Proactive Virtual Machine Scheduling to Optimize the Energy Consumption of Computational Cloud

Shailesh Saxena[1]
Research Scholar, Department of CS and IT
MJP Rohilkhand University, Bareilly, India

Dr. Mohammad Zubair Khan[2]
Department of CS
Taibah University, Medina, KSA

Dr. Ravendra Singh[3]
Department of CS and IT
MJP Rohilkhand University, Bareilly, India

Abdulfattah Noorwali[4]
Department of Electrical Engineering
Umm Al-Qura University, Makkah, Saudi Arabia

***Abstract*—The rapid expansion of communication and computational technology provides us the opportunity to deal with the bulk nature of dynamic data. The classical computing style is not much effective for such mission-critical data analysis and processing. Therefore, cloud computing is become popular for addressing and dealing with data. Cloud computing involves a large computational and network infrastructure that requires a significant amount of power and generates carbon footprints (CO_2). In this context, we can minimize the cloud's energy consumption by controlling and switching off ideal machines. Therefore, in this paper, we propose a proactive virtual machine (VM) scheduling technique that can deal with frequent migration of VMs and minimize the energy consumption of the cloud using unsupervised learning techniques. The main objective of the proposed work is to reduce the energy consumption of cloud datacenters through effective utilization of cloud resources by predicting the future demand of resources. In this context four different clustering algorithms, namely K-Means, SOM (Self Organizing Map), FCM (Fuzzy C Means), and K-Mediod are used to develop the proposed proactive VM scheduling and find which type of clustering algorithm is best suitable for reducing the energy uses through proactive VM scheduling. This predictive load-aware VM scheduling technique is evaluated and simulated using the Cloud-Sim simulator. In order to demonstrate the effectiveness of the proposed scheduling technique, the workload trace of 29 days released by Google in 2019 is used. The experimental outcomes are summarized in different performance matrices, such as the energy consumed and the average processing time. Finally, by concluding the efforts made, we also suggest future research directions.**

Keywords*—Cloud computing; CO_2; proactive scheduling; unsupervised learning; clustering; energy; prediction; cloud-sim; performance assessment

I. Introduction

“Go Green” is the key theme of the proposed investigation. However, a significant amount of digital data is generated and consumed every day. This demand for computation leads to develop such infrastructure that can deal with such a huge load. One such technology is cloud computing which provides computational resources on demand. The proposed work is keenly focused on offering the techniques that improve the VM workload scheduling to reduce environmental loss and preserve the energy for achieving green computing.

Green is the way of life and it teaches us how to live a sustainable and luxurious life. To support the new generation types of equipment and technologies, we depend on cloud computing to deal with bulky data. Cloud servers are an effective solution because a rich amount of data is generated using these devices and processing such big data. The cloud can scale the computational ability according to demand. On the other hand, to perform computation, we need a huge power supply and cooling system that increases the power consumption and emission of harmful gases. Thus, need to achieve green computing by reducing the power consumption of the computational cloud. In this context, in recent literature [1][2][3], we found VM (virtual machine) workload scheduling can be a good strategy to efficiently utilize the computational resources and reducing power consumption of cloud servers.

The physical machines contain several virtual machines. These VMs are used to deal with the workload that appeared for processing. If we better utilize the resources, we can process a large number of jobs in fewer VMs. Additionally, we can also turn off the ideal machines to reduce power consumption [4]. In this context, the proposed work is motivated to work with VM scheduling techniques to achieve green computing. In recent literature, we identify there are two kinds of VM scheduling approaches active and proactive. The proactive technique is more effective than the active approach due to prior knowledge of VM workload. Thus, this approach can be more beneficial for the proposed investigation.

A. Motivation

VM scheduling provides many benefits in different scenarios of the cloud and relevant technologies. Those facts are validated by using identifying some noteworthy contributions related to the proposed domain of study. Some essential of them are discussed in this section.

Proficient VM management is extremely critical for expanding benefit, energy-saving, and forestalling SLA infringement. VM placement plans can be characterized as reactive and proactive to improve VM arrangement by determining future workloads expectations. M. Masdari et al. [5] advances a review of the proactive VM placement and classifies them. They depict how techniques have been applied to lead viable and low overhead. Also, factors like assessment boundaries, simulators, workload, energy-saving, and predictions are analyzed. Finally, the issues and future opportunities are featured.

Fault-aware scheduling is significant for the cloud and identified with the reception of dynamic workload. R. Kaur et al. [6] proposes a pattern similarity-based scheduling for the cloud. To approve the solution, they performed two investigations with conventional technique and with the fault aware technique. The outcomes show the viability of the plan.

The server burns a tremendous measure of energy to fulfill the expanding need of computational assets. Computing and Cooling are the two frameworks that are enormous energy-devouring. Dynamic VM consolidation is a procedure to lessen energy utilization. Forceful union prompts the production of areas of interest that affects energy utilization. S. Ilager et al. [7] propose an Energy and Thermal-Aware Scheduling (ETAS) that merges VMs to limit energy utilization. The ETAS tends to compromise between time and cost-saving and can be tuned dependent on the prerequisite. They perform tests by utilizing real traces. The outcomes show that ETAS decreasing energy.

In a Virtual Symmetric Multiprocessing (VSMP) climate, the conduct of the scheduler can impact I/O responsiveness. The hinder remapping component can use different virtual CPUs. W. Zhang et al. [8] recognized an "Interfere with capacity Holder Pre-emption" (IHP) issue. The IHP presents that a virtual CPU handicapping the interferes with the capacity of the visitor's organization gadget is de-booked by a scheduler. In this manner, the creator proposes CoINT, a gasket organizer dwelling in the hypervisor, to improve the organization I/O execution. CoINT wipes out the IHP issue and lessens I/O intrudes on delay. They execute CoINT in the KVM hypervisor and assess its proficiency utilizing benchmarks. The outcomes show that CoINT can work on the netperf throughput up to 3x.

Cloud research has needed data on the qualities of the creation of VM workloads. A comprehension of these qualities can illuminate the resource management frameworks. E. Cortez et al. [9] present an interpretation of Microsoft Azure's VM workloads, including VMs' lifetime, size, and resource utilization. Then, show that specific VM is genuinely reliable. In light of this perception, present Resource Central (RC) gathers VM telemetry, learns practices, and gives predictions to administrators. They change the VM scheduler to use predictions in oversubscribing servers. Utilizing real traces show that the aware prediction schedules increment use and actual resource fatigue.

Conventional virtualization frameworks can't viably detach shared micro-architectural assets. Processors and memory-concentrated VMs fighting for assets will prompt genuine execution interference. Y. Cheng et al. [10] propose a contention-aware prediction model on the virtualized multi-core frameworks. Start with recognizing performance interference factors and plan benchmarks to get VM's contention affectability and intensity features. Second, based on the features, fabricate a VM performance forecast model utilizing ML. The model can be utilized to streamline VM execution. The outcomes show that the model accomplishes high accuracy, and the MAE is 2.83%.

B. Objectives

This paper investigates the energy-efficient Cloud resources scheduling for reducing the power utilization of datacenters. We focused on energy-aware VM consolidation schemes to reduce the number of running hosts to preserve energy.

We deal with this issue of selecting a power-efficient configuration that may also be suitable for mapping the client request through available VMs. The framework also considers the selection of optimal pair of the server. To scale the solution, considering client requests as time-varying, the framework enables us to virtualize the system to react to workload variations and adapt relevant configurations. The dynamic configuration improves leveraging predictions about upcoming resource demand and availability. The system takes advantage of the correlation between the historical workload and future workload demands in predictions [11].

An increasing number of datacenters are being deployed to support different applications and services. In a cloud, the applications are hosted on physical servers. These servers have great processing capabilities and can fulfill the performance demands, incurring high energy costs and increasing CO_2 generation and environmental losses [12].

The energy utilization for keep servers running became an important issue, which requires major investigation and immediate steps to improve energy efficiency. According to [13] the datacenters contribute to 30% of the world's CO_2 emissions. Thus, energy-efficient servers are a fundamental concern. To allow hosting multiple independent applications, platforms rely on virtualization to better utilize server resources. Virtualization has been adopted for resource usage efficiency; by VM consolidation and on-demand resource allocation and migration [14]. The dynamic workloads consolidation using migration of VMs, helps to increase the server utilization, reducing the use of resources and power. The ability to move workloads enables PMs to be turned off during low requirements. This offers an efficient way of running a data center in terms of energy saving. However, the data center workload often stays around 30%, and the part of under loaded servers can be as high as 70% [15].

The workloads vary with time, and prior knowledge about future demand may help to handle different critical conditions. So, we proposed a clustering scheme to predict the requests and resource requirements with network and traffic. This method improves the decisional accuracy and guaranteeing high quality of service. The aim is to provide a power aware performance management system in a virtualized environment.

Thus, we proposed a prediction approach based on ML clustering to predict the workload regarding the resource demand of requests. We also incorporate improvements that make the tunable parameters in real-time. Using this technique, we are minimizing the power utilization to meeting the performance expectations. In order to implement, we investigate the use of dynamic configuration techniques, such as voltage scaling, the server on/off switching, and migration. We also collect evidence to verify feasibility and effects through simulations. The solution supports different types of virtualization techniques to enable resource provisioning. The framework has three major components: Data clustering, workload prediction, and power management, as shown in Fig. 1.

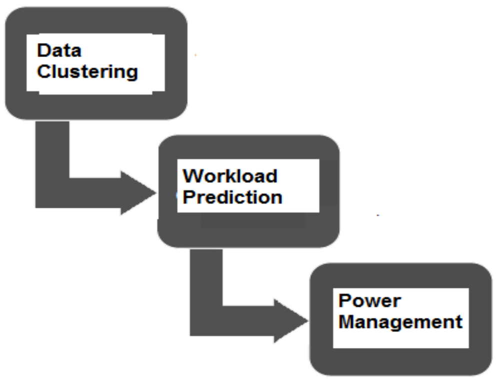


Fig. 1. Life Cycle of Proposed Framework.

1) Data clustering: The approach relies on historical workload dynamics for a time period referred to as the observation window. A VM request consists of multiple cloud resources (e.g., CPU, bandwidth etc.). This poses unique challenges to developing prediction techniques. Also, different clients may request different amount and type of the resource. Therefore, it is difficult to predict the demand for each type of resource. To address this issue of maintaining the personalized QoS requirements, we divide requests into several categories and then apply prediction for each category.

2) Workload prediction: We rely on the traces as training data samples to calculate the weights. One of the problems in implementation is that we will also need to train the model to adjust these weights according to the workload dynamics, which may vary over time. We refer to these predictive models as to be needed training from time to time. We will use a predictive model that increases the accuracy over time and avoids storing large traces. The model predicts a number of requests based on the learned weights. Next, the model observes requests received in each category. The algorithm uses these observations to update the weights. Therefore, we will find the optimal weights. The predictor has an overhead after observing the actual workload to update the weights. These updates increase the accuracy with time and make it dynamic to adopt the latest variations.

3) Power management: The predictions are passed to the Power Management module, which uses this prediction to decide when a PMs go to in ideal state and when become alive. This component is updated with the information of all PMs and maintains their utilizations and states. It uses a heuristic to predict requests and determine how many PMs are alive. The algorithm tries to map requests with live PMs. To do that, it sorts PMs according to the best fit to the least fit. The conventional algorithm could not be used directly. Thus, a modification is made to the algorithm. This limitation has been addressed by mapping these multiple dimensions into a single metric. Furthermore, heuristic considers the energy when sorting the PMs by the following criteria:

a) PMs that are in a working state

b) PMs that have higher utilizations. The utilization is defined as the product of the utilization.

c) PMs that have higher capacities. The capacity is defined as the product of the capacities of an individual resource.

The aim behind sorting conditions is to use the available PMs already being used, so live PMs are ranked first. Then use the utilization metric, since increasing the utilization of the PMs makes the cluster more efficient. Thus model allows for sleeping more PMs. Finally, based on capacities, we can fit more VMs in a PM.

II. LITERATURE REVIEW

In this section we are exploring the recent research work carried out to improve the performance of cloud infrastructure and also contribute for VM scheduling and green computing aspects.

A. Related Work

Cloud Computing (CC) is a multi-tenant framework used by multiple users' concurrently, each exhibiting different and varied behaviour. This heterogeneity shapes highly fluctuating load and creates new usage patterns. VMs interference plays a big part in changes at load. Server load prediction is crucial to ensure efficient resource usage.

B. Literature Summary

The investigation of green computing is being essential for reducing the carbon footprints and for long-term sustainable computing. To understand the effect of green computing in this paper, a survey is performed on green cloud computing techniques. The key focus is paid on proactive techniques of cloud power management, which involve the predictive strategies for regulating the QoS (quality of service) requirements of applications. According to the collected literature, we found that the VM consolidation by improving the scheduling techniques can significantly preserve datacenters power demands. Additionally, proactive methodologies improve the performance on step ahead. In this context, we also summarize the recent VM scheduling approaches in Table I. These tables include the summary of recent contributions in terms of research work for achieving power efficiency. This summary includes the solutions developed for improve the virtual machine scheduling performance to reduce brown energy and those algorithms which are utilized to derive the required power optimization technique. There are two main key findings as a conclusion of this survey. First, the prediction based VM management in the cloud for consolidating the scattered future requests in multiple machines can reduce the power demand of the datacenters in effective manners. And second, the available solutions have a lack of adjustment about the uncertainty of users' demand.

So here the proposed proactive VM scheduling focused on prediction using unsupervised learning algorithms to handle the uncertain future demands of users to maintain the effective consumption of power in cloud datacenters.

TABLE I. REVIEW SUMMARY

Ref. No	Direction	Solution	Significance
[16]	Virtual Machines (VMs) interference	A real-time server load prediction system based on incoming task classification and VM interference.	An improved HAT, with ensemble drift detectors. Able to dealing with changes and good accuracy with less time and memory
[17]	Concerned with balancing servers load	Utilize resource and job response time, enhance scalability, and user satisfaction.	Dynamic resource allocation using EANN to predict VM load and get better performance.
[18]	Prediction of server load	Improving the resource utilization, reducing energy consumption and guaranteeing the QoS.	Integrates the cloud and Markov chain to realize CM-MC algorithm. Results show that algorithm is high accuracy and reduce energy consumption.
[19]	Host load prediction	Proposed a host load prediction method based on Bidirectional Long Short-Term Memory (BiLSTM)	BiLSTM improve the memory capability and nonlinear modeling. 1-month trace of 12K machines is used to validate accuracy.
[20]	Focusing on the issue of host load estimating in mobile cloud	LSTM, for the intricate and long-term arrangement of the cloud condition and dependent on GSO LSTM neural system.	Cloud load forecasting model using LSTM, and GSO to search optimal parameters. And compared with similar algorithms. Results show that the algorithms are offers higher accuracy.
[21]	Performance of dynamic clouds depends on efficiency of its load balancing and resource allocation	A study on the predictive approach for dynamic resource distribution. And a rule-based workload-balancing based on predictions.	Simulation of cloud using CloudSim and used an algorithm with lower computational demand and a faster balancing. The result will show a predictive workload balancing is an effective solution.
[22]	Cloud DBMS load balancing	Distribution of transactions between replicas, load balancing to improve the resources utilization.	A predictive load balancing service for replicated cloud databases.
[23]	a predictive and elastic load balancing service for replicated cloud databases	The distribution of transactions among replicas, load balancing can improve the resources utilization.	Showed that the use of prediction can help to predict possible SLA violations that represent workloads of cloud-replicated databases.
[28]	Datacenter consuming a high amount of energy and increasing carbon emissions.	VM need to be allocated to minimize resource and energy wastage.	Solution for datacenters named E-FPA. Aim is to provide energy-oriented VM allocation using Switching Probability. It enhances energy consumption, Migration, and First Fit Decreasing.
[29]	Green Computing to reduce, power and water consumption, hardware and carbon emission.	Presents an analysis report about green computing and its characteristics.	Discusses about green computing, trending concepts and future challenges. This analysis helps researchers to understand green cloud.
[30]	Economic and environmental costs of data centre and equipped for peak processing.	Idle servers and components of the network consume a significant amount of resources.	Describe the green data centre and metrics. And discuss energy-saving solutions for servers, network, and other green solutions.
[31]	Efficient VM management for energy saving, increasing profit, and preventing SLA violations.	This survey on the proactive VM placement approaches and categorizes them.	Factors such as evaluation parameters, simulation software, workload, power management, and prediction are compared to illuminate VM placement. Issues and future studies are provided.
[32]	VoIP have face a crisis: 1. Lack of resources and, overload; 2. Redundancy of resources and, energy loss.	The SDN can provide a view of the network for resource management. NFV can be used to implement a variety of devices and functions.	GreenVoIP to manage the resources of cloud VoIP centers. It not only prevents overload but also supports green computing. That framework can minimize active devices.
[33]	Communication traffic have imposed a heavy burden on data centers and resulted in high energy consumption.	Edge computing is explored to provision the latency-sensitive applications.	Geo-distribution of edge devices is leverage green computing. It is desirable to maximize utilization of green energy. Investigate cost minimization problem of VM migration, task allocation and scheduling using heuristic algorithm.
[34]	Provisioning Edge computing QoS mainly delay guarantee for delay-sensitive applications. energy consumption in edge servers may be higher	Energy-efficient and delay-guaranteed workload allocation problem in an IoT-edge-cloud computing system are investigated	Offloading workloads to servers, and computation experience, e.g., delay and energy consumption. Delay is un-negligible in a intensive environment. The workload allocations among local, neighbour, minimal energy and delay using algorithm.
[35]	Promote ecosystem services, including mitigation of storm water flooding and water quality degradation	Goals include increasing carbon sequestration, songbird habitat, reducing urban heat effects, and improving the landscape.	GI is improving water and the ecosystem by reducing storm water runoff. Provide design to enable better communication among designers and groups. Demonstrates workflows to facilitate the creation of GI, incorporated models using web applications.
[36]	VM placement	Vector Bin-Packing (VBP) problem to minimize the number of PMs used	FFD variant, Aggregated Rank in FFD is proposed. Experiments using two datasets: based on Amazon and a synthetic dataset. The efficiency of FFD-AR is better.

III. Proposed Proactive VM Scheduling

This section explains the proposed proactive virtual machine scheduling technique for optimizing the cloud datacenter performance in terms of power consumption. In this context, this section provides a discussion about the different algorithms and the data analytics steps to explain the work of the required model.

A. System Overview

The increasing demand for computational and storage resources motivates us to design sustainable and prolong computing technologies. In this direction, one of the most crucial development is green cloud computing. Green computing technology is providing a way to minimize the cost and consumption of infrastructural assets. Therefore, green cloud computing has become one of the most essential and trending areas of research and development. The key idea is to employ various techniques to switch off the additional computational and network devices that are unused or in an ideal state to preserve the power consumption. Additionally, when the load again appeared on datacenters, machines became alive to deal with the workload.

In this context, we are proposing an unsupervised learning technique for designing the physical machine consolidation scheme. The unsupervised learning techniques are suitable for time-critical applications and also for parameter enhancement and optimization. Therefore, to improve the capability of power management the clustering algorithms are adopted. Therefore, first implemented four popular clustering approaches (namely K-Means, SOM (Self Organizing Map), FCM (Fuzzy C-Means), and K-Medoid) that categorize the cloud datacenters workload into three categorize, i.e., low, medium, and high workload groups.

Further, these trained machine learning models are being used for efficient VM allocation and making power on and off decisions based on the predicted outcomes. In this context, a modified VM scheduling algorithm has been proposed. That makes use of one step ahead predicted workload and evaluates the current wearability of data center resources. Based on this decision function, we decide the required resources which fulfill the demand. In this way, we are reducing the migration of processes and maintain less traffic overhead. In addition, by using the predictive method, we estimate the future possible resource demand. Suppose demand is higher than the available one step ahead workload. In that case, we restart the physical machines, or if the future trend shows a low demand, we turn off the machines to utilize the minimum resources.

B. Clustering of Workload

To utilize the proactive VM scheduling for improving the performance of cloud datacenters in terms of energy efficiency, we are intended to use the unsupervised learning techniques for predictions.

In this context, Fig. 2 demonstrates the model for learning with the historical workload patterns. Therefore, the Google workload trace dataset is being used for performing the clustering. The dataset consists of parent ID, Task ID, job type, nrmlTaskCores, nrmlTaskMem, and time. Among them, parent ID and Task ID provide the identification of the process. Thus, we remove one of them, so here we eliminate parent ID for the dataset. The remaining attributes are utilized further for clustering of the tasks. To create clusters, we need to provide a number of clusters here; use k=3 for preparing clusters of three types, to categorize and belong in low, medium and high groups.

Further, we have implemented a provision that will be used to select the clustering algorithm for training. The algorithms are discussed in the previous section. The selected algorithm is used with the trace file and clusters the data. The employed clustering algorithms result in two key outcomes, i.e., validation performance of the clustering algorithm and the centroids identified from the data. These centroids are further being used for predicting the trends of the upcoming workload.

C. Predictive VM Scheduling

The VM scheduling in energy efficient manner need to include the following phases:

1) Predicting the future workload: To demonstrate the effective VM scheduling and consolidation technique, using the proactive manner. First of all, we need to have workload trends. Therefore, to prepare future trends for each one-minute interval, we have trained machine learning algorithms. These algorithms use the cluster centroids and simple linear regression technique to compute the future trends of the host workload. In this context, let we have a set of centroids (C) such that:

$$C = [C_L, C_M, C_H] \tag{1}$$

Where, C_L = centroid for small workload, C_M = centroid for medium workload, C_H = High workload.

Each centroid C_L, C_M, *and* C_H consists of memory and CPU properties of demand. Thus, when talking about the centroids, then we also considering each element of these subsets or centroids. Now, we need to predict a pattern relevant to the upcoming future workload. In this context, first, consider one hour for predicting the workload values. Thus, we extract the last one hour per minute values and keep them in a list L_N where N are minutes.

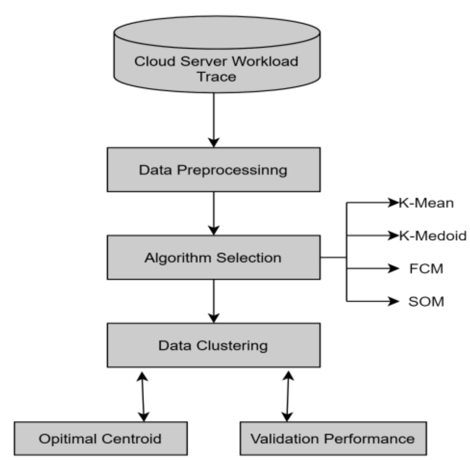


Fig. 2. Workload Clustering.

Algorithm 1: Proposed load Encoding

Input: last recent workload list L_N, list of centroids

$C = [C_L, C_M, C_H]$

Output: Encoded list E_n

Process:

1. *for (i = 1; i < n; i++)*
    1) *temp = L_i*
    2) *for (j = 1; j < m; j++)*
        I. *tempC = C_j*
        II. *d_j = temp- tempC*
    3) *end for*
    4) *S = getShortest(d)*
    5) *e = getlabel(S)*
    6) *E.Add(e)*
2. *End for*

Let considering the one-hour forecast, then N=60. Now we encode the values available in the list with three symbols L, M, and H. The encoding process is demonstrated in algorithm 1. According to the given algorithm 1, we accept a list of likelihood workload list L_N and prepared centroids $C = [C_L, C_M, C_H]$. The algorithm processes both the inputs and prepares a list of encoded patterns of the list. The algorithm extracted each element of the list L_N and compared it with all the centroids. The most minimum distance centroid is labelled as the encoding value of L_N. After assigning a label to the load value, we get a new list E of encoded n values in terms of L, M and H.

The next step proposes to utilize the Encoded list E_N as the initial prediction of the next one-hour workload pattern. Now we need to enhance the prediction. Therefore first, we calculate or difference between the actual extracted value as well as encoded values using the following equation:

$$D_N = L_N - E_N \quad (2)$$

Where D_N is the difference between the last observed value and reference or actual value.

2) VM consolidation: This section explains how the prediction is used in real-time to predict and reduce power utilization by on and off the physical machines. Therefore, let at actual time t_a, we have a basic prediction taken from the list E_N for the given time t_a. That workload is denoted here as P_{t_a}. Additionally, at that time, we also get the actual workload appeared and denoted as W_a. Further, at the next time movement $\boldsymbol{t_b}$, we need to adjust the initial prediction. Thus, the new predicted value is:

$$P_{t_b} = E_{t_b} + E_r \quad (3)$$

where,

$$E_r = W_a - P_{t_a} \quad (4)$$

Here, E_r it can be positive or negative. E_{t_b} is the basic prediction at the time t_b. Further, we make use of a difference list D_N to impure the prediction.

$$F = D_{t_a} - E_r \quad (5)$$

The measured different F is demonstrating the influence in prediction error; therefore, the system needs to adjust this for making the final prediction by restructuring the prediction equation using the following equation:

$$P_{t_b} = E_{t_b} + E_r + F \quad (6)$$

Now, the system utilizing the predicted resource demand and the available resources of the datacenter to decide to switch ON and OFF, the physical machine. In this context, we prepare an algorithm for making decisions regarding the same, steps are described as algorithm 2.

Algorithm 2: Predictive VM Scheduling Technique

Input: list of datacenters host list H_m, P predicted workload, T consolidated resource available

Process:

1. *For each next time event*
2. *if (P < T * 0.66)*
    a. *S = m * 0.2*
    b. *For (i = 1; i < m; i++)*
        1) *if (m < S)*
            1. *MigrateProcessofHost(H_i)*
            2. *HID = H_i. Sleep*
        2) *End if*
    c. *end for*
3. *Else*
    1) *HID .Start(Top,1)*
    2) *H_m.Add (HID)*
4. *End if*
5. *End for*

According to the discussed algorithm 2, we have the host lists H_m with m elements; on the other hand, we compute the total consolidated resources available as T. the algorithm starts with the analysis of each time increment. To compute the predicted demand of resources P. if the P is less than 66% of available resources in the consolidated resource T, we compute the capacity of 20% of machines available, which the existing resources will handle; thus, we turn off one physical machine at a time. These two thresholds depend on the designer and the application's quality of service requirements. Otherwise, each time we turn on one physical machine available in the sleep list of the host.

IV. Results Analysis

This section provides the experimental analysis of the proposed VM scheduling approach. Thus, first the clustering algorithms are applied on the dataset. The predictive performance of ML techniques is measured. Additionally, the performance of cloud infrastructure has also been evaluated and simulated in this section.

A. Experimental Scenarios

The experimental simulation of proposed proactive VM scheduling is done through Cloud-Sim simulator with a datacenter of 5 servers [Host1, Host2, Host3, Host4, Host5] having 20, 5,10 20, and 10 VMs respectively. The computational setup of the simulated datacenter can be elaborate through Table II.

TABLE II. SIMULATION SETUP OF PROPOSED PROACTIVE VM SCHEDULING

Datacenters	**Host1**	**Host2**	**Host3**	**Host4**	**Host5**
CPU per VM (MIPS)	250	250	250	500	250
RAM per VM (MB)	2048	2048	2048	2048	2048
Total Storage (GB)	1M	1M	1M	1M	1M
Bandwidth (Gbits/sec)	1000	1000	1000	1000	1000
VMs	20	5	10	20	10

The VM scheduling ensures the optimal VM resource utilization and reduction in power consumption. The aim is to investigate the proactive VM scheduling techniques using ML techniques to understand the demand of resources and prepare the plan to better schedule the cloud VM for reducing the power consumption and enhancing the running cost of the cloud. Therefore, first, some unsupervised learning approaches are compared to identify the best-performing clustering approach. Further, an algorithm will be proposed to design a proactive resource scheduling technique to achieve Green Computing, reduce power consumption, and reduce carbon emission. Thus, the given experiment includes two experimental scenarios:

1) Comparing different clustering algorithms: Clustering is an unsupervised learning task. It automatically discovers grouping the data. Clustering only interpret the data and find natural groups. In this scenario, we compare four clustering algorithms, namely, K-Means, SOM (Self Organizing Map), FCM (Fuzzy C Means), and K-Mediod over different performance parameters and different data sizes.

2) Comparing simulation performance of cloud for VM scheduling: This simulation scenario implements proactive VM scheduling techniques using the above-discussed algorithms. Thus, using all the algorithms, the prepared VM Scheduling techniques are compared to investigate energy consumption.

B. Experiments Comparing Clustering Algorithms

In this section, a comparative performance study is carried out among different unsupervised learning algorithms. The aim is to obtain an efficient and accurate algorithm for predicting the accurate future VM resource demand and can optimize the scheduling to better resource allocation. The comparative outcomes of the algorithm's performance are reported in this section.

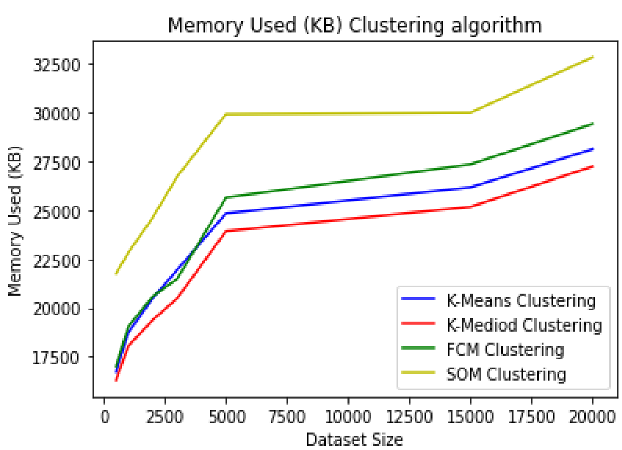
We are investigating the efficiency of the algorithms, so it needs to compute the models' performance in terms of memory and time required to process the workload of different sizes. The memory usage of the algorithms is measured based on java function and using the following formula:

$$M_{Used} = M_{Assigned} - M_{Free} \quad (7)$$

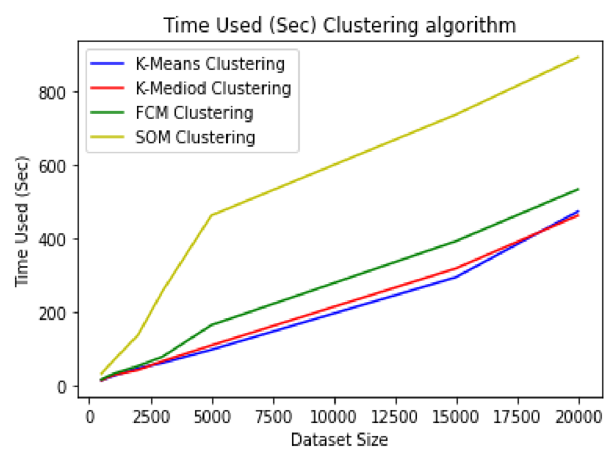
Similarly, the time consumption of the implemented algorithms is given as the amount of time consumed for processing the data using the algorithms is given as the time consumption of the algorithm. That can be calculated using the following formula:

$$Time_{Consumed} = Time_{End} - Time_{Start} \quad (8)$$

The memory and time consumption of the implemented clustering algorithms are described using Fig. 3(A) and (B). Additionally, the observed values are given in Table III. In both the line graphs, the X-axis demonstrate the size of experimental dataset size in terms of instances, additionally, in Fig. 3(A) the Y-axis shows the memory used in terms of kilobytes (KB) and in Fig. 3(B), the Y-axis shows the time consumed in processing the data in terms of seconds (Sec). According to the obtained performance, the SOM and FCM are producing a higher amount of time and memory overhead during the computation. Therefore, these algorithms are also producing a significant delay during the cloud VM allocation. Therefore, according to all the obtained performance parameters, we can say the K-Means and K-Mediod are time and memory efficient. Thus, we recommend being using the K-Means and K-Mediod algorithms to reduce the delay in scheduling.



(A)



(B)

Fig. 3. Comparative Performance of Clustering Algorithms in (A) Memory used (B) Time Consumed.

TABLE III. PERFORMANCE OF CLUSTERING ALGORITHMS IN MEMORY USAGE AND TIME CONSUMED

Dataset Size	Memory Usage				Time Consumed			
	K-Mean	K-Mediod	FCM	SOM	K-Mean	K-Mediod	FCM	SOM
500	16712	16279	16992	21746	14	14	16	32
1000	18718	18028	19028	22817	26	28	32	68
2000	20484	19372	20582	24617	47	42	53	137
3000	21942	20478	21477	26725	61	66	78	256
5000	24837	23927	25649	29927	97	109	164	462
15000	26173	25173	27352	30002	294	318	392	736
20000	28134	27248	29428	32846	473	462	533	892

C. Proactive VM Scheduling for Green Computing

The VM scheduling by proactive techniques are demonstrated using different clustering algorithms, namely K-Mean, K-Mediod, FCM (Fuzzy C Mean), and SOM (Self Organizing Map). The aim is to find the impact on scheduling using these ML techniques over different parameters. The simulation consists of two major goals:

a) Compare the Impact of four different clustering based proactive scheduling over average processing time (Pt AVG).

b) Compare the impact of four different clustering based proactive scheduling over efficient power consumption (PC).

1) Average processing time: The process time is the amount of time for which a central processing unit (CPU) was used for processing instructions. The processing time is a combination of the total time a process resides in the processor and the time required to wait for the resource. Processing time (Pt) is generally calculated in MS (Milliseconds).

$$Pt_{AvG} = \frac{1}{N}\sum_{i=1}^{N} Pt_i \quad (9)$$

Where, N number of processes to be scheduled, and Pt_i is the processing time of the i^{th} process.

Table IV have the simulation results of average processing time for proposed proactive VM scheduling based on four different clustering to evaluate which one take minimum time in VM scheduling.

The comparative evaluation of four different clustering based proactive VM scheduling using average processing time (Pt_{AvG}) is also shown in Fig. 4. That is a bar graph that is prepared using the observations collected from the simulation. The processing time of the implemented simulation is measured here in terms of Minutes. In order to show the performance, the X-axis contains the number of VMs, and in the Y axis, the average processing time in minutes is reported. According to the system's performance, K-Mediod clustering algorithm requires less processing time than other similar algorithms.

2) Power consumption: The entire VM scheduling optimization techniques are works to enhance the profitability of cloud service providers. In this context, the low energy or power consumption is tried to reduce by optimal resource allocation of the cloud. Table V is used to show the simulation results of power usage during proactive VM scheduling based on four different clustering to evaluate which one gives the efficient usage of power during VM scheduling.

TABLE IV. SIMULATION RESULT OF AVG. PROCESSING TIME

Dataset Size	K-Mean	K-Mediod	FCM	SOM
1000	10.2	9.3	11.2	9.3
2500	21.5	18.7	23.5	20.9
5000	43.3	39.1	46.3	29.5
7500	60.4	55.9	68.4	56.8
10000	80.3	73.2	90.3	76.3
15000	115.9	101.2	126.9	108.1
20000	173.3	158.8	182.3	163.2

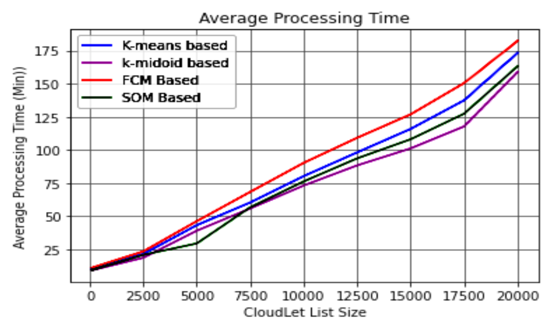


Fig. 4. Performance of Proactive VM Scheduling's on Average Processing Time.

TABLE V. SIMULATION RESULTS OF POWER CONSUMPTION

Dataset Size	K-Mean	K-Mediod	FCM	SOM
1000	60.3	54.8	58.9	53.8
2500	98.3	108.2	94.1	86.6
5000	124.5	128.7	119.2	109.5
7500	160.3	155.5	153.1	136.9
10000	198.3	201.3	183.7	166.3
15000	225.9	241.2	216.4	201.9
20000	263.3	266.2	252.3	245.4

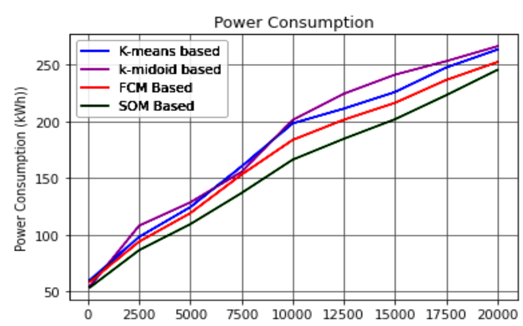


Fig. 5. Performance of Proactive VM Scheduling's on Power Consumption.

The Power Consumption is measured here in terms of KW/h. The recorded power consumption for all clustering based proactive VM scheduling's is shown in Fig. 5. To show the performance, the X axis shows the number of VMs, and the Y-axis shows the cloud's power consumption. According to the performance of the system, the FCM and SOM-based models report the lowest consumption.

V. Conclusion

The proposed work is implementing a simulation-based experimental study for designing an effective and efficient proactive VM scheduling approach to minimize the energy consumption of the cloud infrastructure. The scheduling techniques are essential for allocating the resources to the users' request to limit the unwanted resource utilization of cloud datacenters. That also increases the energy consumption, maintenance cost, and running cost of the datacenters. Therefore, we need some effective resource allocation strategy that can deal with the dynamic nature of workload. In addition, by implementing the host switch ON and OFF technique, we can also control the power consumption and running cost of datacenters.

Therefore, the proposed work is motivated to design and develop a proactive VM scheduling model which utilizes four different unsupervised learning techniques to predict the future demand of the cloud. Additionally, based on the predicted demand trend, the proposed algorithm plans to keep switching OFF the resources when the model predicts the less resource demand trend. In this way, by replacing, we prepared four resource scheduling techniques by using the reported clustering algorithms. Based on the experimental analysis and obtained findings, we can see the model successfully preserves the energy by making inactive hosts in datacenters to minimize energy consumption. Additionally, the obtained performance of the models also justifies our argument for the proposed proactive VM scheduling model.

VI. Future Scope

In near future we are motivated to perform more extensive experiments on the given resource provisioning model, additionally we also compare the technique with the classical resource scheduling model to find how the proposed technique can improve the productivity of the model. Finally, in near future we also work for involving the deep learning models for making the accurate future resource demand prediction.

References

[1] J. M. T. I. Jayalath, E. J. A. P. C. Chathumali, K. R.M . Kothalawala, N. Kuruwitaarachchi, “Green Cloud Computing: A Review on Adoption of Green-Computing attributes and Vendor-Specific Implementations”, Smart Computing and Systems Engineering, 2019, Paper No: SC 24.

[2] Q. Zhang, X. Lin, Y. Hao, J. Cao, “Energy-Aware Scheduling in Edge Computing Based on Energy Internet”, IEEE access VOLUME 8, 2020.

[3] N. Gholipour, E. Arianyan, R. k. Buyya, “A novel energy-aware resource management technique using joint VM and container consolidation approach for green computing in cloud data centers”, Simulation Modelling Practice and Theory 104 (2020) 102127.

[4] M. A. Alsaih, R. Latip, A. Abdullah, S. K. Subramaniam, K. A, Alezabi, “Dynamic Job Scheduling Strategy Using Jobs Characteristics in Cloud Computing”, Symmetry 2020, 12, 1638; doi:10.3390/sym12101638.

[5] M. Masdari, M. Zangakani, “Green Cloud Computing Using Proactive Virtual Machine Placement: Challenges and Issues”, J Grid Computing, https://doi.org/10.1007/s10723-019-09489-9.

[6] R. Kaur, G. Kaur, “Proactive Scheduling in Cloud Computing”, Bulletin of Electrical Engineering and Informatics ISSN: 2302-9285 Vol. 6, No. 2, June 2017, pp. 174~180, DOI: 10.11591/eei.v6i2.649.

[7] S. Ilager, K. Ramamohanarao, R. K. Buyya, “ETAS: Energy and thermal-aware dynamic virtual machine consolidation in cloud data center with proactive hotspot mitigation”, Concurrency Computat Pract Exper. 2019;31:5221. 2019 John Wiley & Sons, Ltd. 1 of 15.

[8] W. Zhang, X. Hu, J. Li, H. Guan, “CoINT: Proactive Coordinator for Avoiding Interruptability Holder Preemption Problem in VSMP Environment”, IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, 978-1-5386-4128-6/18/$31.00.

[9] E. Cortez, A. Bonde, A. Muzio, M. Russinovich, M. Fontoura, R. Bianchini, “Resource Central: Understanding and Predicting Workloads for Improved Resource Management in Large Cloud Platforms”, SOSP'17, October 28, 2017, Shanghai, China © 2017 Association for Computing Machinery, ACM ISBN 978-1-4503-5085-3/17/10.

[10] Y. Cheng, W. Chen, Z. Wang, Y. Xiang, “Precise contention-aware performance prediction on virtualized multi-core system”, Journal of Systems Architecture 72 (2017) 42–50.

[11] L. He, D. Zou, Z. Zhang, K. Yang, H. Jin, S. Jarvis, ”Optimizing Resource Consumption in Clouds”, in Proc. of the 12th IEEE/ACM International Conference on Grid Computing(Grid 2011), 2011, PP.42-49.

[12] Maarten van Steen, Andrew S. Tanenbaum, “A brief introduction to distributed systems”, Computing (2016) 98:967–1009, DOI 10.1007/s00607-016-0508-7.

[13] Auday Al-Dulaimy, Wassim Itani, Ahmed Zekri1, and Rached Zantout, “Power management in virtualized datacenters: state of the art”, Journal of Cloud Computing: Advances, Systems and Applications (2016) 5:6 DOI 10.1186/s13677-016-0055-y.

[14] Leila Helali, Mohamed Nazih Omri, “A survey of data center consolidation in cloud computing systems”, Computer Science Review 39 (2021) 100366.

[15] Dzmitry Kliazovich, Pascal Bouvry, Samee Ullah Khan, “DENS: data center energy-efficient network-aware scheduling”, Cluster Comput, DOI 10.1007/s10586-011-0177-4.

[16] H. Toumi, Z. Brahmi, M. M. Gammoudi, “Real time server load prediction system for the ever-changingcloud computing environment”, ScienceDirect – 2019.

[17] K. S. Qaddoum , N. N. El Emam , M. A Abualhaj, “Elastic neural network method for load prediction in cloud computing grid”, International Journal of Electrical and Computer Engineering (IJECE)- April 2019.

[18] X. Xiaolong, Z. Qitong, M. Yiqi, L. Xinyuan, “Server load prediction algorithm based on CM-MC for cloud systems”, IEEE - Journal of Systems Engineering and Electronics Vol. 29, No. 5, October 2018.

[19] H. Shen, X. Hong, “Host Load Prediction with Bi-directional Long Short-Term Memory in Cloud Computing”, arXiv – 2020.

[20] P. Sudhakaran, S. Swaminathan, D. Yuvaraj, S. S. Priya, “Load Predicting Model of Mobile Cloud Computing Based on Glowworm Swarm Optimization LSTM Network”. IJIM – 2020.

[21] M. Jodayreea, M. Abazab, Q. Tan, “A Predictive Workload Balancing Algorithm in Cloud Services”, ScienceDirect -2019.

[22] C. S. S. Marinho, E. F. Coutinho, J. S. C. Filho , L. O. Moreira, F. R. C. Sousa, J. C. Machado, “A Predictive Load Balancing Service for Cloud-Replicated Databases”, IEEE – 2017.

[23] C. S. S. Marinho, L. O. Moreira, E. F. Coutinho, J. S. C. Filho, F. R. C. Sousa, J. C. Machado, “LABAREDA: A Predictive and Elastic Load Balancing Service for Cloud-Replicated Databases”, Journal of Information and Data Management, Vol. 9, No. 1, June 2018.

[24] H. Moradi, W. Wang, A. Fernandez, D. Zhu, “uPredict: A User-Level Profiler-Based Predictive Framework for Single VM Applications in Multi-Tenant Clouds”, arXiv – 2019.

[25] S. Mittal, Prof. M. Dubey, “AMO Based Load Balancing Approach in Cloud Computing”, IOSR Journal of Computer Engineering (IOSR-JCE) – 2017.

[26] S. Manaseer, M. Alzghoul, M. Mohmad, “An Advanced Algorithm for Load Balancing in Cloud Computing using MEMA Technique”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) January 2019.

[27] X. Zhou, F. Lin, L. Yang, J. Nie, Q. Tan, W. Zeng, N. Zhang, “Load balancing prediction method of cloud storage based on analytic hierarchy process and hybrid hierarchical genetic algorithm”, SpringerPlus (2016).

[28] U. M. Joda, A. S. Ismail, H. Chizari, G. A. Salaam, A. M. Usman, A. Y. Gital, O. Kaiwartya, A. Aliyu, “Virtualization oriented Green Computing in Cloud Datacenter: Flower Pollination Approach”, Springer – 2018.

[29] A. Patil, R. Patil, “An Analysis Report on Green Cloud Computing Current Trends and Future Research Challenges”. Elsevier SSRN – 2019.

[30] P. A. Naidu, P. Chadha, V Nalina, “Efficient Strategies for Green Cloud Computing”, Journal of Network Communications and Emerging Technologies (JNCET) Volume 10, Issue 6, June (2020).

[31] M. Masdari, M. Zangakani, “Green Cloud Computing Using Proactive Virtual Machine Placement: Challenges and Issues”, Springer Nature B.V. 2019.

[32] A. Montazerolghaem, M. H. Yaghmaee, A. L. Garcia, “Green Cloud Multimedia Networking: NFV/SDN based Energy-efficient Resource Allocation”, IEEE Transactions on Green Communications and Networking – 2017.

[33] L. Gu, J. Cai, D. Zeng, Y. Zhang, H. Jin, W. Dai, “Energy efficient task allocation and energy scheduling in green energy powered edge computing”, Elsevire – 2019.

[34] M. Guo, L. Li, Q. Guan, “Energy-Efficient and Delay-Guaranteed Workload Allocation in IoT-Edge-Cloud Computing Systems”, 2019 IEEE.

[35] L. Leonard, B. Miles, B. Heidari, L. Lin, A. M. Castronova, B. Minsker, J. Lee, C. Scaifed, L. E. Band, “Development of a participatory Green Infrastructure design, visualization and evaluation system in a cloud supported jupyter notebook computing environment”, 2018 Elsevier Ltd.

[36] S. Jangiti, E. S. Ram, V. S. S. Sriram, “Aggregated Rank in First-Fit-Decreasing for Green Cloud Computing”, Springer Nature Singapore Pte Ltd. 2019.