International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 5-Nov-Dec -2026, ISSN (Online): 2395-566X

Green Cloud Computing: A Survey of Energy-Efficient Resource Management and Carbon-Aware Scheduling in Cloud Data Centres

Bhor Aniket Madhav[1] Boddawar Kunal Chandrakant[2] Walke Yashraj Dinkar[3]
Prof. S. L. Rahane[4] Prof. M. S. Nalawade[5] Dr. V. M. Dhede[6]

[123]Department Of Electronics And Telecommunication Jaihind College Of Engineering Kuran, Pune, Maharashtra, India
Seminar Guide Department Of Electronics And Telecommunication Jaihind College Of Engineering Kuran, Pune, Maharashtra, India; Seminar Coordinator Department Of Electronics And Telecommunication Jaihind College Of Engineering Kuran, Pune, Maharashtra, India; Head Of Department Department Of Electronics And Telecommunication Jaihind College Of Engineering Kuran, Pune, Maharashtra, India

Abstract—Cloud platforms deliver compute, storage and software as on-demand services, yet the data centres behind them draw large and continuous amounts of electricity for servers, network fabrics and cooling plants. Green cloud computing is the body of techniques that lowers this energy and carbon footprint without breaking the service-level commitments made to users. This survey organises the field into four layers of intervention: efficient hardware and facility design, virtualization with dynamic consolidation of virtual machines, energy- and carbon-aware scheduling of workloads, and integration of renewable supply. For each layer the paper explains the underlying principle, reviews representative studies, and compares their objectives, metrics and limitations. Building on this review, a layered conceptual framework is described in which a monitoring module gathers utilization, power and grid-carbon signals, a prediction module estimates near-term demand, and a decision engine places and migrates workloads on the most suitable servers while idle hosts are switched to low-power states. Evaluation through simulation toolkits or a small private-cloud test bed is discussed together with metrics such as power usage effectiveness, energy per request and SLA violation rate. Open issues, including the energy cost of migration, accurate carbon accounting and the trade-off between consolidation and performance, are identified as directions for further study.

Index Terms—Green cloud computing, energy efficiency, virtualization, VM consolidation, carbon-aware scheduling, power usage effectiveness, data centre.

I. INTRODUCTION

Cloud computing has changed the way organisations obtain computing resources. Instead of purchasing and maintaining their own servers, users rent processing capacity, storage and ready-made software from a provider and pay only for what they consume [1]. This model is attractive because it converts large capital investment into a flexible operating expense, and because resources can be scaled up or down within minutes. Education platforms, online banking, streaming services, e-commerce and, more recently, the training of artificial intelligence models all rely on this elastic infrastructure.

The convenience of the cloud hides a physical reality. Every virtual machine finally runs on a real server located inside a data centre, and that facility must also power storage arrays, switches, uninterruptible power supplies, lighting and, above all, cooling equipment that removes the heat produced by electronic components [2], [3]. Studies of global data-centre consumption show that, although efficiency gains have so far kept total energy use from growing as fast as computing demand, the absolute amount remains significant and is sensitive to future growth in workloads [4]. A further concern is that servers in many facilities operate at low average utilization while still consuming a large share of their peak power, so a substantial part of the electricity is spent on capacity that does little useful work [5].

Green cloud computing addresses this gap. Its aim is to deliver the same quality of service with less energy and lower carbon emission, by combining better hardware, smarter software and cleaner power sources [6]. The topic is broad, and published work is scattered across computer architecture, distributed systems, power engineering and operations research. The purpose of this survey is to bring these strands together in a form suitable for academic study.

The main contributions of this paper are as follows:

- A classification of green cloud techniques into hardware and facility, virtualization and consolidation, energy- and carbon-aware scheduling, and renewable-energy integration.

© 2026 IJSRET

- A structured review and comparison of representative studies in each category, including their metrics and limitations.
- A layered conceptual architecture for a green cloud management system that combines monitoring, demand prediction and energy-aware placement.
- A discussion of evaluation methods, open challenges and future research directions.

The remainder of the paper is organised as follows. Section II explains the sources of energy use in a data centre and the metrics used to measure efficiency. Section III presents the taxonomy and literature survey. Section IV describes the proposed conceptual framework. Section V compares the approaches, Section VI discusses evaluation, Section VII outlines challenges and future work, and Section VIII concludes the paper.

II. BACKGROUND

1. Where The Energy Goes

The power drawn by a data centre can be divided into the IT load and the facility overhead. The IT load consists of servers, storage and network equipment. The overhead includes chillers, computer-room air handlers, pumps, power distribution losses and lighting [3]. Inside a server, the processor is typically the largest single consumer, followed by memory, storage devices, fans and the power supply. An important observation is that the power of a conventional server does not fall in proportion to its load: an idle machine may still draw roughly half of its peak power. Barroso and Hölzle called for *energy-proportional* computing, in which power consumption tracks utilization closely, because real servers spend most of their time between 10% and 50% load [5].

2. Efficiency Metrics

The most widely quoted facility metric is power usage effectiveness (PUE), defined by The Green Grid as the ratio of total facility power to the power delivered to IT equipment [7]:

$$PUE = P_{total} / P_{IT} \quad (1)$$

An ideal facility would have a PUE of 1.0, meaning that no energy is spent outside the IT equipment. PUE, however, says nothing about how usefully the IT equipment itself is used. For that reason researchers also report server utilization, energy consumed per request or per job, and SLA violation rate. Where carbon is the concern, the carbon intensity of the electricity supply, measured in grams of CO_2 per kilowatt-hour, is multiplied by the energy consumed to obtain the emission attributable to a workload [8].

A commonly used linear model expresses server power as a function of CPU utilization u [9]:

$$P(u) = P_{idle} + (P_{max} - P_{idle}) \cdot u \quad (2)$$

This simple relationship is the basis of many consolidation algorithms, because it shows that packing load onto fewer machines and switching off the rest removes the large idle term from every powered-down host.

3. Virtualization As The Enabler

Hypervisors such as Xen allow several isolated virtual machines (VMs) to share one physical host [10]. Live migration moves a running VM between hosts with only a brief pause, which makes it possible to rearrange workloads at run time without interrupting users [11]. These two capabilities together are the foundation of most software-level energy savings in the cloud.

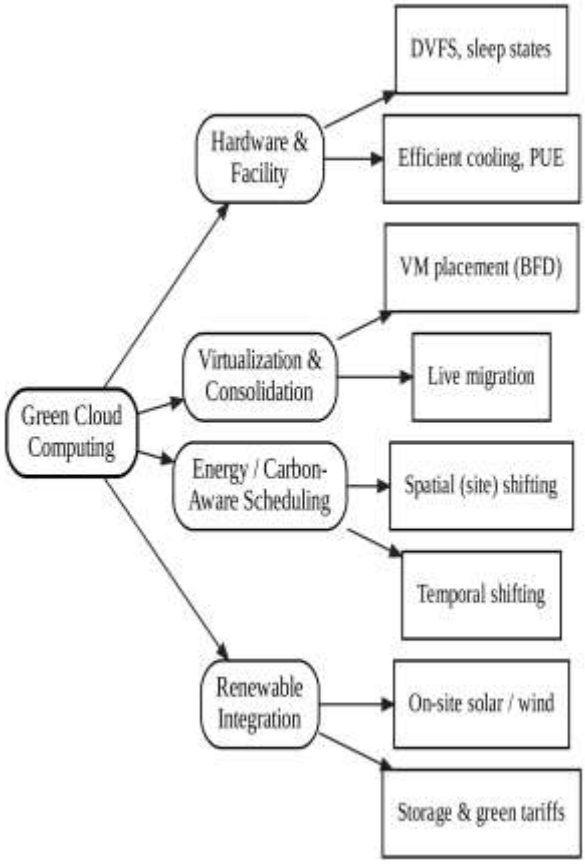


Fig. 1. Taxonomy of green cloud computing techniques considered in this survey.

© 2026 IJSRET

III. TAXONOMY AND LITERATURE SURVEY

Fig. 1 groups the reviewed techniques into four categories. The categories are complementary: a facility can apply all of them at once, and the best results reported in the literature usually come from combining several layers.

1. Hardware And Facility Level

At the hardware level, dynamic voltage and frequency scaling (DVFS) lowers processor voltage and clock frequency when demand is light, reducing dynamic power roughly in proportion to frequency and the square of voltage. Meisner *et al.* proposed PowerNap, in which an entire server moves rapidly between a fully active state and a near-zero-power nap state, removing most idle consumption for workloads with frequent short gaps [12]. At the facility level, hot-aisle and cold-aisle containment, free cooling with outside air, and better power-distribution equipment reduce the overhead captured by PUE [3]. These measures are effective but require capital investment and are mainly under the control of the data-centre operator.

2. Virtualization And Vm Consolidation

Dynamic consolidation is the most studied software technique. Beloglazov, Abawajy and Buyya formulated energy-aware allocation as a bin-packing problem and introduced a modified best-fit-decreasing heuristic that places each VM on the host whose power would increase the least [9]. In a follow-up study, Beloglazov and Buyya proposed adaptive thresholds that decide when a host is overloaded, based on statistical analysis of past utilization, together with VM-selection policies such as minimum migration time [13]. Their experiments, performed with real workload traces in the CloudSim simulator [14], showed considerable energy savings while keeping SLA violations bounded. The main weakness of consolidation is that aggressive packing can overload hosts and that each migration itself consumes network bandwidth and energy [11].

3 Energy- And Carbon-Aware Scheduling

Scheduling decides where and when a job runs. Garg *et al.* studied high-performance computing workloads distributed across several cloud sites and showed that carbon emission can be reduced by sending jobs to sites with lower carbon intensity and better cooling efficiency, while still meeting deadlines [15]. More recently, Radovanović *et al.* described a carbon-intelligent computing system deployed in production, which shifts flexible, delay-tolerant workloads in time towards hours when the local grid is cleaner [8]. Network-aware placement is also relevant: Kliazovich *et al.* built the GreenCloud simulator to capture the energy of switches and links, showing that communication energy cannot be ignored in data-centre studies [16].

4. Renewable Energy Integration

Data centres can lower their emissions by drawing power from on-site solar panels, wind contracts or green tariffs. Because renewable output varies with weather and time of day, scheduling and storage become essential partners: deferrable work can be timed to match generation, and batteries can smooth short gaps [8], [15]. Surveys of cloud energy efficiency point out that combining supply-side and demand-side flexibility offers more benefit than either alone [6].

Table I Summary Of Representative Studies In Green Cloud Computing

Ref.	Category	Technique	Key Finding	Limitation
[5]	Hardware	Energy-proportional design argument	Servers spend most time at 10–50% load where efficiency is poor	Conceptual; requires new hardware
[12]	Hardware	PowerNap rapid sleep/wake	Idle power largely eliminated for bursty loads	Needs fast transition support
[9]	Consolidation	Modified best-fit decreasing placement	Fewer active hosts and lower energy	Static thresholds in base form
[13]	Consolidation	Adaptive overload thresholds, VM selection	Better energy–SLA trade-off on real traces	Simulation-based evaluation
[15]	Scheduling	Carbon- and cooling-aware multi-site placement	Lower emission with deadlines met	Focused on HPC batch jobs
[16]	Scheduling	Packet-level energy simulation	Network energy is significant	Simulation only
[8]	Carbon-aware	Temporal shifting by grid carbon forecast	Flexible load moved to cleaner hours	Only delay-tolerant jobs benefit

© 2026 IJSRET

IV. PROPOSED CONCEPTUAL FRAMEWORK

Based on the reviewed literature, this section describes a layered green-cloud management framework intended for academic study and simulation. Its structure is shown in Fig. 2.

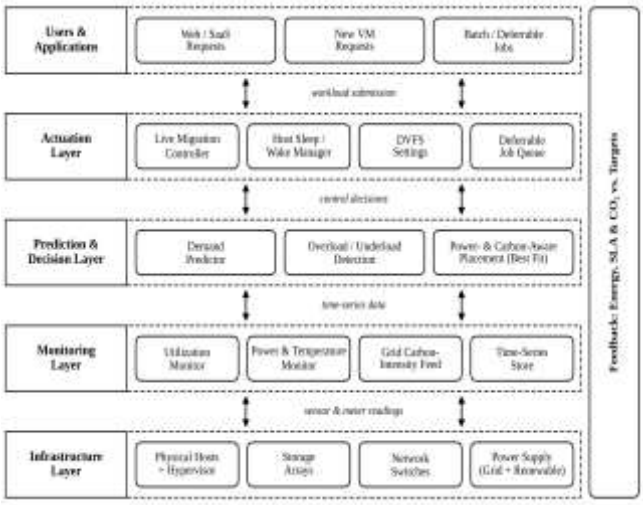


Fig. 2. Layered architecture of the proposed green-cloud management framework.

1. Infrastructure Layer
The lowest layer contains the physical hosts, storage and network equipment together with power meters and temperature sensors. Each host exposes CPU, memory and network utilization through the hypervisor and its instantaneous power through an intelligent power-distribution unit or a model such as (2).

2. Monitoring Layer
The monitoring module samples utilization, power and inlet temperature at a fixed interval , for example every five minutes, and also obtains the carbon intensity of the electricity supply from the grid operator or an on-site renewable controller. These readings are stored as time series.

3. Prediction And Decision Layer
A demand estimator forecasts the utilization of each VM for the next interval. Simple moving averages or regression are sufficient for an academic prototype. The decision engine then performs three tasks. First, it detects overloaded hosts whose predicted utilization exceeds an upper threshold and selects VMs to move away from them. Second, it detects underloaded hosts whose VMs can all be accommodated elsewhere, so that the host can be switched to sleep. Third, it places new and migrating VMs using a power-aware best-fit rule that minimises the increase in power, with the carbon intensity used as a tie-breaker or as an extra cost term when several sites are available [9], [15].

4. Actuation Layer
Decisions are carried out through the cloud controller: live migration commands, host sleep or wake commands, and DVFS settings. A deferrable-job queue holds batch work that can wait, and releases it when renewable supply is high or carbon intensity is low [8]. A feedback loop compares the achieved energy and SLA figures with targets and adjusts thresholds.

5. Comparative Analysis
Table II compares the four technique families on several qualitative criteria. Hardware and facility measures give steady savings but need investment. Consolidation offers large savings with no new hardware, but carries a risk to performance. Carbon-aware scheduling may not reduce energy at all, yet it reduces emission by changing when and where energy is used. Renewable integration lowers emission directly but depends on location and weather.

Table Ii Qualitative Comparison Of Green Cloud Technique Families

Criterion	Hardware / Facility	VM Consolidation	Energy/Carbon-Aware Scheduling	Renewable Integration
Main target	PUE, idle power	Active host count	Energy and CO2 per job	Grid CO2 intensity
Capital cost	High	Low	Low	High
Performance risk	Low	Medium to high	Low to medium	Low
Control location	Operator	Cloud manager	Scheduler	Operator and scheduler
Typical evaluation	Measurement	Simulation, trace replay	Simulation, deployment	Deployment, modelling
Key references	[5], [12]	[9], [13]	[8], [15]	[6]

© 2026 IJSRET

V. EVALUATION METHODOLOGY

Because experiments on a production cloud are rarely possible for students, two evaluation routes are suitable. The first is simulation using CloudSim, which models hosts, VMs, power models and allocation policies, and which has been used in many of the reviewed works [14]. GreenCloud can be used when network energy must be included [16]. The second route is a small test bed built from a few computers running an open hypervisor, with a plug-in power meter on each host. In both cases the proposed policy should be compared with a baseline, such as round-robin placement without consolidation, using total energy in kilowatt-hours, number of migrations, average host utilization, SLA violation percentage and estimated CO_2 emission.

VI. CHALLENGES AND FUTURE DIRECTIONS

Several issues remain open. Migration overhead is the first: every live migration transfers memory pages over the network and briefly degrades the moved VM, so frequent rebalancing can cancel the expected savings [11]. Second, prediction errors lead either to overload, when demand is underestimated, or to wasted capacity, when it is overestimated. Third, accurate carbon accounting is difficult because the carbon intensity of electricity changes by region and by hour, and providers do not always expose it to tenants [8]. Fourth, a narrow focus on PUE can hide inefficiency inside the IT equipment; metrics that combine facility and server efficiency are needed [3]. Finally, the rising share of AI training and inference workloads, which run on power-hungry accelerators, requires new scheduling models that consider GPU power states. Future research is expected to combine machine-learning-based prediction, carbon-aware placement across regions and on-site storage into unified control systems.

VII. CONCLUSION

This survey has examined green cloud computing as a layered set of techniques that reduce the energy and carbon cost of cloud services. Hardware and facility improvements lower overhead, virtualization and consolidation increase utilization and allow idle machines to sleep, energy- and carbon-aware scheduling moves work to the most suitable place and time, and renewable integration cleans the supply itself. The reviewed studies show that the largest benefits come from combining these layers. The conceptual framework presented here brings monitoring, prediction and energy-aware placement together and can be evaluated through simulation or a small test bed, making it a practical basis for further academic work in this area.

REFERENCES

1. P. Mell and T. Grance, "The NIST definition of cloud computing," National Institute of Standards and Technology, Gaithersburg, MD, USA, Special Publication 800-145, 2011.
2. A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, "The cost of a cloud: Research problems in data center networks," ACM SIGCOMM Computer Communication Review, vol. 39, no. 1, pp. 68–73, 2009.
3. M. Dayarathna, Y. Wen, and R. Fan, "Data center energy consumption modeling: A survey," IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 732–794, 2016.
4. E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, "Recalibrating global data center energy-use estimates," Science, vol. 367, no. 6481, pp. 984–986, 2020.
5. L. A. Barroso and U. Hölzle, "The case for energy-proportional computing," Computer, vol. 40, no. 12, pp. 33–37, 2007.
6. T. Mastelic, A. Oleksiak, H. Claussen, I. Brandic, J.-M. Pierson, and A. V. Vasilakos, "Cloud computing: Survey on energy efficiency," ACM Computing Surveys, vol. 47, no. 2, Art. no. 33, 2015.
7. The Green Grid, "PUE: A comprehensive examination of the metric," White Paper no. 49, 2012.
8. A. Radovanović et al., "Carbon-aware computing for datacenters," IEEE Transactions on Power Systems, vol. 38, no. 2, pp. 1270–1280, 2023.
9. A. Beloglazov, J. Abawajy, and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing," Future Generation Computer Systems, vol. 28, no. 5, pp. 755–768, 2012.
10. P. Barham et al., "Xen and the art of virtualization," in Proc. 19th ACM Symp. Operating Systems Principles (SOSP), 2003, pp. 164–177.
11. C. Clark et al., "Live migration of virtual machines," in Proc. 2nd USENIX Symp. Networked Systems Design and Implementation (NSDI), 2005, pp. 273–286.
12. D. Meisner, B. T. Gold, and T. F. Wenisch, "PowerNap: Eliminating server idle power," in Proc. 14th Int. Conf. Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2009, pp. 205–216.
13. A. Beloglazov and R. Buyya, "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, vol. 24, no. 13, pp. 1397–1420, 2012.

© 2026 IJSRET

14. R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, and R. Buyya, "CloudSim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms," Software: Practice and Experience, vol. 41, no. 1, pp. 23–50, 2011.
15. S. K. Garg, C. S. Yeo, A. Anandasivam, and R. Buyya, "Environment-conscious scheduling of HPC applications on distributed cloud-oriented data centers," Journal of Parallel and Distributed Computing, vol. 71, no. 6, pp. 732–749, 2011.
16. D. Kliazovich, P. Bouvry, and S. U. Khan, "GreenCloud: A packet-level simulator of energy-aware cloud computing data centers," The Journal of Supercomputing, vol. 62, no. 3, pp. 1263–1283, 2012.

© 2026 IJSRET