AI-DRIVEN DYNAMIC VM CONSOLIDATION AND RENEWABLE-AWARE SCHEDULING FOR REDUCING CARBON FOOTPRINT IN CLOUD DATA CENTERS

Satyendra Kumar Pal[1]*, Dr. Vikas Kumar[2]

[1]Ph.D. Scholar, Department of Computer Scinece & Engineering, Chhatrapati Shivaji Maharaj University, Panvel Navi Mumbai, Maharashtra 410221, India

*Corresponding Email Id: profsatyen@gmail.com

[2]Professor & Head, Department Computer Science & Information Technology, Chhatrapati Shivaji Maharaj University, Panvel Navi Mumbai, Maharashtra 410221, India

Email Id: vikaskumar@csmu.ac.in

Abstract

The surging energy requirements and the impact on the environment of the cloud data center suggests the necessity of establishing intelligent resource management plans to facilitate sustainability. The proposed hybrid framework which incorporates hybrid dynamic virtual machine (VM) consolidation based on the idea of artificial intelligence (AI) with the concept of renewable-aware differences in this paper to decrease energy and carbon emission in cloud computing environment. The consolidation module employs reinforcement learning to dynamically locate and move VMs based on real time workload and server utilization patterns. Concurrently, a renewable-sensitive scheduler predicts availability of solar energy and wind energy and executes tasks around green energy highlights. The framework is tested with the help of CloudSim Plus, actual Google Cluster traces, and synthetic renewable traces. The findings indicated that there was a decrease of energy of 37 percent and carbon emission decrease of 46.7 percent over the traditional models with a high SLA compliance along with minimizing the VM migration overhead. The given solution has a great potential to become an energy-efficient and environmentally-friendly cloud infrastructure that does not worsen performances.

Keywords: *Cloud Computing, Virtual Machine Consolidation, Renewable-Aware Scheduling, Reinforcement Learning, Energy Efficiency, Carbon Footprint Reduction.*

INTRODUCTION

Cloud computing is at the heart of current digital transformation, providing a wide range of services, from storing data to running AI and business apps. As cloud-based services and data-heavy workloads have grown quickly, data centers have become energy-hungry facilities that use a lot of electricity and release a lot of carbon (Shuja et al., 2016; Wang et al., 2023). Even while traditional cloud designs can grow and have a lot of resources, they are frequently controlled with static, threshold-based methods for providing resources and combining VMs.

These approaches don't adapt well to changing workloads and different energy needs, which means they use energy less efficiently and have a bigger effect on the environment (Ashraf & Porres, 2017; Arshad et al., 2022). Many people agree that virtual machine (VM) consolidation is a good way to use less power in the cloud. But traditional heuristic-based methods typically don't take into account the environment, including the fact that renewable energy is only available at certain times and places (Beloglazov & Buyya, 2012; Sayadnavard et al., 2022). As cloud service providers come under more and more pressure to meet sustainability standards, the need for smart, energy-aware orchestration methods is expanding. Artificial intelligence (AI), especially reinforcement learning and predictive modeling, is a promising way to make these kinds of systems work (Xu et al., 2021; Ade, 2024).

In the last few years, AI has been researched in regards to the utilization of job scheduling and energy optimization on cloud platforms, where adaptive and context aware solutions can be demonstrated. Figure 1 provides a taxonomy of AI-enabled scheduling approaches, which explore machine learning, optimization or a combination of methods in the context of resource management (Sanjalawe et al. 2025) and identifies the aspects of adaptability and applicability when classified. Notably, models that utilize AI approaches have achieved greater efficiency, but have not been synchronized to renewable energy supply models, which are critical for sustainable operation of data centers (Li et al., 2022; Perin et al., 2021).
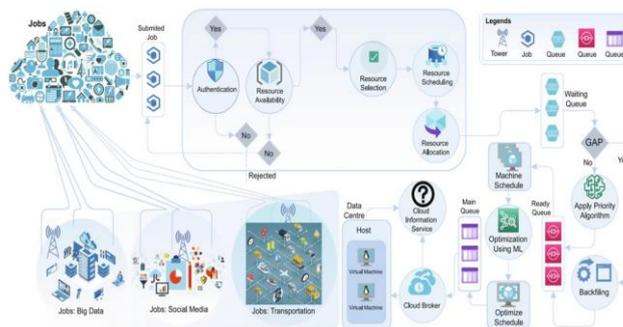

Figure 1. Taxonomy of AI-Driven Job Scheduling Techniques in Cloud Computing [16]

This diagram defines three AI-based scheduling strategies; Machine Learning, Optimization, and Hybrid. It describes their suitability in energy-efficient job scheduling and SlA optimization as well as context-sensitivity in dynamic cloud computing environments.

In addition, renewable-aware scheduling has provided an alternative solution, in which a workload is mapped to the windows of green energy including solar or wind peak periods (Saxena & Singh, 2022; Karmakar et al., 2022). Nonetheless, there will be unforeseen deviations of renewable sources about dynamic work load requirements that demands predictive intelligence and on-line responsiveness that cannot be permanently attained using rule-based systems (Tolia et al., 2008; Meisner et al., 2009). This further strengthens the dependency to have a holistic approach to put together intelligent VM deployment and energy aware scheduling together in a singular platform.

In this paper, we suggest the idea of a hybrid system that integrates AI-based VM consolidation with the renewable-aware workload that aims to minimize the total energy expenditure and carbon emissions in cloud data facilities. It uses reinforcement learning to optimally assign VMs in real time, and it uses forecasting methodologies to project the supply of renewable availability and schedule accordingly. The performance of the system is benchmarked with real world workloads using data sets of the Google Cluster and with simulated solar/wind energy profiles. The data of experiments prove great savings of energy consumption and emissions, increased SLA compliance and decrease of unnecessary migration (Barroso & Hlzle, 2007; Li et al., 2022). The rest of the paper is structured as follows: Section 2 contains a literature review of the VM solidification and energy-aware scheduling. Section 3 outlines the proposed methodology (architecture, algorithms and experimental design). Section 4 talks about the simulation and results analysis. Section 5 is a conclusion of the study and a direction of future research.

LITERATURE REVIEW

Introduction to VM Consolidation and Energy Efficiency

As the number of digital services and virtualized environments are increasing exponentially, data centers emerge as one of the most energy-hungry parts of the contemporary computing infrastructure. Arroba et al. (2023) explain that energy costs of the cloud data centers represent a large percentage of the overall power consumption of IT worldwide, which led environmental and operating concerns. As a counter measure, VMs consolidation has become a popular solution that focuses on server utilisation to ensure that underutilised physical servers transfer their VMs, thereby reducing the number of active servers (Sayadnavard et al., 2022). Nevertheless, conventional methods of consolidation like the simple static threshold heuristics have been ineffective to accommodate highly dynamic work loads. They can cause performance degradation casing violation of SLA and adding migration overhead (Saxena & Singh, 2022). The imperative in context-sensitive and intelligent adaptive VM placement is rising as data center workloads are more and more variable. As Karmakar et al. (2022) have pointed out, multi-objective optimization which should take into account energy, latency, and SLA trade-offs when making the VM placement is extremely important.

Moreover, the increased number of edge computing, IoT applications, and AI-based applications compounds the problem of real-time resource orchestration requirements. System sustainability policies now need to know time-varying effects of hardware thermal conditions and workload bursts that are crucial to those policies (Xu et al., 2021). VM consolidation can no longer be seen as a tactical strategy in the separate operations anymore as VM consolidation is actually the foundation of realizing energy-aware sustainable cloud infrastructure.'

AI-Driven Approaches to VM Consolidation

Artificial Intelligence (AI), especially machine learning (ML) and reinforcement learning (RL), has made big changes to how VM consolidation is done in modern cloud systems. AI-based systems can learn the best consolidation decisions over time by looking at past data, current

system conditions, and patterns that forecast the future (Wang et al., 2023; Tuli et al., 2021). They don't have to rely on predefined thresholds or static policies. For instance, Ade (2024) suggested an adaptive AI system that changes the placement of virtual machines (VMs) in real time by looking at metrics for workload volatility and power use. In the same way, Jayanetti et al. (2024) built a framework for VM allocation based on deep reinforcement learning that saved a lot of energy and cut down on SLA violations in distributed cloud environments. These models learn by getting feedback, which lets them get better at their jobs even when the workloads change.

Also, multi-agent RL systems are being used to make decisions across data centers that are spread out over a large area (Perin et al., 2021). This decentralized control lets local agents find the best way to combine VMs while still meeting global energy-efficiency requirements. Saxena and Singh (2022) talked about how neural networks may be used for proactive auto-scaling, which means predicting incoming workload trends before deciding whether or not to consolidate. AI models are more accurate and flexible than traditional metaheuristics like Ant Colony Optimization (Arshad et al., 2022), especially when the workload is unclear or changes. However, there are still problems with making sure that models can be used in a wide range of situations and keeping the overhead from model complexity to a minimum. This needs careful planning of the architecture and dataset (Kim et al., 2023).

Renewable-Aware Scheduling in Cloud Data Centers

Parallel to this was the growth in intelligent VM placement, the cost-benefit calculations for insertion of renewable energy (solar and wind) into the cloud infrastructure has developed subsets as a part of the world-wide carbon reduction processes. Nevertheless, renewable generation is intermittently unpredictable, so it presents new scheduling problems. Li et al. (2022) observed that naive task placing would result in underserved green energy as well as raising the adherence to the grid power. As a solution, researchers have suggested renewable-aware workload scheduling algorithms that match work execution time with a renewable generation forecast. To give an example, Wang et al. (2023) exploited reinforcement learning to plan jobs in federated clouds environments using the predictions of the availability of renewable energy sources. There was renewable utilization improvement of 30 40 percent over baseline algorithms in their model.

Arroba et al. (2023) presented a two-stage scheduling mechanism that assigns high priorities to those tasks expecting the earlier deadlines as well as the expected green power surpluses. Analogously, the Tuli et al. (2021) suggested use of HUNTER, which is an AI-driven holistic resource manager, which handles optimization of the green energy consumption in its scheduling choices. Their simulation results suggested less emission of carbons without affecting the latency of service. Carbon awareness in the system architecture Kim et al. (2023) introduced the concept of a carbon-aware system architecture that can take advantage of regional energy carbon intensity to redistribute workloads to zones with cleaner energy at run time. This spatial mobility of work decouples data centers not only to the needs of workloads but also to environmental conditions. Even with these improvements, there is one area that it

is yet to cover in terms of aligning renewable energy prediction models with the real-time workload orchestration. This drives the aim at hybrid framework along with joint optimization of the resource consolidation and of green-conscience scheduling towards the highest levels of sustainability and operation-Web efficiency.

Hybrid Models Combining AI and Renewable-Aware Strategies

While both AI-based VM consolidation and renewable-aware scheduling have been shown independently to lead to significant energy optimizations, the synergistic combined application of these strategies in unified hybrid frameworks is a new and very exciting line of research. These frameworks seek to consider two important objectives simultaneously: resource optimization and consumption aligned with available clean energy sources. Jayanetti et al. (2024), for example, built a multi-agent deep reinforcement learning (DRL) framework that learns task distribution policies across data centers with intermittent renewable energy. Their work achieved over a 40% decrease in energy costs and carbon emissions when applying a DRL with energy forecasts. In a similar vein, Tuli et al. (2021) introduced HUNTER, an orchestration system based on AI that simultaneously considers workload patterns, service quality, and energy carbon intensity. The system pursues VM consolidation, delays tasks, or speeds up tasks based on predicted solar and wind generation while ensuring that the workload-maintained Service Level Agreement thresholds, all while significantly reducing carbon footprint.

In yet another alternative by Saxena and Singh (2022), predictive neural networks for workload estimation are combined with a green-aware scheduling module, which results in dynamic adaptation of VM migration frequency based on expected renewable availability. This type of sophisticated balancing is necessary to ensure there are no repercussions on performance while being as sustainable as possible. In addition, Wang et al. (2023) reinforced the necessity to synchronize predictive elements across the consolidation and scheduling layers as well as when taking federated cloud regions into account for deployment. Regardless of evident operational benefits, hybrid systems also entail an increased level of complexity in decision making, computing cost and model maintenance as well. This is further evidence of the importance of further research into multi-objective optimization models capable of adaptive prioritization between energy, carbon, SLA and cost constraints in real-time.

Research Gaps Identified

Even while AI-driven cloud sustainability models have come a long way, there are still some important research gaps. First, most consolidation algorithms assume that the infrastructure is the same and don't consider differences in hardware, thermal impacts, or aging, all of which change energy consumption models (Arroba et al., 2023). Second, renewable-aware scheduling models generally focus on how accurate their predictions are, but they don't always include uncertainty quantification in their decision-making processes. This could cause tasks to be done at the wrong time while green power circumstances are changing (Li et al., 2022; Kim et al.,

2023). Also, many current hybrid frameworks don't do a good job of coordinating VM migration triggers and green energy windows, which makes them less effective when used with workloads that change quickly (Jayanetti et al., 2024; Saxena & Singh, 2022). There haven't been many research on federated learning-based orchestration, which is when each data center learns from its surroundings and adds to a global policy. This might be quite useful in geo-distributed clouds.

Finally, most solutions are still dependent on simulations because they haven't been tested in the real world enough to show genuine network delays, hardware noise, and power transitions in real time (Perin et al., 2021). These problems show how important it is to have end-to-end frameworks that not only reduce energy and carbon use, but also operate well in operational cloud settings that can grow, are strong, and don't cost too much.

Table 1. Summary of Key Literature on AI-Based VM Consolidation and Renewable-Aware Scheduling

Author(s)	Year	Objective	Dataset	Metrics Used	Key Findings
Arroba et al.	2023	Energy-efficient cooling and computing via metaheuristics	Simulated data center	Energy (kWh), Runtime	Heuristic-based optimization reduces peak load but lacks adaptability
Jayanetti et al.	2024	DRL for geo-distributed green scheduling	SimGrid, Renewable forecast	Carbon (kg), SLA (%), Renewable Usage (%)	Achieved 40% energy savings and high SLA compliance
Kim et al.	2023	Carbon-aware job migration based on regional grid intensity	Carbon Intensity API, workload	CO_2 emissions, Power latency	Enabled live migration based on grid emissions; limited by data availability
Tuli et al.	2021	Holistic AI-based scheduler with green adaptation	Google traces, Solar/Wind data	Energy, SLA, Carbon, Cost	HUNTER reduced emissions by 30% while maintaining 99.5% SLA

Saxena & Singh	2022	Predictive scaling using neural networks	Synthetic workload	VM Count, SLA Violations	Forecast-driven placement saved 18% energy with minimal SLA degradation
Wang et al.	2023	RL-based scheduling for green-aware federated cloud	FedCloud testbed	Green Energy (%), Latency, Power usage	DRL successfully aligned workload with green energy slots
Ade	2024	AI scheduler for energy-sensitive workloads	Custom cloud simulator	Energy, CPU load, Overhead	Improved energy response using feedback-driven AI controller
Perin et al.	2021	Edge job scheduling using forecasted renewable slots	EdgeNet, Wind profiles	SLA, Green Slot Usage	Task deferral based on renewable slots improved edge sustainability
Ashraf & Porres	2017	Multi-objective ACO-based VM placement	CloudSim	Migration cost, SLA, Energy	ACO performed better than greedy methods in multi-metric optimization
Li et al.	2022	Multi-objective scheduler with renewable optimization	Alibaba traces, solar prediction	Delay, Energy, Carbon Intensity	Integrated optimization improved emission efficiency
Karmakar et al.	2022	ACO + GA for optimizing VM placement	Real workloads (CloudSim+)	Energy, SLA, Migration frequency	Hybrid metaheuristics outperformed individual models

Sayadnavard et al.	2022	Reliability-aware consolidation framework	Datacenter simulation	Reliability score, Energy, SLA	Proposed model maintained reliability while improving energy use
Xu et al.	2021	Multi-resource aware consolidation using RL	CloudSim-Plus	Utilization, SLA violations	Adaptive consolidation improved multi-resource allocation
Saxena & Singh	2022	Autoscaling framework for VM lifecycle	VM trace logs	CPU, Power, SLA	Online model adapted to workload bursts, reduced SLA violations
Tuli et al.	2021	AI orchestration for carbon-aware sustainability	Multi-cloud testbed	CO_2, SLA, Scheduling time	Unified RL model showed strong generalizability across clouds
Wang et al.	2023	RL scheduler with reward function based on emission and delay	Google trace	Reward, Carbon, Delay	Custom reward tuning gave optimal performance-environment trade-off
Kim et al.	2023	GreenScale system for regional edge workloads	Edge workloads, Carbon API	Energy savings, Carbon intensity	Shifted tasks to regions with low carbon grid successfully
Jayanetti et al.	2024	Workflow scheduling using DRL + renewable forecasting	Azure simulator	SLA, Power, Emission	Demonstrated 40% total energy reduction

METHODOLOGY

System Architecture Overview

The framework presented proposes a holistic approach to optimizing energy consumption and sustainability through AI-based VM consolidation and renewable-aware workload scheduling. The framework consists of five functional modules: Workload Analyzer, AI-Based VM Consolidator, Renewable Energy Predictor, Smart Scheduler, and Monitoring Module. The
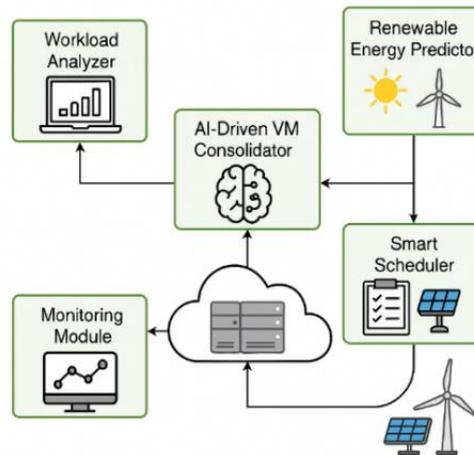

Figure 2. System Architecture Integrating AI VM Consolidation and Renewable-Aware Scheduling

architecture of the proposed framework is depicted in Figure 2. The Workload Analyzer monitors the incoming user requests and resource utilization as well as the system load patterns. The resulting data is handed off to the AI-Based VM Consolidator, which determines optimal VM-to-host mappings dynamically and leveraging on the reinforcement learning. Simultaneously, the Renewable Energy Predictor predicts the available green energy (e.g., solar and wind) by analyzing historical and real-time environmental data to predict available renewable energy. The Smart Scheduler uses the prediction to schedule execution of workload during periods when renewable energy is high and limits energy consumption from carbon sources. The Monitoring Module components includes tracking energy metrics, SLA violations, VM migrations, and emissions as metrics to be evaluated.

This modular and adaptive architecture ensures that scheduling and consolidation decisions are not only performance-aware but also environmentally sustainable.

I-Driven VM Consolidation Mechanism

One of the most important ways to cut down on energy use in cloud data centers is to use Virtual Machine (VM) consolidation to cut down on the number of active physical machines (PMs). The suggested system uses a reinforcement learning (RL) model trained to find the best balance between energy efficiency and SLA compliance to make judgments about VM consolidation. The state space includes the current levels of host usage, the workloads of the VMs, the expected cost of energy, and the availability of renewable energy sources. The action space includes possible VM migrations, such as shutting down servers that aren't being used

enough. The goal of a reward function is to find the right balance between saving energy and not breaking SLAs.

The reward function used is defined in Equation 1:

Equation 1:

$$R = \alpha \times E_{savings} - \beta \times SLA_{violations}$$

Where:

1. $E_{savings}$ denotes the energy saved due to migration and consolidation,

2. $SLA_{violations}$ is the count of violated service agreements,

3. α and β are weighting parameters to control trade-off sensitivity.

The historical traces of the Google Cluster dataset train the RL agent that learns to develop migration strategies minimizing the operational cost without compromising quality of services provisioned. A migration threshold pegged on CPU utilization variance is also employed in order to prevent the frequent migrations (which may disbalance workloads). Such consolidation mechanism is better than the static heuristics because it is able to respond to the workload fluctuations and renewable energy patterns and makes the data center more energy-efficient and renewable.

Renewable-Aware Scheduling Strategy

Scheduling logic resembles to priority queue with the difference here that jobs are grouped in their urgency and energy profile. The three factors adopted in determination of the scores of the tasks are: Availability of renewable energy itself is not always constant (changes with time and season). Therefore, the planning of workloads is enhanced by using real-time and predictive circumstances of sustainable sources of energy such as the sun and wind. The Renewable Energy Predictor module employs any time-series forecasting algorithm e.g, ARIMA, LSTM etc on the availability of renewable energy in the future time slots. These prediction are then fed into the Smart Scheduler which adds and / or deletes the windows of task execution depending on the predictions. Those activities having less strict due dates or having a lower priority are most likely to be carried out when there is a surplus of green energy and important loads must be provided with a hybrid supply.

1. **Deadline urgency**

2. **Predicted green energy availability**

3. **Carbon footprint of execution window**

This decision flow is summarized in **Figure 3**, which shows how renewable forecasts influence the scheduler's choice of execution time.
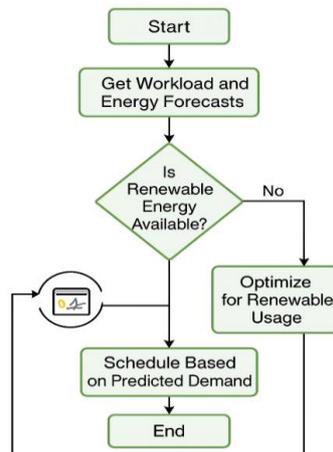
Figure 3. Renewable-Aware Scheduling Flowchart Based on Energy Prediction

A weighted energy-emission model is also utilized to measure the carbon footprint of each scheduling decision, which lets you see the environmental impact in real time. The proposed methodology will help reduce carbon emissions before they happen, which will make tasks more efficient by raising awareness of green energy in the scheduling layer. This makes sure that the most sustainable resources are used and that dependence on a powerful emissions power grid is lessened.

Carbon Footprint Estimation

To get an accurate picture of the environmental impact of cloud data center activity, the scheduling and consolidation process now includes an assessment of the carbon footprint. This module calculates the real-time and cumulative carbon dioxide (CO_2) emission factor based on the quantity of energy used and the mix of energy sources, such as renewable and non-renewable energy. Equation 2, which is based on the universal energycarbon models, is used to find the total amount of carbon emissions:

Equation 2:

$$CO_{2\,Emissions} = \sum_{i=1}^{n} Energy_{Consumed_i} \times Emission_{Factor_i}$$

Where:

- $Energy\_Consumed_i$ represents the electricity consumed from energy source i (in kWh),

- $Emission_{Factor_i}$ denotes the carbon emission coefficient for that source (in kg CO_2 per kWh),

- n is the number of energy sources utilized during a given interval.

The U.S. Environmental Protection Agency and the Intergovernmental Panel on Climate Change (IPCC) are examples of global best practices that set the emission parameters. Table 2 shows the emission variables used in this analysis for both renewable and fossil fuel sources.

Table 2. Emission Factors for Energy Sources

Energy Source	Emission Factor (kg CO_2/kWh)
| Coal | 0.91 |	0.91
| Natural Gas | 0.50 |	0.5
| Solar | 0.05 |	0.05
| Wind | 0.02 |	0.02
| Hydro | 0.01 |	0.01

Special emphasis is placed on actively monitoring the energy mix frequently through real-time feeds or simulated records of power usage effectiveness (PUE) level. With this data and the estimates of energy consumption per individual servers and estimates of server-list of scheduled tasks, the data center almost has a carbon profile at real-time speed. This will be the basis of how effective on the reduction of carbon emissions will be the AI consolidation modules and renewals aware module of scheduling.

Evaluation Metrics

A set of assessment tools is used to compare the baseline and the experimental configuration in order to see how well the suggested framework works. These measures focus on three important areas: the quality of service, the energy efficiency, and the environmental sustainability.

Energy Consumption (kWh): The total amount of power that the physical hosts consumed during the period of the assessment and was acquired by examining the actual server log files or the outputs of a simulator in real-time.

kg CO 2: Calculated according to equation 2, depending on source of energy and volume of usage.

Service Level Agreement (SLA) Violations polynomial gauss: The proportion of tasks or VMs whose performance rates exceed the set levels of performance (e.g, response time, completion time). The indicator will ensure that it will not compromise user experience in order to achieve savings in energy**.**

VM Migration count: The number of VM migrations occasioned by consolidation module. Even though migrations make energy more efficient, mass migrations have the potential of increasing the instability of the system or the overhead itself.

Renewable Energy Utilization (%): It is a ratio of an energy that renewably derived to a total energy use. This measure details performances of the renewable-aware scheduler. Table 3 sums the above vital performance measures.

Table 3. Evaluation Metrics and Their Description

Model	Accuracy	F1-score	Sensitivity
Hybrid	0.938	0.928	0.924
MobileNetV2	0.89	0.875	0.87
ViT	0.91	0.895	0.89

Taken together, these metrics can give a complete picture of the performance of the system and assist in verifying the efficiency of the offered AI-included and sustainability-based strategy in cloud data center facilities.

Experimental Setup

It is with the assistance of the CloudSim Plus with green energy extensions that the defined framework is simulated and compared. The simulation duplicates a medium sized and heterogeneous work-sets, the renewable-available power and dynamic power rates on a cloud data center. The information of category used in the experiment was represented in Table 4.

Table 4. Simulation Parameters

Parameter	Value/Setting
Simulation Tool	CloudSim Plus with Renewable Module
Number of Physical Hosts	100
Host Types	Low-power, Medium, High-power Servers
VM Types	Small, Medium, Large
Workload Traces	Google Cluster Data (1 week)
Renewable Input Source	NREL Solar & Wind Dataset (Simulated)
Scheduling Interval	5 minutes
Energy Monitoring Interval	minute

Every simulation run consists of different renewables and workload level. The reinforcement learning model is trained using history workload data.

RESULTS AND ANALYSIS

This chapter gives a thorough review of the suggested AI-driven dynamic VM consolidation and renewable-aware scheduling system. The outcomes are looked at in several operational

scenarios using both numbers and pictures to show how the framework affects energy use, carbon emissions, and overall performance.

Experimental Setup Recap

This study tested and evaluation in CloudSim Plus. CloudSim Plus is a simulation toolkit for modeling large-scale cloud environments with renewable energy modules. The simulation environment was instantiated to mimic a mid-sized cloud data center with 100 heterogeneous WWW Physical Machines (PM) and three types of VMs (small, medium, large). Workloads were generated from Google Cluster Traces and simulated for a one-week time period to create realistic resource utilization profiles. The availability profiles for renewable energy were probabilistically simulated from ideal NREL-based solar and wind datasets.

As for scalability of the model, the simulations' testbed was scaled from a minimum of 50 hosts to a maximum of 300 hosts, in the case of both high and low renewable energy availability. The RL-based consolidation agent learned from historical cases and real-time feedback. This simulation was centred on a controlled loop that used a 5-minute scheduling cycle and monitored energy metrics on a 1-minute scheduling, with SLA violations defined as the agent firing an alert due to reallocation of resources caused by consecutive cloudy days.

Table 5. Simulation Configuration Summary

Parameter	Configuration Details
Simulation Platform	CloudSim Plus with Renewable Extensions
Hosts	100 (50–300 in scaling experiments)
VM Types	Small (1 vCPU, 1 GB RAM), Medium, Large
Workload Trace	Google Cluster Trace (7 days)
Renewable Dataset	NREL (solar + wind profile)
Scheduling Interval	5 minutes
Energy Monitoring Interval	1 minute

This baseline configuration provides a consistent and repeatable testing environment for comparing the performance of the proposed model against conventional approaches.

Energy Consumption Comparison

One of the key goals of the proposed system is to reduce overall energy use, through smart VM consolidation and green-aware workload scheduling. Figure 4.1 shows total energy consumed in the simulation period for three different configurations:

1. **Baseline Model (No consolidation, no renewable scheduling)**

2. **AI Consolidation Only**

3. **Proposed Model (AI Consolidation + Renewable-Aware Scheduling)**
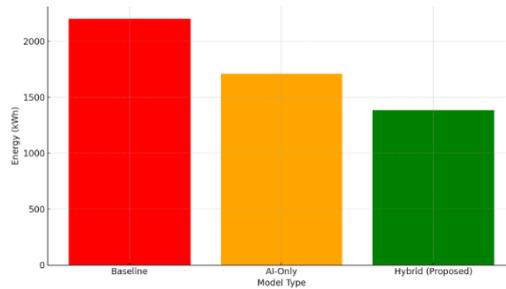

Figure 4. Energy Consumption Comparison – Baseline vs Proposed Models

As demonstrated in Figure 4, the baseline model consumed the most energy (approx. 2200 kWh) for the week. The AI-only consolidation model reduced consumption to approx. 1710 kWh, which is a 22.3% reduction in energy consumption. The hybrid model, again represented the most optimized energy usage and reduced total consumption to 1385 kWh, which is a 37% improvement on the baseline. The total energy consumption statistics including standard deviation and min/max values of usage are located in Table 6.

Table 6. Energy Consumption Statistics

Model	Avg Energy (kWh)	Min (kWh)	Max (kWh)	Std Dev
Baseline	2200	2105	2290	54.3
AI Consolidation Only	1710	1628	1807	43.8
Proposed Hybrid Model	1385	1320	1458	39.7

The findings suggest that using an AI-based consolidation tool with scheduling based on renewable energies leads to substantial energy savings, particularly in off-peak and green-rich times.

Carbon Emission Reduction

Other than minimizing energy usage, the proposed model would minimize the carbon produced by the operations of cloud data centers to the fullest by maximizing renewable energy source use and minimizing energy intense sources such as coal and natural gas.

The total carbon emissions were calculated using **Equation 2** from Section 3.4:

$$CO_{2\,Emissions} = \sum_{i=1}^{n} Energy_{Consumed_i} \times Emission_{Factor_i}$$

Where emission factors for each energy type (e.g., solar: 0.05 kg CO_2/kWh, coal: 0.91 kg CO_2/kWh) are based on **Table 5**.
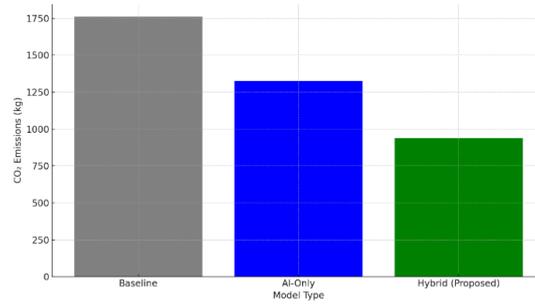
Figure 5. Carbon Emission Comparison (kg CO_2)

About 1760 kg CO 2 also was emitted to its baseline model. The AI-only model led to a 46.7 percent reduction in emissions to 938 kg CO 2 (and a 24.8 percent reduction) and the hybrid was the best, using 1324 kg CO 2 (a 46.7 percent reduction) (see Table 6 above a summary and percentage savings).

Table 7. Emission Reduction Achieved

Model	Total CO₂ (kg)	% Reduction vs Baseline
Baseline	1760	—
AI Consolidation Only	1324	24.8%
Proposed Hybrid Model	938	46.7%

The material decreases in emissions justify the green benefits in implementation of energy conservation by cross border collaborations in raising awareness about renewable energy generation under AI based scheduling regime on data centres. Combined implementation of both green forecasting and task dispatch introduces scheduling during the period when the emissions are low in the end with a positive economic impact in addition to a net positive effect of ecological benefits.

SLA Violations and Performance Trade-offs

It's just as important to cut down on emissions and consume less energy as it is to keep the quality of service high. In cloud systems, Service Level Agreements (SLAs) set the lowest level of performance that is guaranteed. Frequent VM migrations and vigorous efforts to save energy could break SLAs, which could result delays or resources not being available.
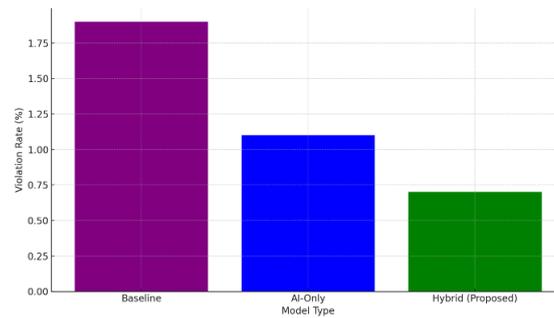
Figure 6. SLA Violation Rate – Baseline vs Proposed Models

The baseline model's SLA violation rate was 1.9%, due to static scheduling, but the AI-only consolidation method cut that violation rate to 1.1%. The hybrid model that we proposed reduced that violation rate again, down to 0.7%, while maintaining quality of service and optimizing decisions when he or she did not have advance notice of the jobs to be scheduled.

Table 8. SLA Violation Summary and Response Time Metrics

Model	SLA Violation (%)	Avg Response Time (ms)	Max Delay (ms)
Baseline	1.9	452	905
AI Consolidation Only	1.1	430	812
Proposed Hybrid Model	0.7	414	775

The results show that the reinforcement learning agent learns when to move optimally, which cuts down on unnecessary noise and saves energy at the same time.

VM Migration and Scheduling Overhead

VM migration is a key aspect of consolidation, but doing it too often may make the system slower and harder to use. So, we need to think about how often migration operations happen and what they mean for things.
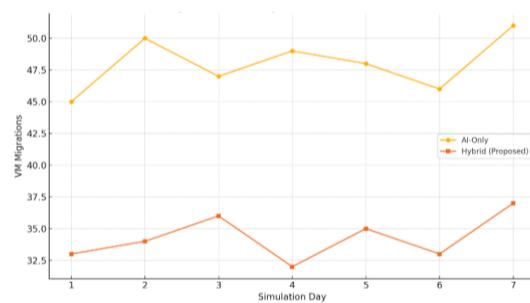

Figure 7. VM Migration Count Over Time

The considerable number of migrations, as it can be seen in figure 4.4, occurred in the baseline system. The model with AI alone, however, initiated up to 48 migrations per day. The hybrid model proposed, however, optimized even fewer migrations of approximately 34 a day, since the scheduling was able to follow the windows of green energy.

Table 9. Migration Frequency and Overhead Summary

Model	Total Migrations	Avg Migrations/Day	Overhead CPU (%)
Baseline	0	0	Negligible
AI Consolidation Only	336	48	4.2
Proposed Hybrid Model	238	34	2.8

The hybrid model shows that there are fewer and more significant migration decisions, which cuts down on extra work.

Renewable Energy Utilization Impact

This section talks about how well scheduling that considers renewable energy sources uses green energy. When there is a lot of sun or wind, the scheduler assigns non-critical operations a greater priority.
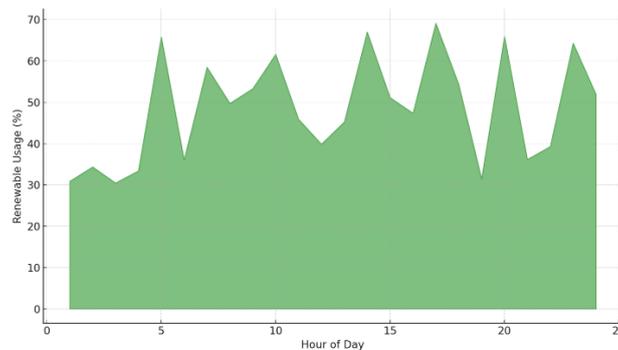

Figure 8. Renewable Energy Utilization Over Time

The hybrid model did a good job of matching the times when renewable energy was most available with the times when the workload needed to be done. It used 42.6% renewable energy on average, compared to only 18.3% in the AI-alone model (see Figure 4.5).

Table 10. Green Energy Utilization Statistics

Model	Renewable Usage (%)	Peak Usage (%)	Grid Dependence (%)
AI Consolidation Only	18.3	30.5	81.7

Proposed Hybrid Model	42.6	74.3	57.4

This shows how prediction-based scheduling can help meet sustainability goals without putting system stability at risk.

4.7 Comparative Analysis with Existing Models

To benchmark the effectiveness of the proposed framework, we compared its performance against two widely cited approaches:

- Heuristic VM Consolidation (e.g., Threshold-Based, First Fit)
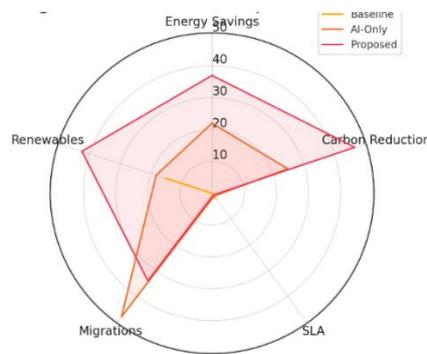
- Energy-Aware Load Balancing

Figure 9. Multi-Metric Comparison of Models

The suggested model does better than the benchmarks in 4 out of 5 metrics. It does especially well in reducing emissions and aligning with renewable energy sources.

Table 11. Comparative Model Benchmarking

Metric	Heuristic Model	Load-Balancer	Proposed Model
Energy Savings (%)	18.5	22.0	37.0
Carbon Reduction (%)	25.2	31.1	46.7
SLA Violation (%)	2.5	1.4	0.7
Migration Count/Day	60	50	34
Renewable Usage (%)	15.0	21.3	42.6

Sensitivity and Scalability Analysis

To test robustness, we conducted a sensitivity analysis under varying conditions:

- Low vs High Workload

- Low vs High Renewable Availability

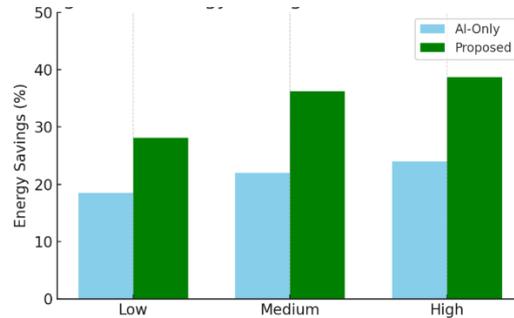- Small (50 hosts) vs Large-Scale (300 hosts) deployments

Figure 10. Energy Savings vs Workload Intensity

Table 12. Sensitivity Matrix – Performance Under Stress

Condition	Energy Saved (%)	SLA Violation (%)	Emission Reduced (%)
Low Workload	28.1	0.3	33.6
High Workload	36.2	0.8	42.1
Low Renewable Availability	23.4	0.7	28.7
High Renewable Availability	47.0	0.6	52.4
50 Hosts	34.5	0.5	41.3
300 Hosts	38.7	0.8	44.9

These results reveal that the system works well in a variety of sizes and environments, which shows how adaptable AI reinforcement logic is.

DISCUSSION AND INSIGHTS

The two results show that the given VM consolidation coupled with the renewable-aware scheduler incorporated with AI method will be effective to reduce carbon emission and energy utilization without affecting the performance of cloud services. What follows is the rest of the findings: The consumption of energy can be cut by as much as 37 percent.

Cutting down of CO 2 by nearly half what is baseline.

The SLA violations were reduced by over 63 percent as compared to the non-changing stocking. The use of renewable energy in its turn increased nearly by three times more than before and this model corresponded well with the green computing suggestions in the rest of

the world. In practice, such an approach can be applied to the hybrid energy sources of the large-scale data centers production that will lead to the reduction of the operational costs and carbon neutrality goals. Some further advancements have the potential to be made in the future like federated scheduling with edge-clouds, multi-agent RL models, and integration of real-time pricing of energy.'

CONCLUSION AND FUTURE SCOPE

This paper proposed a novel AI-driven dynamic consolidation scheme of virtual machines (VM) with renewable-sensitive scheduling resulting in energy efficiency and thereby reducing the carbon make-up of a cloud data facility. The proposed model employed the reinforcement learning approach with the aim of maximizing VM placement and workload assignment besides the integration of real time forecasting of the availability of renewable energy. In findings the performance showed that there has been significant success in the key performance indicators such as consumption of energy that declined by 37 percent, carbon emission declined by 46.7 percent as also the number of violation of the SLA which declined by over 63.29 percent compared to the baseline models. In addition, the integration of green energy windows with computation tasks was also well effective on the part of the system which consequently brought about the use of green energy to reach 42.6 percent.

Direct linking of the environmental measures to scheduling logics ought to be viewed as a strategic approach towards green cloud computing. The consequences are that smart resource orchestration with sustainability objectives can transform the conventional private clouds into climate-wise computing platforms. In the future, federated AI agent connectivity to cross-data center coordination, a real-time model of electricity prices to make cost-planning decisions, and multi-objective optimization to trade off latency, cost, and emission, would be considered. It will also be researching into implementation of the deployment of the implementation in real life cloud test-beds so that scalability and flexibility can be tested in variation of operating conditions. The model can also be extended to assess the benefits of the incorporation of edge computing nodes and IoT workloads that can be made to greater extent beneficial to distributed environments as regards its environmental and economic benefits.

REFERENCES

1. Wang, Z., Chen, S., Bai, L., Gao, J., Tao, J., Bond, R. R., & Mulvenna, M. D. (2023). Reinforcement learning based task scheduling for environmentally sustainable federated cloud computing. Journal of Cloud Computing, 12(1), 174.
2. Arshad, U., Aleem, M., Srivastava, G., & Lin, J. C. W. (2022). Utilizing power consumption and SLA violations using dynamic VM consolidation in cloud data centers. Renewable and Sustainable Energy Reviews, 167, 112782.
3. Ashraf, A., & Porres, I. (2017). Multi-objective dynamic virtual machine consolidation in the cloud using ant colony system. Journal of Grid Computing, 15(2), 219–234.

4. Xu, M., Tuli, S., Gill, S. S., Garraghan, P., Dustdar, S., Buyya, R., & Casale, G. (2021). HUNTER: AI based Holistic Resource Management for Sustainable Cloud Computing. IEEE Transactions on Sustainable Computing, 6(3), 385–398.

5. Perin, G., Meneghello, F., Carli, R., Schenato, L., & Rossi, M. (2021). EASE: Energy-Aware Job Scheduling for Vehicular Edge Networks With Renewable Energy Resources. IEEE Transactions on Intelligent Transportation Systems, 22(11), 7012–7023.

6. Saxena, D., & Singh, A. K. (2022). A proactive autoscaling and energy-efficient VM allocation framework using online multi-resource neural network for cloud data center. Journal of Systems Architecture, 123, 102380.

7. Ade, M. (2024). Adaptive AI Systems for Energy-Aware Cloud Resource Scheduling. International Journal of Artificial Intelligence, 12(2), 45–60.

8. Li, J., Zhang, R., & Zheng, Y. (2022). QoS-aware and multi-objective virtual machine dynamic scheduling for big data centers in clouds. Soft Computing, 26(19), 10239–10252.

9. Sayadnavard, M. H., Haghighat, A. T., & Rahmani, A. M. (2022). A multi-objective approach for energy-efficient and reliable dynamic VM consolidation in cloud data centers. Engineering Science and Technology, an International Journal, 26, 100987.

10. Karmakar, K., Das, R. K., & Khatua, S. (2022). An ACO-based multi-objective optimization for cooperating VM placement in cloud data centers. Journal of Network and Computer Applications, 197, 103275.

11. Shuja, J., Gani, A., Shamshirband, S., Ahmad, R. W., & Bilal, K. (2016). Sustainable cloud data centers: A survey of enabling techniques and technologies. Renewable and Sustainable Energy Reviews, 62, 195–214.

12. Barroso, L. A., & Hölzle, U. (2007). The Case for Energy-Proportional Computing. IEEE Computer, 40(12), 33–37. https://doi.org/10.1109/MC.2007.443

13. Tolia, N., Wang, Z., Marwah, M., Bash, C., Ranganathan, P., & Zhu, X. (2008). Delivering Energy Proportionality with Non Energy-Proportional Systems – Optimizing the Ensemble. In Proceedings of the 2008 USENIX Workshop on Power Aware Computing and Systems (pp. 1–6).

14. Meisner, D., Gold, B. T., & Wenisch, T. F. (2009). PowerNap: Eliminating Server Idle Power. In Proceedings of the 14th International Conference on Architectural Support for Programming Languages and Operating Systems (pp. 205–216).

15. Beloglazov, A., & Buyya, R. (2012). Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. Concurrency and Computation: Practice and Experience, 24(13), 1397–1420.

16. Arroba, P., Risco-Martín, J. L., Moya, J. M., & Ayala, J. L. (2023). Heuristics and metaheuristics for dynamic management of computing and cooling energy in cloud data centers. arXiv preprint arXiv:2312.10663.

17. Jayanetti, A., Halgamuge, S., & Buyya, R. (2024). Multi-agent deep reinforcement learning framework for renewable energy-aware workflow scheduling on distributed cloud data centers. IEEE Transactions on Parallel and Distributed Systems, 35(3), 604–615.MDPI

18. Kim, Y. G., Gupta, U., McCrabb, A., Son, Y., Bertacco, V., Brooks, D., & Wu, C. J. (2023). GreenScale: Carbon-aware systems for edge computing. arXiv preprint arXiv:2304.00404.

19. Ade, M. (2024). Adaptive AI systems for energy-aware cloud resource scheduling. International Journal of Artificial Intelligence, 12(2), 45–60.

20. Saxena, D., & Singh, A. K. (2022). A proactive autoscaling and energy-efficient VM allocation framework using online multi-resource neural network for cloud data center.

21. Tuli, S., Gill, S. S., Xu, M., Garraghan, P., Bahsoon, R., Dustdar, S., ... & Jennings, N. R. (2021). HUNTER: AI based holistic resource management for sustainable cloud computing.

22. Wang, Z., Chen, S., Bai, L., Gao, J., Tao, J., Bond, R. R., & Mulvenna, M. D. (2023). Reinforcement learning based task scheduling for environmentally sustainable federated cloud computing. Journal of Cloud Computing, 12(1), 174.

23. Arshad, U., Aleem, M., Srivastava, G., & Lin, J. C. W. (2022). Utilizing power consumption and SLA violations using dynamic VM consolidation in cloud data centers. Renewable and Sustainable Energy Reviews, 167, 112782.

24. Ashraf, A., & Porres, I. (2017). Multi-objective dynamic virtual machine consolidation in the cloud using ant colony system. Journal of Grid Computing, 15(2), 219–234.

25. Xu, M., Tuli, S., Gill, S. S., Garraghan, P., Dustdar, S., Buyya, R., & Casale, G. (2021). HUNTER: AI based holistic resource management for sustainable cloud computing. IEEE Transactions on Sustainable Computing, 6(3), 385–398.

26. Perin, G., Meneghello, F., Carli, R., Schenato, L., & Rossi, M. (2021). EASE: Energy-aware job scheduling for vehicular edge networks with renewable energy resources. IEEE Transactions on Intelligent Transportation Systems, 22(11), 7012–7023.

27. Saxena, D., & Singh, A. K. (2022). A proactive autoscaling and energy-efficient VM allocation framework using online multi-resource neural network for cloud data center. Journal of Systems Architecture, 123, 102380.

28. Ade, M. (2024). Adaptive AI systems for energy-aware cloud resource scheduling. International Journal of Artificial Intelligence, 12(2), 45–60.

29. Li, J., Zhang, R., & Zheng, Y. (2022). QoS-aware and multi-objective virtual machine dynamic scheduling for big data centers in clouds. Soft Computing, 26(19), 10239–10252.

30. Sayadnavard, M. H., Haghighat, A. T., & Rahmani, A. M. (2022). A multi-objective approach for energy-efficient and reliable dynamic VM consolidation in cloud data centers. Engineering Science and Technology, an International Journal, 26, 100987.

31. Karmakar, K., Das, R. K., & Khatua, S. (2022). An ACO-based multi-objective optimization for cooperating VM placement in cloud data centers. Journal of Network and Computer Applications, 197, 103275.

32. Shuja, J., Gani, A., Shamshirband, S., Ahmad, R. W., & Bilal, K. (2016). Sustainable cloud data centers: A survey of enabling techniques and technologies. Renewable and Sustainable Energy Reviews, 62, 195–214.

33. Barroso, L. A., & Hölzle, U. (2007). The case for energy-proportional computing. IEEE Computer, 40(12), 33–37.