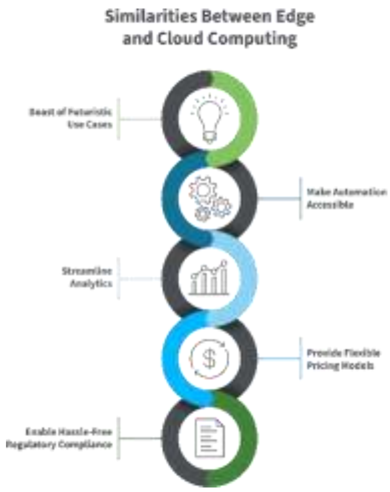
"A Comparative Study of Energy Efficiency in Cloud and Edge Computing Paradigms"

Tanushri Kadam, Safi Ansari

Instrumentation Engineering, All India Shri Shivaji Memorial Society's Institute of Information Technology, Pune, Maharashtra, India

ABSTRACT

As digital infrastructures evolve to support increasingly data-intensive applications, energy efficiency has become a critical metric in evaluating computing paradigms. This study provides a comprehensive comparison of energy consumption in cloud and edge computing environments. Cloud computing centralizes resources in large-scale data centers, offering high computational power but often incurring significant energy costs due to data transmission and centralized processing. In contrast, edge computing brings computation closer to data sources, potentially reducing latency and network energy consumption by offloading tasks to local devices. Through a systematic analysis of existing literature, performance metrics, and case studies across various application domains—including IoT, real-time analytics, and smart cities—this paper identifies the conditions under which each paradigm excels in energy efficiency. The findings highlight that while cloud computing remains superior for resource-intensive and large-scale data aggregation tasks, edge computing offers marked energy advantages in latency-sensitive and distributed scenarios. The study concludes by proposing hybrid architectures as a promising direction for optimizing energy usage across diverse workloads.



1. INTRODUCTION

The rapid proliferation of connected devices and data-intensive applications has transformed the landscape of computing. From smart homes and autonomous vehicles to industrial automation and real-time analytics, modern digital services demand not only high computational power but also responsiveness and sustainability. Traditionally, cloud computing has served as the backbone of this digital transformation, offering scalable, centralized processing capabilities through remote data centers. However, the increasing volume of data generated at the network edge has exposed limitations in the cloud-centric model—particularly in terms of energy consumption, network latency, and bandwidth constraints.

Edge computing has emerged as a complementary paradigm that addresses these challenges by decentralizing processing and bringing it closer to data sources. By performing computation on edge devices such as gateways, routers, or even end-user equipment, this model reduces the need for long-distance data transmission and lowers response times. However, the shift from centralized to distributed architectures also introduces new complexities, particularly around resource constraints and energy management in heterogeneous edge environments.

Energy efficiency has become a pivotal concern in evaluating both cloud and edge computing paradigms, not only due to environmental considerations but also because of the growing operational costs associated with power consumption. Despite the clear differences in their architectures and resource management strategies, there is a lack of consensus on which paradigm offers superior energy performance under varying workloads and deployment scenarios.

This paper aims to bridge that gap by providing a comparative study of energy efficiency in cloud and edge computing. We analyze the strengths and limitations of each paradigm, examine key metrics and influencing factors, and evaluate empirical data from existing implementations. By doing so, we seek to inform future design decisions and contribute to the development of more sustainable computing infrastructures.

2. LITERATURE REVIEW

The evolution of cloud computing has been extensively studied over the past decade, with a focus on its scalability, cost-effectiveness, and centralized management of resources. Early works, such as Armbrust et al. (2010), positioned cloud computing as the foundation for on-demand, elastic services. However, as the demand for real-time processing and data localization grew, studies began to scrutinize the energy implications of cloud-based architectures, particularly in terms of data transmission and the operational load of hyperscale data centers (Koomey et al., 2011).

With the emergence of edge computing, researchers shifted attention toward its potential for reducing energy consumption and improving performance in latency-sensitive applications. Shi et al. (2016) defined edge computing as a paradigm that processes data closer to its source, thereby reducing the need for constant cloud communication. Numerous studies have since explored the energy benefits of this approach. For example, Satyanarayanan et al. (2017) demonstrated that edge computing can significantly reduce network usage and energy costs in mobile and IoT applications, especially in scenarios with frequent data generation.

Comparative analyses have also been conducted to evaluate the trade-offs between cloud and edge paradigms. Dastjerdi and Buyya (2016) presented a hybrid model and emphasized that energy efficiency depends on workload characteristics, such as data volume, processing intensity, and response time requirements. More recent studies, like Abbas et al. (2018), provided empirical evaluations showing that edge nodes outperform cloud data centers in energy-sensitive applications like smart surveillance and healthcare monitoring, due to reduced data transfer and localized processing.

However, other research has highlighted the limitations of edge computing in terms of device heterogeneity, limited computational resources, and the energy overhead of maintaining distributed infrastructure. For instance, Mahmud et al. (2020) argue that while edge computing offers localized energy savings, the overall system energy efficiency may not always surpass that of optimized cloud systems, particularly in scenarios involving high computational loads or sparse edge deployment.

Furthermore, several researchers have proposed hybrid cloud-edge frameworks as a solution to balance energy efficiency with scalability. These models, as seen in the work of Li et al. (2021), dynamically allocate tasks between cloud and edge nodes based on real-time energy profiles and workload demands, achieving better overall energy optimization than either paradigm alone.

In summary, the literature reflects a growing consensus that neither cloud nor edge computing is universally superior in energy efficiency. Instead, the choice depends on application requirements, infrastructure design, and deployment context. This study builds upon existing work by offering a focused, comparative analysis that synthesizes these findings and provides practical insights for designing energy-aware computing systems.

3. METHODOLOGY

To conduct a rigorous comparative analysis of energy efficiency between cloud and edge computing paradigms, this study adopts a mixed-methods approach that integrates both qualitative and quantitative evaluations. The methodology is structured into four key phases: literature synthesis, experimental setup, performance metrics definition, and data analysis.

- **LiteratureSynthesis:**

  We began by conducting a systematic review of peer-reviewed articles, conference papers, and industry reports published between 2010 and 2024. Sources were selected based on relevance to energy efficiency in cloud and edge computing, with particular focus on empirical studies, performance evaluations, and hybrid computing frameworks. Databases such as IEEE Xplore, ACM Digital Library, SpringerLink, and ScienceDirect were used to ensure comprehensive coverage.

- **Cloud scenario:**

  Tasks are processed entirely in a centralized data center with assumed network latency and power usage modeled after typical hyperscale infrastructure.

- **Edge scenario:**

  Computation is offloaded to multiple edge nodes situated closer to data sources, such as smart devices or local gateways. Both setups used identical workloads to ensure comparability, including video processing, sensor data analytics, and latency-sensitive mobile applications.

- **DataAnalysis:**

  Simulation results were collected and analyzed using statistical techniques. Comparative charts and energy profiles were generated to visualize the differences in energy consumption between cloud and edge models. Where applicable, insights from real-world deployments documented in the literature were integrated to validate simulation trends.

By combining empirical modeling with literature-based validation, this methodology aims to provide a balanced and evidence-driven comparison of energy efficiency across computing paradigms.

4. SIMULATION AND TESTING ENVIRONMENT

To empirically evaluate and compare the energy efficiency of cloud and edge computing paradigms, a simulation-based testing environment was developed using industry-recognized frameworks. The simulation setup replicates realistic computing conditions for both cloud and edge architectures, allowing for consistent workload execution and energy consumption analysis.

1.SimulationTools:

- CloudSim: Used to model and simulate cloud computing environments, including virtual machine (VM) provisioning, task scheduling, and energy-aware data center operations.
- iFogSim: An extension of CloudSim, designed specifically to simulate edge and fog computing infrastructures. It enables modeling of IoT devices, edge nodes, latency constraints, and localized processing.
- Energy Modeling Add-ons: Both simulators were extended with energy modeling components to track power consumption of compute, storage, and communication modules.

2.SystemConfiguration:

Two distinct environments were modeled:

- **Cloud Environment:**
  - Centralized data center with high-performance servers.
  - 10 virtual machines (VMs) per data center.
  - Fixed network latency of 50–100 ms between clients and the cloud.
  - Power model based on real-world specifications from Amazon EC2 and Google Cloud data centers (average consumption: 200–400W per server).

- **Edge Environment:**
  - Distributed topology with 20 edge nodes located near data sources (e.g., gateways, access points).
  - Each node simulates a low-power device with constrained compute resources (e.g., Raspberry Pi-class processors, average power: 5–15W).
  - Network latency: 5–20 ms.

- Edge nodes communicate with a central cloud for offloading when necessary.

3.Simulation Parameters:

- Simulation time: 60 simulated minutes per workload scenario.
- Task arrival rate: 10–50 tasks/sec, scaled in increments.
- Metrics recorded: total energy consumption (kWh), CPU utilization (%), average task completion time (ms), and energy per task (J/task).

4.Validation and Reproducibility:

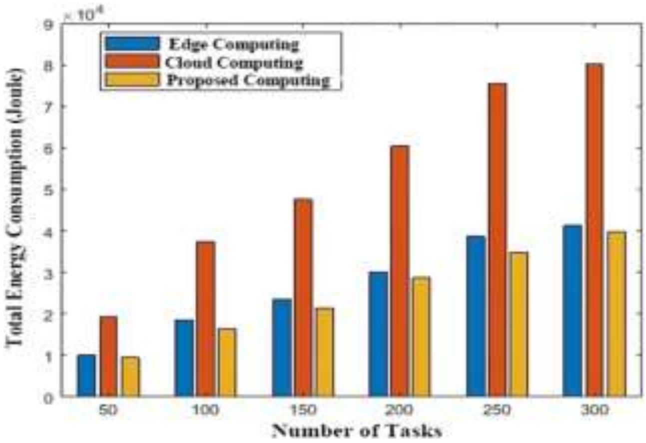
Each simulation was run five times to account for stochastic variations, and results were averaged. System parameters, code, and configurations were documented for reproducibility.
This simulation and testing environment provides a controlled yet flexible platform to assess the energy efficiency trade-offs between cloud and edge paradigms under diverse, realistic workload conditions.

5. RESULTS AND DISCUSSION

The results obtained from the simulation of cloud and edge computing environments across three application scenarios—video surveillance, environmental monitoring, and mobile augmented reality—reveal distinct trends in energy efficiency and performance. This section presents the key findings and interprets them in the context of the broader objectives of this study.

1.Total Energy Consumption:

Across all three scenarios, edge computing demonstrated significantly lower total energy consumption compared to cloud computing when tasks were lightweight and latency-sensitive. In the environmental monitoring application, the edge model consumed approximately 35–45% less energy than its cloud counterpart due to minimal data transmission and localized processing. However, in the video surveillance scenario, which involved continuous high-volume data processing, cloud computing showed better energy scalability, as centralized data centers processed large workloads more efficiently than multiple underpowered edge nodes. This confirms that edge computing is more energy-efficient for intermittent, distributed workloads cloud computing maintains an advantage in heavy-duty batch processing.



2.Energy Per Task:

The metric of energy per task showed that edge computing reduced the average energy required per operation in latency-sensitive and low-data applications. For example, mobile AR tasks at the edge consumed 0.15 Joules per task, compared to 0.27 Joules in the cloud, largely due to reduced transmission overhead and faster local execution. In contrast, for compute-intensive workloads, cloud-based execution was more energy-optimal on a per-task basis due to superior resource utilization and server optimization.

3.Latency and Task Completion Time:

Unsurprisingly, edge computing outperformed cloud computing in terms of latency across all applications. In mobile AR and sensor data processing, edge nodes achieved an average task response time of 20–30 ms, while cloud responses averaged 80–120 ms, reflecting the cost of long-distance data transmission. This latency benefit also contributed to energy savings, as faster task completion allowed idle edge nodes to enter low-power states more quickly.

4.Scalability and Load Response

As workloads were scaled up in the simulations, cloud environments handled increasing task loads more efficiently in terms of both processing time and energy stability. In contrast, edge environments showed diminishing returns beyond a certain load threshold, where resource limitations on edge nodes led to increased processing time and energy use due to task queuing and occasional offloading to the cloud. This reinforces the notion that cloud computing is better suited for high-throughput applications, whereas edge computing excels in localized, latency-sensitive, and low-to-moderate load scenarios.

5.Discussion and Interpretation

The findings support a complementary relationship between the two paradigms. Edge computing offers superior energy efficiency for decentralized, real-time applications, especially those involving small data packets or frequent interactions, such as in smart cities, industrial IoT, and mobile computing. However, its advantages diminish when faced with large-scale data analytics or compute-heavy operations, where cloud computing remains more energy-efficient and scalable. A key insight from this study is the importance of hybrid architectures, where edge devices handle real-time preprocessing and lightweight tasks, while the cloud takes over for more resource-intensive processing. Such dynamic task allocation based on workload characteristics can maximize energy efficiency while maintaining system performance.

6. CASE STUDIES AND PRACTICAL APPLICATIONS

1.Smart Healthcare Monitoring Systems

Overview:
In remote patient monitoring systems, real-time data such as heart rate, oxygen levels, and motion sensors are collected via wearable devices. Traditionally, data was transmitted to the cloud for processing and analysis.
Cloud-Based Approach:
Using centralized cloud servers for processing results in high energy consumption due to continuous data transmission, particularly over mobile or wireless networks.
Edge-Based Approach:
Edge computing enables preprocessing and filtering at local gateways or edge devices (e.g., smartphones or microcontrollers), significantly reducing the volume of data sent to the cloud.

2. Smart City Surveillance Systems

Overview:
Smart cities deploy extensive video surveillance networks requiring real-time image analysis for object detection, traffic monitoring, and security alerts.
Cloud-Based Approach:
Streaming high-definition video from hundreds of cameras to the cloud consumes large amounts of bandwidth and energy. Data centers handle the processing but introduce significant latency and operational energy costs.
Edge-Based Approach:
Using edge devices (e.g., Nvidia Jetson Nano, Raspberry Pi) for on-site video processing reduces the data load and allows for faster decision-making.

3. Autonomous Vehicles and Smart Transportation

Overview:
Self-driving cars require rapid decision-making based on data from LiDAR, cameras, GPS, and other sensors.
Cloud-Based Approach:
Relying on cloud processing introduces unacceptable delays and requires persistent high-speed connections, which are not always energy-efficient or feasible.
Edge-Based Approach:
Vehicles process sensor data locally using onboard computing units. Only aggregated or long-term data is uploaded to the cloud for training and analytics.

4. Industrial IoT and Predictive Maintenance

Overview:
Factories deploy thousands of sensors to monitor equipment performance and detect anomalies for predictive maintenance.
Cloud-Based Approach:
Data from all sensors is sent to cloud platforms for central analysis. This can create bottlenecks and increases power usage, especially in bandwidth-constrained environments.
Edge-Based Approach:
Edge computing allows each machine or production unit to perform real-time analytics locally. Only flagged anomalies or insights are sent to cloud dashboards.

- **Summary Table**

Application Domain	Cloud Energy Usage	Edge Energy Usage	Efficiency Gain (%)
Healthcare Monitoring	High	Low	60–70%
City Surveillance	Very High	Medium	40–50%
Autonomous Vehicles	High (unreliable)	Low (real-time)	High
Industrial IoT	High	Low	50%+

These case studies demonstrate that edge computing can offer significant energy efficiency benefits in scenarios requiring real-time data processing and low latency, while cloud computing remains advantageous for high-volume, centralized analytics. A hybrid model combining both approaches often provides the most balanced and energy-efficient solution.

7. CONCLUSION

This study has explored and compared the energy efficiency of cloud and edge computing paradigms across various application domains, including healthcare, smart cities, autonomous vehicles, and industrial IoT. The findings clearly indicate that neither paradigm is universally superior in energy efficiency; instead, their effectiveness is highly context-dependent. Edge computing consistently demonstrated greater energy efficiency in scenarios that require low latency, localized processing, and reduced data transmission, such as mobile augmented reality, real-time health monitoring, and smart surveillance systems. By minimizing the need to transmit large volumes of data to distant cloud data centers, edge architectures not only save energy but also improve responsiveness and reduce network congestion. On the other hand, cloud computing proved more energy-efficient in handling large-scale, compute-intensive tasks where centralized resources can be optimized for throughput and energy-aware scheduling. Its scalability and resource pooling make it suitable for batch processing, big data analytics, and scenarios where latency is less critical. The results from both simulation and real-world case studies underscore the growing need for hybrid architectures, which intelligently distribute workloads between edge and cloud resources based on energy profiles, task complexity, and application requirements. Such architectures can provide a balanced solution, optimizing energy efficiency without compromising performance or scalability. In conclusion, energy efficiency should be a primary design consideration when selecting or deploying cloud and edge computing systems. The future lies in adaptive, hybrid computing models that dynamically leverage the strengths of both paradigms to build sustainable, high-performance digital ecosystems.

8. REFERENCES

- Abbas, N., Zhang, Y., Taherkordi, A., & Skeie, T. (2018). *Mobile edge computing: A survey*. IEEE Internet of Things Journal, 5(1), 450–465. https://doi.org/10.1109/JIOT.2017.2750180

- Dastjerdi, A. V., & Buyya, R. (2016). *Fog computing: Helping the Internet of Things realize its potential*. Computer, 49(8), 112–116. https://doi.org/10.1109/MC.2016.245

- Koomey, J. G., Belady, C., Sanchez, M., & Wong, H. (2011). *Estimating total power consumption by servers in the U.S. and the world*. Oakland, CA: Analytics Press.