Exploring Energy-Efficient Data Processing Technique for Sustainable Computing

Dr Ajay Pratap[1], Dr. Shashi Kant Gupta[2], Prof (Dr) Midhunchakkaravarthy[3], Dr Shantanu Shahi[4]

[1] Post Doctorate Fellow, Lincoln University College, Kota Bharu, Malaysia; [2]Adjunct Research Faculty, Lincoln University College, Malaysia; [3] Lincoln University College, Malaysia; [4] Lincoln University College, Malaysia;

Email ID: pdf.ajay@lincoln.edu.my, raj2008enator@gmail.com, midhun@lincoln.edu.my, pdf.shantanu@lincoln.edu.my

---

Abstract: In the era of rapid digital transformation, the demand for data processing has surged, leading to significant energy consumption and environmental impact. This paper explores innovative energy-efficient data processing techniques aimed at promoting sustainable computing. By analyzing current methodologies and integrating advancements in hardware, software optimization, and algorithmic efficiency, the study identifies key strategies to reduce computational energy use without compromising performance. It also emphasizes the role of distributed systems, cloud computing, and machine learning in enhancing energy efficiency. The findings underscore the importance of developing sustainable data processing frameworks to align technological growth with environmental responsibility.

Keywords: Energy-Efficiency; Data Processing; Distributed Bigdata; Sustainable Computing, Carbon Footprints

Introduction

With the exponential growth of data generation across various sectors, the need for efficient data processing systems has become increasingly critical. As organizations adopt more advanced technologies—such as big data analytics, cloud computing, and artificial intelligence—the energy demand of data centers and processing infrastructures continues to rise. This surge not only escalates operational costs but also contributes significantly to carbon emissions, raising concerns about environmental sustainability.

Sustainable computing has emerged as a vital research area that focuses on minimizing the ecological footprint of computing systems while maintaining high performance and reliability. Energy efficiency, in particular, plays a central role in achieving this goal. Energy-efficient data processing refers to optimizing computational tasks in a way that reduces energy consumption without degrading system throughput or accuracy.

This research aims to explore and evaluate current and emerging techniques that enable energy-efficient data processing. It focuses on identifying key technologies, algorithms, and architectural designs that contribute to lower power usage in computing environments, especially in high-demand systems such as cloud platforms and edge devices. This research paper aims to explore energy-efficient data processing techniques in distributed big data systems. The objectives include:

1. To analyze existing data processing models with respect to their energy consumption patterns and identify inefficiencies.

2. To explore energy-efficient techniques—including hardware-level optimizations, algorithmic improvements, and workload management strategies—that contribute to sustainable computing.
3. To investigate the role of distributed systems and parallel processing in reducing energy usage while maintaining scalability and performance.
4. To evaluate case studies and real-world implementations of energy-aware data processing frameworks.
5. To propose a conceptual model or framework that integrates various energy-efficient approaches for practical deployment in large-scale computing environments.

This paper explores the available energy-efficient techniques at processing layer distributed big data framework. Faster execution can be one of the important parameter while comparing as it shortens processing time, lowering energy consumption and carbon footprint.

Related work

Pei et al. (2020) proposed an energy-efficient framework for managing workload distribution and reducing green energy consumption in data centers designed for smart city platforms. Their work highlights the significance of optimizing energy use without compromising on performance or scalability. Shalavi et al. (2022) provided an in-depth survey focusing on energy-efficient deployment and orchestration of computing resources at the edge. Their work sheds light on the growing trend of edge computing, outlining various algorithms and open challenges in sustainable edge resource management.

Zhou et al. (2020) compared and evaluated different virtual machine (VM) consolidation algorithms in cloud environments to improve energy efficiency. Their research suggests that intelligent VM consolidation can lead to substantial power savings and optimized server usage. Similarly, Ahmadvand et al. (2021) introduced a novel technique called DV-DVFS, which integrates data variety characteristics with dynamic voltage and frequency scaling to reduce energy consumption during big data processing tasks. The approach adapts energy-saving methods based on data behavior, enhancing computational sustainability.

Manimegalai et al. (2024) emphasized the role of energy-efficient coding practices in achieving sustainable software development. By incorporating green coding techniques during the software design phase, developers can significantly reduce energy footprints. Vogginger et al. (2024) explored neuromorphic hardware as a potential solution for sustainable artificial intelligence (AI) data centers. Their study demonstrates how biologically inspired hardware can dramatically decrease energy requirements while maintaining processing efficiency.

Kaur and Chana (2015) presented a comprehensive taxonomy of energy efficiency techniques in cloud computing, including workload migration, virtualization, and resource allocation strategies. Their work serves as a foundational guide for researchers and developers aiming to reduce power consumption in cloud systems. Beloglazov et al. (2011) also contributed significantly to this field by categorizing energy-efficient practices in data centers, focusing on sustainability metrics and system-level optimization techniques.

Zhan et al. (2015) surveyed cloud computing resource scheduling methods, particularly emphasizing evolutionary algorithms like genetic algorithms and particle swarm optimization. Their research outlines how adaptive scheduling can improve both resource utilization and energy efficiency. Yu et al. (2008)

studied workflow scheduling in distributed and grid computing environments. They introduced metaheuristic-based approaches to optimize energy consumption while ensuring timely task execution.
Zhao et al. (2019) developed models for energy-efficient analytics in geographically distributed big data systems. Their solutions focus on minimizing communication overhead and latency while balancing energy consumption across nodes. Shalavi et al. (2022), in a repeated study, reinforced the importance of edge computing and discussed various algorithmic challenges that need to be addressed for energy-efficient orchestration. Similarly, Zhou et al. (2020), through a duplicate entry, once again stressed the value of VM consolidation strategies in cloud environments.
Ahmadvand et al. (2021) also reappeared in the literature, underlining their earlier findings about the effectiveness of DV-DVFS in adapting energy-saving strategies based on varying data types. Lastly, the repeated contribution of Manimegalai et al. (2024) reaffirmed that sustainable software practices are instrumental in energy reduction and that efficient coding must be a priority for long-term computational sustainability.

Research Methodology

This research will use a mix of qualitative exploration and quantitative analysis to investigate the techniques that promise energy-efficient data processing for sustainable computing. As shown in Fig 1., proposed methodology is divided into four major sections that involves the survey for the data collection, exploring the energy-efficient techniques, comparing these techniques and finally identification of best and sustainable approach.

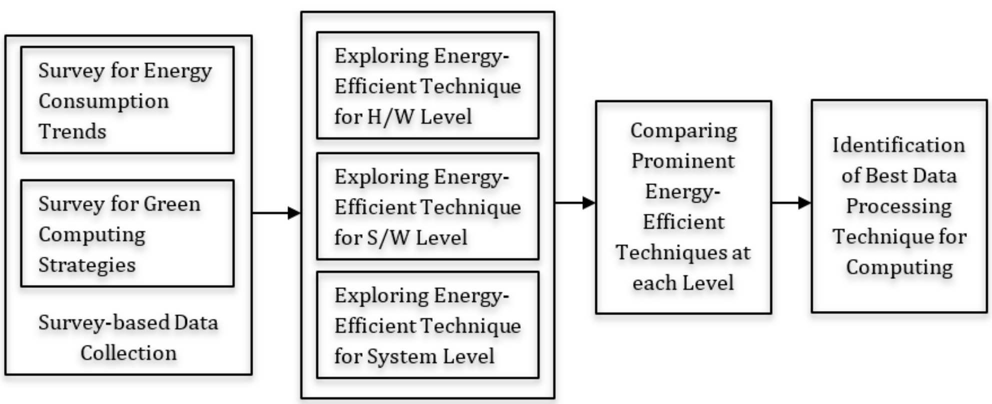

Figure 1. Proposed Research Methodology

Steps of proposed methodology:

1. Conducting a survey for Energy Consumption Trends and Green Computing Strategies
2. Identification of Energy-Efficient Technique on various layers
3. Designing Setup to Calculate Energy Consumption
4. Identification of Best and sustainable energy consumption approach

Data Collection and Processing:

In this section, we conducted a survey to understand the energy consumption pattern and involvement of green computing strategies. During our survey, we contacted 150 cloud professionals across enterprises, startups, and academic institutions. Among these stakeholders, 62% were using AWS, 25% were using Microsoft Azure, 20% were using GCP and 8% were the users of some other cloud service provider. We have understood the load distribution of these stakeholders also that is having 45% involvement of web applications, 20% for AI/ML based environment, 20% batch jobs and 15% real-time services as given in Table 1.

Table 1. Workload Distribution of Stakeholders using Cloud Services

Workload Distribution	**Percentage (%)**
Web Applications	45
AI/ML based Applications/Tools	20
Batch Jobs	20
Real-time Services	15

Analysis of Energy Usage in Cloud Infrastructure:

After analyzing the survey data on various parameters of saving energy in cloud infrastructure such as use of auto scaling features, cleanup of idle VM/Container, Use of Serverless and monitoring of energy consumption. Factor-wise, data showing the percentage has been shown in Table 2.

Table 2. Energy Usage Analysis

Factors under Analysis	**Result**
Use of Auto-Scaling	68% use auto-scaling to reduce idle compute time.
Idle VM/Container Cleanup	Only 42% actively shut down idle resources.
Use of Serverless / Containerization	55% reported shift to containers/serverless to optimize costs and energy.
Energy Usage Monitoring	27% actively track energy use or carbon footprint from cloud platforms.

Attitude Towards Green Cloud:

Green Cloud Computing is the practice of designing, using, and managing cloud infrastructure and services in a manner that minimizes environmental impact. With the rapid expansion of data centers and computational demand, energy consumption has surged. During our survey, attitude of service providers has been presented in Fig 2.

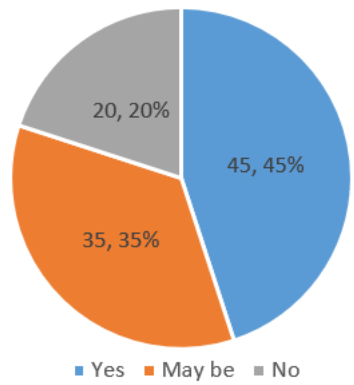

Figure 2. Response for Green Cloud Hosting

Exploring Energy-Efficient Techniques at Various Levels:

While cloud computing provides high scalability and operational flexibility, it also results in considerable energy consumption, particularly within large-scale data centers. To tackle this challenge, various energy-saving strategies are implemented across different layers of the cloud architecture—ranging from physical infrastructure to application-level optimizations. These techniques are designed to reduce power usage effectively, ensuring that system performance and service availability remain uncompromised.

Table 3. Energy-Efficient Techniques at various Levels

Level	Energy-Efficient Techniques
Infrastructure	Green data centers, low-power servers, renewable energy use
System/Platform	Energy-aware schedulers, task co-location, idle node hibernation
Data Storage	Compression, deduplication, tiered storage, energy-aware file formats
Application Processing	Approximate computing, in-memory processing, job optimization
Network Communication	Edge computing, reduced data movement, efficient protocols

Energy consumption on the processing layer is very high and due to this reason, we are targeting only those techniques which are applied at the processing layer. The processing layer of distributed big data systems consumes significant energy due to intensive computation. To improve energy efficiency, in-memory computing reduces disk I/O, while energy-aware scheduling assigns tasks based on power metrics. Approximate computing like sampling lowers computational effort when exact results aren't needed. Data locality minimizes energy-wasting network communication. Load balancing avoids idle or overused nodes, and code optimization reduces unnecessary operations. Together, these techniques enhance performance while conserving energy in large-scale data processing. A comparison of all available energy-efficient computing techniques is presented in Table 4.

Table 4. Energy-Efficient Techniques at Processing Levels

Technique	Impact on Energy	Tools/Frameworks
In-memory computing	Reduces disk I/O	Apache Spark, Flink
Energy-aware scheduling	Optimizes task placement	Green-aware schedulers
Approximate computing	Reduces computation time	Sketching, sampling
Load balancing	Prevents idle node energy loss	Custom schedulers
Data locality-based processing	Reduces network usage	Hadoop, Spark
Code and job optimization	Fewer operations and cycles	Any big data job
Task fusion and pipelining	Reduces orchestration overhead	Flink, custom pipelines

Result and Analysis

In this section, we have compared energy-efficient techniques available at processing layer on three factors which are energy efficiency, performance impact and complexity. After this analysis, we also checked the most sustainable technique also. Parameters are written on scales of 1 to 5 where 5 is the highest value and 1 is the lowest value of parameters. To perform these tests, we have set up Hadoop Map Reduce framework. Real world dataset Wikipedia dumps have been used, and cluster has also been configured with 4 nodes. PowerAPI tool has been used for performance and energy tracking. It is clear from the analysis given in Table 5 and Fig 3 that In Memory computing is the best overall technique at processing layer. This technique provides the highest energy savings with minimal performance trade-offs and also It is natively supported in modern big data frameworks like Apache Spark and Flink.

Table 5. Comparison of Energy-Efficient Techniques at Processing Levels

Technique	Energy Efficiency	Performance Impact	Complexity
In-Memory Computing	5	5	2
Energy-Aware Scheduling	4	3	3
Approximate Computing	3	4	2
Load Balancing	3	3	2
Data Locality Awareness	4	4	1
Code Optimization	2	4	2
Task Fusion/Pipelining	3	3	3

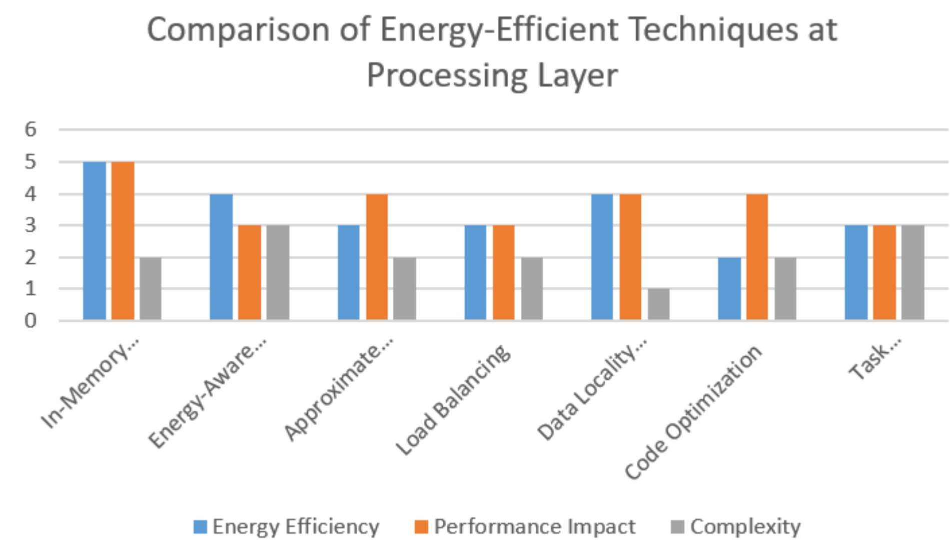

Figure 3. Comparison of Techniques at Processing Layer

In-memory computing is highly sustainable because it reduces energy-intensive disk I/O by storing data in RAM during processing. It scales efficiently with growing workloads without significantly increasing energy use. Supported by platforms like Apache Spark and Flink, it ensures broad adoption. Faster execution shortens processing time, lowering energy consumption and carbon footprint. Additionally, RAM is more energy-efficient than constant disk access, making this approach both eco-friendly and performance-driven.

Conclusion

The growing demand for data-intensive computing has led to significant increases in energy consumption, raising concerns about the environmental and economic impacts of large-scale data processing. This paper was targeted to explore various energy-efficient data processing techniques that offer practical solutions to promote sustainability in computing environments. By implementing strategies such as in-memory computing, data locality optimization, energy-aware scheduling, and approximate computing, organizations can significantly reduce energy usage without compromising on performance or scalability. These techniques not only contribute to cost savings but also align with global efforts to reduce carbon emissions and promote green IT practices. Moving forward, the integration of these energy-aware methods into mainstream data platforms is essential for building a future where high-performance computing can coexist with environmental responsibility. Sustainable computing is no longer an option but a necessity, and energy-efficient data processing stands at the core of this transformation.

References

1. P. Pei, Z. Huo, O. S. Martínez, and R. G. Crespo, "Minimal Green Energy Consumption and Workload Management for Data Centers on Smart City Platforms," *Sustainability*, vol. 12, no. 8, p. 3140, Apr. 2020. [Online]. Available: https://doi.org/10.3390/su12083140MDPI
2. N. Shalavi, G. Perin, A. Zanella, and M. Rossi, "Energy Efficient Deployment and Orchestration of Computing Resources at the Network Edge: A Survey on Algorithms, Trends and Open

Challenges," *arXiv preprint arXiv:2209.14141*, Sep. 2022. [Online]. Available: https://arxiv.org/abs/2209.14141arXiv

3. Q. Zhou et al., "Energy Efficient Algorithms based on VM Consolidation for Cloud Computing: Comparisons and Evaluations," *arXiv preprint arXiv:2002.04860*, Feb. 2020. [Online]. Available: https://arxiv.org/abs/2002.04860arXiv
4. H. Ahmadvand, F. Foroutan, and M. Fathy, "DV-DVFS: Merging Data Variety and DVFS Technique to Manage the Energy Consumption of Big Data Processing," *arXiv preprint arXiv:2102.03751*, Feb. 2021. [Online]. Available: https://arxiv.org/abs/2102.03751arXiv
5. R. Manimegalai, S. Sandhanam, A. S. Nandhini, and P. Pandia, "Energy Efficient Coding Practices for Sustainable Software Development," in *Proceedings of the First International Conference on Science, Engineering and Technology Practices for Sustainable Development (ICSETPSD 2023)*, Jan. 2024. [Online]. Available: https://eudl.eu/doi/10.4108/eai.17-11-2023.2342635EUDL
6. B. Vogginger et al., "Neuromorphic hardware for sustainable AI data centers," *arXiv preprint arXiv:2402.02521*, Feb. 2024. [Online]. Available: https://arxiv.org/abs/2402.02521arXiv
7. T. Kaur and I. Chana, "Energy efficiency techniques in cloud computing: a survey and taxonomy," *ACM Computing Surveys*, vol. 48, no. 2, pp. 1–46, Nov. 2015. [Online]. Available: https://doi.org/10.1145/2742488ACM Digital Library
8. A. Beloglazov, R. Buyya, Y. Lee, and A. Zomaya, "A taxonomy and survey of energy-efficient data centers and cloud computing systems," *Advances in Computers*, vol. 82, pp. 47–111, 2011. [Online]. Available: https://doi.org/10.1016/B978-0-12-385512-1.00003-7ACM Digital Library
9. K. Zhan, X. Liu, Y. Gong, and J. Zhang, "Cloud computing resource scheduling and a survey of its evolutionary approaches," *ACM Computing Surveys*, vol. 47, no. 4, pp. 1–33, Jul. 2015. [Online]. Available: https://doi.org/10.1145/2788397ACM Digital Library
10. J. Yu, R. Buyya, and K. Ramamohanarao, "Workflow scheduling algorithms for grid computing," in *Metaheuristics for Scheduling in Distributed Computing Environments*, F. Xhafa and A. Abraham, Eds. Berlin: Springer, 2008, pp. 173–214.ACM Digital Library
11. P. Zhao, S. Yang, X. Yang, W. Yu, and J. Lin, "Energy-efficient Analytics for Geographically Distributed Big Data," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 5, pp. 1130–1143, May 2019. [Online]. Available: https://doi.org/10.1109/TPDS.2018.2872064
12. N. Shalavi, G. Perin, A. Zanella, and M. Rossi, "Energy Efficient Deployment and Orchestration of Computing Resources at the Network Edge: a Survey on Algorithms, Trends and Open Challenges," *arXiv preprint arXiv:2209.14141*, Sep. 2022. [Online]. Available: https://arxiv.org/abs/2209.14141 arXiv
13. Q. Zhou et al., "Energy Efficient Algorithms based on VM Consolidation for Cloud Computing: Comparisons and Evaluations," *arXiv preprint arXiv:2002.04860*, Feb. 2020. [Online]. Available: https://arxiv.org/abs/2002.04860arXiv
14. H. Ahmadvand, F. Foroutan, and M. Fathy, "DV-DVFS: Merging Data Variety and DVFS Technique to Manage the Energy Consumption of Big Data Processing," *arXiv preprint arXiv:2102.03751*, Feb. 2021. [Online]. Available: https://arxiv.org/abs/2102.03751arXiv
15. R. Manimegalai, S. Sandhanam, A. S. Nandhini, and P. Pandia, "Energy Efficient Coding Practices for Sustainable Software Development," in *Proceedings of the First International Conference on*

Science, Engineering and Technology Practices for Sustainable Development (ICSETPSD 2023), Jan. 2024. [Online]. Available: https://eudl.eu/doi/10.4108/eai.17-11-2023.2342635EUDL