
| RESEARCH ARTICLE

Green Cloud Computing: Reducing the Carbon Footprint of Microsoft Azure Workloads

Shekar Rao Lakavath

Independent Researcher, Bonneylake, Washington, USA

Corresponding Author: Shekar Rao Lakavath, **E-mail:** Shekar.lakavath@gmail.com

| ABSTRACT

Data centers underpinning cloud computing consume a substantial and growing share of global electricity, and the resulting carbon emissions have become a material sustainability concern for organizations that operate at cloud scale. Green cloud computing addresses this concern by combining resource-efficiency techniques, such as workload consolidation and right-sizing, with carbon-aware practices that account for when and where computation occurs relative to the availability of low-carbon electricity. This article examines how these techniques can be applied to reduce the carbon footprint of workloads hosted on Microsoft Azure. The article proposes a carbon-aware workload management architecture spanning grid carbon intensity signal ingestion, workload characterization, a central scheduling layer, and three green optimization actions, namely regional placement, right-sizing and consolidation, and deferred scheduling, compares six green computing techniques against their underlying principles and Azure implementation, and maps common sustainability challenges to specific mitigation approaches. Two tables and one architecture figure illustrate this analysis. The findings indicate that resource-efficiency techniques such as consolidation and autoscaling remain the most mature and broadly deployable levers available to Azure customers today, while carbon-aware temporal and spatial workload shifting, though demonstrated to meaningfully reduce emissions in research settings, requires workloads with sufficient scheduling flexibility to apply effectively. The article concludes with practical recommendations for organizations seeking to reduce the carbon footprint of their Azure workloads.

| KEYWORDS

Green cloud computing, carbon footprint, Microsoft Azure, energy efficiency, carbon-aware computing, virtual machine consolidation, sustainability

| ARTICLE INFORMATION

ACCEPTED: June 28, 2024

PUBLISHED: August 30, 2024

DOI: <https://doi.org/10.61424/jcsit.v1i1.1076>

1. Introduction

The rapid growth of cloud computing has been accompanied by a corresponding rise in the electricity consumption of the data centers that underpin it. Early research into this relationship established that data centers were, even a decade ago, responsible for a significant and growing share of global electricity use, driven by both the direct energy demand of computing hardware and the substantial cooling and power-distribution overhead required to keep that hardware operating reliably (Baliga, Ayre, Hinton, & Tucker, 2011). As organizations have shifted a growing share of their computing workloads to public cloud providers such as Microsoft Azure, responsibility for managing this energy and carbon footprint has become distributed between the cloud provider, which controls physical data center design and energy sourcing, and the customer, whose workload placement, sizing, and scheduling decisions materially affect how efficiently that underlying infrastructure is used.

Green cloud computing research has historically emphasized two complementary strategies. The first, resource efficiency, focuses on reducing the total computing capacity required to serve a given workload, primarily through virtual machine consolidation and dynamic right-sizing (Beloglazov, Abawajy, & Buyya, 2012). The second, more recent strategy, carbon-aware computing, focuses not merely on how much energy is consumed but on when and where that energy is drawn from the electricity grid, shifting workloads in time or space toward periods and regions with a cleaner generation mix (Radovanović et al., 2023). Both strategies are directly actionable by Azure customers without requiring changes to Microsoft's underlying data center infrastructure.

This article addresses three questions. First, what techniques has the literature established for reducing the energy consumption and carbon footprint of cloud computing workloads, and how have these techniques evolved from early resource-efficiency research to more recent carbon-aware scheduling approaches? Second, how can these techniques be organized into a coherent architecture for managing Azure workloads sustainably? Third, which specific green computing technique addresses which category of sustainability challenge, and what are the practical tradeoffs involved? The remainder of the article reviews the relevant literature, proposes a carbon-aware workload management architecture for Azure, compares green computing techniques against their Azure implementation, maps sustainability challenges to mitigations, and discusses practical implications for organizations operating Azure workloads.

2. Literature Review

2.1 Foundations of Energy-Efficient Cloud Computing

Baliga, Ayre, Hinton, and Tucker (2011) provided one of the most widely cited early analyses of energy consumption across the full cloud computing stack, decomposing energy use into processing, storage, and network transport components and finding that network transport energy could rival or exceed data center processing energy for certain cloud service categories, a finding with direct implications for how workload placement decisions affect total energy consumption. Beloglazov, Buyya, Lee, and Zomaya (2011) produced a comprehensive taxonomy and survey of energy-efficient data centers and cloud computing systems, organizing the field according to whether techniques operate at the hardware, virtualization, or scheduling layer, and identifying dynamic virtual machine consolidation as the technique with the most immediately actionable potential for cloud customers.

Garg, Yeo, Anandasivam, and Buyya (2011) extended this foundation with an environment-conscious scheduling approach specifically for high-performance computing applications on distributed cloud data centers, demonstrating that incorporating carbon-footprint awareness directly into job scheduling algorithms could reduce emissions without proportionally increasing job completion time. Uddin and Rahman (2012) proposed a broader green information technology framework for data centers, arguing that energy efficiency and carbon reduction should be treated as an explicit design objective across data center facilities, hardware selection, and workload management, rather than emerging only as a byproduct of cost optimization.

2.2 Virtual Machine Consolidation and Resource Allocation

A substantial body of research has focused specifically on algorithms for consolidating virtual machines onto fewer physical hosts to reduce idle capacity. Beloglazov, Abawajy, and Buyya (2012) proposed a set of energy-aware resource allocation heuristics for cloud data centers, demonstrating that dynamic consolidation guided by adaptive utilization thresholds could substantially reduce energy consumption relative to static allocation while maintaining acceptable service-level performance. Beloglazov and Buyya (2012) extended this work with optimal online deterministic algorithms and adaptive heuristics specifically for the dynamic consolidation problem, formally characterizing the tradeoff between energy savings and the performance degradation risk introduced by aggressive consolidation.

Beloglazov and Buyya (2013) further addressed the specific challenge of managing overloaded hosts during dynamic consolidation under explicit quality-of-service constraints, proposing detection and migration strategies that reduce the risk of consolidation-induced performance violations. This body of work collectively established virtual machine consolidation as one of the most mature and widely implemented green cloud computing

techniques, directly reflected in the autoscaling and automated shutdown capabilities that Azure and other major cloud platforms now provide natively.

2.3 Carbon-Aware Workload Scheduling

More recent research has extended energy-efficiency techniques to explicitly account for the carbon intensity of electricity, which varies significantly by time of day and by region depending on the underlying generation mix. Khosravi, Andrew, and Buyya (2017) proposed a dynamic virtual machine placement method specifically designed to minimize combined energy and carbon cost across geographically distributed cloud data centers, demonstrating that carbon-aware placement decisions could achieve meaningful emissions reductions even without any reduction in total energy consumption, simply by shifting where that energy was drawn from the grid.

Wiesner, Behnke, Scheinert, Gontarska, and Thamsen (2021) proposed a temporal workload-shifting approach demonstrating that delaying delay-tolerant cloud workloads to periods of lower grid carbon intensity could meaningfully reduce emissions without violating typical service-level objectives, a finding that directly informs the deferred-scheduling component of the architecture proposed later in this article. Wiesner, Scheinert, Wittkopp, Thamsen, and Kao (2022) extended this line of research with a renewable-aware admission control mechanism specifically designed for delay-tolerant cloud and edge workloads, allowing a scheduling system to distinguish between latency-sensitive requests that must be served immediately and delay-tolerant requests that can be deferred to more carbon-favorable periods.

Radovanović et al. (2023) described a large-scale, production carbon-aware computing system, demonstrating that shifting the timing of flexible compute load in response to real-time grid carbon-intensity forecasts produced measurable emissions reductions across a portfolio of data center clusters, providing empirical validation, at commercial operating scale, of the temporal shifting principles established in earlier academic research.

2.4 Data Center Energy and Carbon Measurement

Accurately measuring energy consumption and associated carbon emissions has been a persistent methodological challenge underlying green cloud computing research. Dayarathna, Wen, and Fan (2016) conducted a comprehensive survey of data center energy consumption modeling approaches, finding substantial variation in model granularity and accuracy across the literature and cautioning that energy-saving claims should be interpreted carefully when the underlying measurement methodology is not clearly specified. Shuja et al. (2016) surveyed techniques and architectures specifically for designing energy-efficient data centers, organizing the field according to hardware, cooling, and workload management interventions and highlighting cooling system optimization as a comparatively underexploited lever relative to server-level power management.

Masanet, Shehabi, Lei, Smith, and Koomey (2020) provided an influential recalibration of global data center energy-use estimates, finding that despite a roughly six-fold increase in compute instances between 2010 and 2018, global data center energy consumption grew only modestly over the same period, a finding they attribute substantially to the efficiency gains from virtualization and consolidation techniques such as those documented by Beloglazov and Buyya. Shehabi, Smith, Masanet, and Koomey (2018) reached a similar conclusion specifically for United States data centers, finding that the shift toward hyperscale cloud data centers, which operate at higher average utilization and better power usage effectiveness than smaller enterprise data centers, was a primary driver of this decoupling between service growth and energy growth.

Jin, Bai, Yang, Mao, and Xu (2020) reviewed power consumption modeling approaches for data center servers specifically, providing a technical foundation for the utilization-to-power mapping that underlies most consolidation and carbon-aware scheduling algorithms. Liu et al. (2020) extended this measurement focus specifically to emissions, proposing a prediction model linking data center network traffic and power usage effectiveness to energy consumption and associated carbon emissions at a global scale.

2.5 Broader Sustainability Frameworks and Standards

Beyond individual technical interventions, researchers have proposed broader frameworks for organizing green computing efforts. Orgerie, Assunção, and Lefèvre (2014) surveyed techniques for improving the energy efficiency

of large-scale distributed systems more broadly, situating cloud-specific consolidation and scheduling techniques within a wider set of approaches spanning grid computing, high-performance computing, and large-scale distributed systems generally. Mastelic et al. (2014) conducted a widely cited survey specifically on energy efficiency in cloud computing, proposing a taxonomy spanning hardware, virtualization, and software-level techniques that has served as a reference framework for subsequent green cloud computing research. Zakarya and Gillam (2017) extended this taxonomy specifically to energy-efficient computing across clusters, grids, and clouds, emphasizing that energy efficiency techniques developed for one computing paradigm often transfer with modification to another.

Koronen, Åhman, and Nilsson (2020) examined data centers within the broader context of future European energy systems, arguing that data center energy efficiency and demand flexibility should be integrated into energy system planning and policy rather than treated purely as an internal information-technology optimization problem. Acun et al. (2022) proposed a holistic approach to carbon-aware data center design, arguing that operational carbon-aware scheduling, of the kind demonstrated by Wiesner et al. and Radovanović et al., should be combined with embodied-carbon-aware hardware procurement and standardized emissions accounting to address the full carbon lifecycle of cloud infrastructure rather than operational emissions alone.

Taken together, this literature indicates that reducing the carbon footprint of cloud workloads requires combining resource-efficiency techniques, which are mature and already substantially embedded in commercial cloud platforms, with carbon-aware temporal and spatial scheduling techniques, which are more recent, require workload scheduling flexibility to apply effectively, but have been shown to produce meaningful additional emissions reductions at both research and production scale. The following sections apply this combined understanding specifically to Microsoft Azure workloads.

3. A Carbon-Aware Workload Management Architecture for Azure

Building on the combined resource-efficiency and carbon-awareness approach established in the literature (Beloglazov, Abawajy, & Buyya, 2012; Radovanović et al., 2023), this article proposes a workload management architecture for reducing the carbon footprint of Azure workloads. Figure 1 illustrates the architecture, in which grid carbon intensity signals, workload characteristics, and Azure resource telemetry feed a central carbon-aware scheduler, which coordinates three green optimization actions.

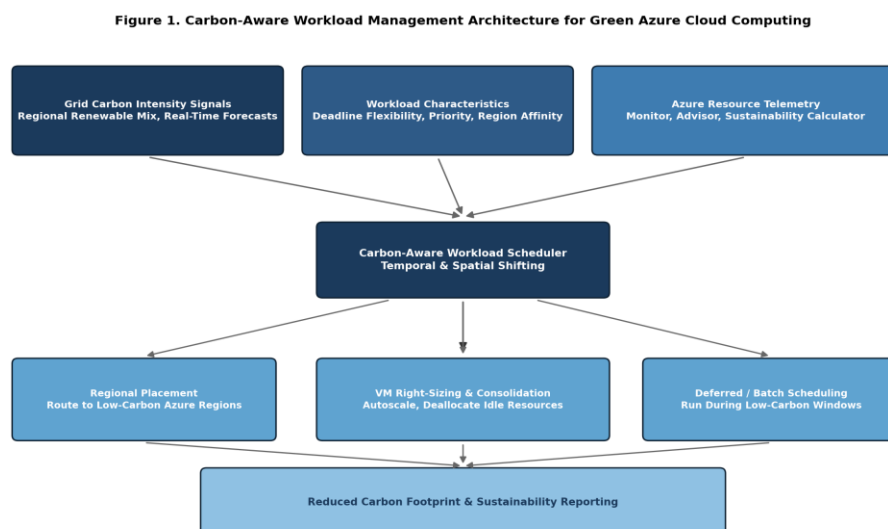


Figure 1. Carbon-aware workload management architecture for green Azure cloud computing, showing how grid carbon intensity signals, workload characteristics, and Azure resource telemetry feed the central scheduler, which coordinates regional placement, right-sizing, and deferred scheduling to reduce the carbon footprint of Azure workloads.

3.1 Signal Ingestion

The scheduler ingests three categories of signal: regional grid carbon intensity and renewable mix forecasts, workload characteristics including deadline flexibility and region affinity, and Azure resource telemetry from Azure Monitor, Advisor, and the Sustainability Calculator. This multi-signal design reflects the combination of grid-awareness and workload-awareness that Wiesner, Scheinert, Wittkopp, Thamsen, and Kao (2022) identify as necessary for effective renewable-aware admission control.

3.2 Carbon-Aware Scheduling

The central scheduler determines, for each workload, whether it should run immediately in its current location, be shifted temporally to a lower-carbon window, or be shifted spatially to a lower-carbon Azure region, applying the temporal and spatial shifting principles validated by Khosravi, Andrew, and Buyya (2017) and Wiesner, Behnke, Scheinert, Gontarska, and Thamsen (2021).

3.3 Green Optimization Actions

Three coordinated actions implement the scheduler's decisions: regional placement routes workloads toward Azure regions with a cleaner electricity mix; right-sizing and consolidation applies the dynamic VM consolidation heuristics established by Beloglazov and Buyya (2012) to minimize active capacity; and deferred or batch scheduling delays delay-tolerant workloads to lower-carbon windows, following the admission-control approach of Wiesner, Scheinert, Wittkopp, Thamsen, and Kao (2022).

3.4 Measurement and Reporting

Outcomes feed into sustainability measurement and reporting, addressing the visibility gap that Masanet, Shehabi, Lei, Smith, and Koomey (2020) and Shehabi, Smith, Masanet, and Koomey (2018) identify as necessary for organizations to track whether efficiency gains are keeping pace with growth in cloud workload demand.

4. Comparing Green Computing Techniques and Their Azure Implementation

Table 1 compares six green computing techniques established in the literature against their underlying principle and corresponding Azure implementation.

Table 1. Green Computing Techniques Compared Against Underlying Principles and Azure Implementation

Green Computing Technique	Underlying Principle	Azure Implementation
Dynamic VM consolidation	Migrate workloads onto fewer active hosts and power down idle servers during low-utilization periods	Azure VM auto-shutdown, Azure Automation, Azure Kubernetes Service cluster autoscaler
Right-sizing and autoscaling	Match provisioned capacity to actual demand rather than static peak-based allocation	Azure Monitor autoscale, Azure Advisor cost and sustainability recommendations
Carbon-aware workload scheduling	Shift delay-tolerant workloads in time to periods of higher grid renewable supply	Azure Batch scheduling policies, custom Azure Functions triggered by carbon-intensity APIs
Carbon-aware regional placement	Route workloads to Azure regions with a higher share of renewable or low-carbon electricity	Azure Front Door / Traffic Manager region routing, multi-region deployment templates

Sustainability measurement and reporting	Quantify operational and embodied carbon to identify optimization targets and track progress	Microsoft Sustainability Manager, Emissions Impact Dashboard
Efficient data center design	Improve power usage effectiveness (PUE) through cooling, hardware, and airflow optimization	Microsoft-operated data center infrastructure (outside direct customer control, but reflected in regional carbon intensity)

As Table 1 indicates, resource-efficiency techniques such as consolidation and right-sizing are directly actionable through native Azure tooling, while carbon-aware scheduling and regional placement require additional custom logic layered on top of Azure's platform capabilities, consistent with the more recent and less standardized status of carbon-aware scheduling research relative to the mature virtual machine consolidation literature (Beloglazov & Buyya, 2012; Radovanović et al., 2023).

5. Mapping Sustainability Challenges to Mitigation Approaches

Table 2 synthesizes the sustainability challenges most frequently documented in the literature reviewed above and maps each to a specific mitigation approach, along with representative supporting sources.

Table 2. Green Cloud Computing Challenges and Mitigation Approaches

Green Cloud Computing Challenge	Mitigation Approach	Supporting Literature
Static resource allocation leads to persistent over-provisioning and idle capacity	Dynamic VM consolidation and threshold-based host deactivation reduce active server count during low demand	Beloglazov, Abawajy, & Buyya (2012); Beloglazov & Buyya (2012)
Workload timing does not account for variation in grid carbon intensity	Temporal shifting of delay-tolerant workloads to low-carbon-intensity windows	Wiesner, Behnke, Scheinert, Gontarska, & Thamsen (2021); Radovanović et al. (2023)
Geographic distribution of data centers ignores regional differences in renewable energy availability	Carbon-aware placement routes workloads toward regions with a cleaner electricity mix	Khosravi, Andrew, & Buyya (2017)
Organizations lack visibility into the carbon impact of their own cloud usage	Automated sustainability measurement and reporting tools quantify operational carbon at the workload level	Masanet, Shehabi, Lei, Smith, & Koomey (2020); Shehabi, Smith, Masanet, & Koomey (2018)
Admission control does not distinguish delay-tolerant from latency-sensitive workloads	Renewable-aware admission control defers non-urgent jobs without violating service commitments	Wiesner, Scheinert, Wittkopp, Thamsen, & Kao (2022)

Rising overall data center energy demand offsets efficiency gains from individual optimizations	Combining consolidation, scheduling, and efficient hardware design to constrain aggregate energy growth	Jin, Bai, Yang, Mao, & Xu (2020); Dayarathna, Wen, & Fan (2016)
Carbon accounting methods vary, making cross-provider or cross-workload comparison difficult	Standardized emissions measurement frameworks applied consistently across regions and workloads	Acun, Lee, Maeng, Chakkaravarthy, Gupta, Brooks, & Wu (2022)

The mapping in Table 2 illustrates that resource-efficiency and carbon-awareness address distinct but complementary failure modes: consolidation reduces the total capacity an organization must operate, while carbon-aware scheduling and placement reduce the emissions intensity of the capacity that remains in operation, consistent with the holistic, full-lifecycle framing that Acun et al. (2022) argue is necessary to address carbon impact comprehensively.

6. Discussion

The analysis presented above supports several practical conclusions for organizations seeking to reduce the carbon footprint of their Azure workloads. First, resource-efficiency techniques remain the most mature, broadly applicable, and immediately actionable lever available; the virtual machine consolidation algorithms established by Beloglazov, Abawajy, and Buyya (2012) more than a decade ago are directly reflected in Azure's native autoscaling and automated shutdown capabilities, meaning organizations can capture substantial efficiency gains without custom engineering effort simply by enabling and correctly configuring these existing platform features.

Second, carbon-aware temporal and spatial scheduling offers meaningful additional emissions reductions but depends on workloads having genuine scheduling flexibility. Wiesner, Behnke, Scheinert, Gontarska, and Thamsen's (2021) temporal shifting results and Khosravi, Andrew, and Buyya's (2017) spatial placement results both apply most directly to batch, delay-tolerant, or non-latency-critical workloads; organizations should inventory which of their Azure workloads genuinely tolerate deferral or geographic flexibility before investing in carbon-aware scheduling infrastructure, since applying it to latency-sensitive, customer-facing workloads risks service-level violations for comparatively modest emissions benefit.

Third, measurement and reporting are prerequisites for informed optimization, not optional add-ons. Masanet, Shehabi, Lei, Smith, and Koomey's (2020) finding that global data center energy growth decoupled from compute growth was only demonstrable because sufficiently detailed measurement data existed to support the analysis; organizations that lack equivalent visibility into their own Azure carbon footprint cannot reliably determine whether their optimization efforts are producing the intended effect, underscoring the practical importance of tools such as the Emissions Impact Dashboard and Microsoft Sustainability Manager.

Fourth, the scale at which carbon-aware scheduling operates matters. Radovanović et al.'s (2023) production deployment demonstrates that meaningful emissions reductions are achievable at commercial operating scale, but their results also depend on access to real-time, granular grid carbon-intensity data that may not be equally available or accurate across all Azure regions; organizations operating primarily in regions with less mature grid carbon-intensity reporting infrastructure should expect correspondingly less precise carbon-aware scheduling decisions.

Finally, addressing carbon impact comprehensively requires looking beyond operational energy consumption alone. Acun et al.'s (2022) holistic framing, encompassing embodied carbon from hardware manufacturing alongside operational carbon from electricity consumption, suggests that organizations focused exclusively on workload scheduling and consolidation are addressing only part of their cloud computing carbon footprint, and that

hardware refresh cycles and instance-type selection also carry a carbon cost that pure operational optimization does not capture.

7. Recommendations

Organizations seeking to reduce the carbon footprint of their Azure workloads should: (1) enable and correctly configure native Azure autoscaling, automated VM shutdown, and Azure Advisor sustainability recommendations as a first, low-effort step, capturing the resource-efficiency gains established by Beloglazov, Abawajy, and Buyya (2012); (2) inventory workloads by scheduling flexibility to identify which are genuine candidates for temporal or spatial carbon-aware shifting, following the admission-control approach of Wiesner, Scheinert, Wittkopp, Thamsen, and Kao (2022); (3) route delay-tolerant, non-latency-critical workloads toward Azure regions with a cleaner electricity mix where regulatory and data-residency requirements permit, applying the placement principles of Khosravi, Andrew, and Buyya (2017); (4) deploy Microsoft Sustainability Manager or the Emissions Impact Dashboard to establish a measurement baseline before making optimization claims, addressing the visibility gap identified by Masanet, Shehabi, Lei, Smith, and Koomey (2020); (5) treat carbon-aware scheduling as complementary to, not a substitute for, resource-efficiency techniques, since the two address distinct components of total carbon impact; and (6) consider the embodied carbon implications of instance-type and hardware-generation selection alongside operational scheduling decisions, following the holistic framing proposed by Acun et al. (2022).

8. Conclusion

Reducing the carbon footprint of cloud computing workloads requires combining two complementary strategies: resource-efficiency techniques that minimize the total computing capacity required to serve a workload, and carbon-aware scheduling techniques that shift computation in time and space toward periods and regions of cleaner electricity generation. This article has reviewed the evolution of green cloud computing research from foundational energy-efficiency and virtual machine consolidation work through more recent carbon-aware scheduling and measurement research, proposed a carbon-aware workload management architecture for Azure, and mapped specific sustainability challenges to their corresponding mitigation techniques. The literature reviewed indicates that resource-efficiency techniques are mature and already substantially available through native Azure tooling, while carbon-aware temporal and spatial scheduling, though demonstrated to produce meaningful additional emissions reductions at both research and commercial scale, requires organizations to first identify which workloads have sufficient scheduling flexibility to benefit. Future research should evaluate the combined emissions impact of resource-efficiency and carbon-aware scheduling techniques applied jointly across representative Azure workload portfolios, and should extend measurement frameworks to capture embodied as well as operational carbon across the full cloud computing lifecycle.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers

Artificial Intelligence (AI) Use Disclosure: The authors declare that no artificial intelligence tools were used in the preparation of this manuscript.

References

- [1] Acun, B., Lee, B., Maeng, K., Chakkaravarthy, M., Gupta, U., Brooks, D., & Wu, C.-J. (2022). A holistic approach for designing carbon aware datacenters (arXiv:2201.10036). arXiv. <https://doi.org/10.48550/arXiv.2201.10036>
- [2] Baliga, J., Ayre, R. W. A., Hinton, K., & Tucker, R. S. (2011). Green cloud computing: Balancing energy in processing, storage, and transport. *Proceedings of the IEEE*, 99(1), 149–167. <https://doi.org/10.1109/JPROC.2010.2060451>
- [3] Beloglazov, A., Abawajy, J., & Buyya, R. (2012). Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. *Future Generation Computer Systems*, 28(5), 755–768. <https://doi.org/10.1016/j.future.2011.04.017>
- [4] 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. <https://doi.org/10.1002/cpe.1867>

- [5] Beloglazov, A., & Buyya, R. (2013). Managing overloaded hosts for dynamic consolidation of virtual machines in cloud data centers under quality of service constraints. *IEEE Transactions on Parallel and Distributed Systems*, 24(7), 1366–1379. <https://doi.org/10.1109/TPDS.2012.240>
- [6] Beloglazov, A., Buyya, R., Lee, Y. C., & Zomaya, A. (2011). A taxonomy and survey of energy-efficient data centers and cloud computing systems. *Advances in Computers*, 82, 47–111. <https://doi.org/10.1016/B978-0-12-385512-1.00003-7>
- [7] Dayarathna, M., Wen, Y., & Fan, R. (2016). Data center energy consumption modeling: A survey. *IEEE Communications Surveys & Tutorials*, 18(1), 732–794. <https://doi.org/10.1109/COMST.2015.2481183>
- [8] Garg, S. K., Yeo, C. S., Anandasivam, A., & Buyya, R. (2011). Environment-conscious scheduling of HPC applications on distributed cloud-oriented data centers. *Journal of Parallel and Distributed Computing*, 71(6), 732–749. <https://doi.org/10.1016/j.jpdc.2010.04.004>
- [9] Jin, C., Bai, X., Yang, C., Mao, W., & Xu, X. (2020). A review of power consumption models of servers in data centers. *Applied Energy*, 265, 114806. <https://doi.org/10.1016/j.apenergy.2020.114806>
- [10] Khosravi, A., Andrew, L. L. H., & Buyya, R. (2017). Dynamic VM placement method for minimizing energy and carbon cost in geographically distributed cloud data centers. *IEEE Transactions on Sustainable Computing*, 2(2), 183–196. <https://doi.org/10.1109/TSUSC.2017.2709980>
- [11] Koronen, C., Åhman, M., & Nilsson, L. J. (2020). Data centres in future European energy systems—energy efficiency, integration and policy. *Energy Efficiency*, 13(1), 129–144. <https://doi.org/10.1007/s12053-019-09833-8>
- [12] Liu, Y., Wei, X., Xiao, J., Liu, Z., Xu, Y., & Tian, Y. (2020). Energy consumption and emission mitigation prediction based on data center traffic and PUE for global data centers. *Global Energy Interconnection*, 3(3), 272–282. <https://doi.org/10.1016/j.gloi.2020.07.008>
- [13] Masanet, E., Shehabi, A., Lei, N., Smith, S., & Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science*, 367(6481), 984–986. <https://doi.org/10.1126/science.aba3758>
- [14] Mastelic, T., Oleksiak, A., Claussen, H., Brandic, I., Pierson, J.-M., & Vasilakos, A. V. (2014). Cloud computing: Survey on energy efficiency. *ACM Computing Surveys*, 47(2), 1–36. <https://doi.org/10.1145/2656204>
- [15] Orgerie, A.-C., Assunção, M. D. de, & Lefèvre, L. (2014). A survey on techniques for improving the energy efficiency of large-scale distributed systems. *ACM Computing Surveys*, 46(4), 47. <https://doi.org/10.1145/2532637>
- [16] Radovanović, A., Koningstein, R., Schneider, I., Chen, B., Duarte, A., Roy, B., Xiao, D., Haridasan, M., Hung, P., Care, N., Talukdar, S., Mullen, E., Smith, K., Cottman, M., & Cirne, W. (2023). Carbon-aware computing for datacenters. *IEEE Transactions on Power Systems*, 38(2), 1270–1280. <https://doi.org/10.1109/TPWRS.2022.3173250>
- [17] Shehabi, A., Smith, S. J., Masanet, E., & Koomey, J. (2018). Data center growth in the United States: Decoupling the demand for services from electricity use. *Environmental Research Letters*, 13(12), 124030. <https://doi.org/10.1088/1748-9326/aaec9c>
- [18] Shuja, J., Bilal, K., Madani, S. A., Othman, M., Ranjan, R., Balaji, P., & Khan, S. U. (2016). Survey of techniques and architectures for designing energy-efficient data centers. *IEEE Systems Journal*, 10(2), 507–519. <https://doi.org/10.1109/JSYST.2014.2315823>
- [19] Uddin, M., & Rahman, A. A. (2012). Energy efficiency and low carbon enabler green IT framework for data centers considering green metrics. *Renewable and Sustainable Energy Reviews*, 16(6), 4078–4094. <https://doi.org/10.1016/j.rser.2012.03.014>
- [20] Whitehead, B., Andrews, D., Shah, A., & Maidment, G. (2014). Assessing the environmental impact of data centres part 1: Background, energy use and metrics. *Building and Environment*, 82, 151–159. <https://doi.org/10.1016/j.buildenv.2014.07.020>
- [21] Wiesner, P., Behnke, I., Scheinert, D., Gontarska, K., & Thamsen, L. (2021). Let's wait awhile: How temporal workload shifting can reduce carbon emissions in the cloud. In *Proceedings of the 22nd International Middleware Conference* (pp. 260–272). ACM. <https://doi.org/10.1145/3464298.3493399>
- [22] Wiesner, P., Scheinert, D., Wittkopp, T., Thamsen, L., & Kao, O. (2022). Cucumber: Renewable-aware admission control for delay-tolerant cloud and edge workloads. In *European Conference on Parallel Processing (Euro-Par 2022)* (pp. 218–233). Springer. https://doi.org/10.1007/978-3-031-12597-3_14
- [23] Zakarya, M., & Gillam, L. (2017). Energy efficient computing, clusters, grids and clouds: A taxonomy and survey. *Sustainable Computing: Informatics and Systems*, 14, 13–33. <https://doi.org/10.1016/j.suscom.2017.03.002>