Computing*, 24, pp. 1825–1853.

[24] Maddikunta, P.K.R., Gadekallu, T.R., Kaluri, R., Srivastava, G., Parizi, R.M. and Khan, M.S., 2020. Green communication in IoT networks using a hybrid optimization algorithm. *Computer Communications*, *159*, pp.97-107.

[25] Chauhan, S., Singh, M. and Aggarwal, A.K., 2021. Cluster Head Selection in Heterogeneous Wireless Sensor Network Using a New Evolutionary Algorithm. *Wireless Personal Communications*, pp.1-32.

[26] Vashishtha, G. and Kumar, R., 2022. An amended grey wolf optimization with mutation strategy to diagnose bucket defects in Pelton wheel. *Measurement*, *187*, p.110272.

[27] Chauhan, S., Singh, M. and Aggarwal, A.K., 2021. Design of a Two-Channel Quadrature Mirror Filter Bank Through a Diversity-Driven Multi-Parent Evolutionary Algorithm. *Circuits, Systems, and Signal Processing*, *40*(7), pp.3374-3394.