Environmental Footprints of Query Processing: A Vision for Sustainable Database Architectures

Michail Bachras
University of Toronto
Toronto, Canada
michalis.bachras@mail.utoronto.ca

Hans-Arno Jacobsen
University of Toronto
Toronto, Canada
jacobsen@eecg.toronto.edu

ABSTRACT

Database systems underpin modern computing infrastructure, yet their environmental impact remains a significant blind spot in both industry and research. As data volumes grow exponentially, the energy consumption, carbon emissions, and water usage of database operations increasingly threaten global sustainability goals. Our paper explores this multidimensional environmental footprint and proposes a vision where sustainability becomes a first-class design criterion alongside traditional performance metrics. We reimagine database architectures that incorporate environmental awareness throughout both hardware and software layers. By identifying critical research challenges, we establish a foundation for database systems that can deliver high performance while meeting the environmental demands of our resource-constrained world.

PVLDB Reference Format:
Michail Bachras and Hans-Arno Jacobsen. Environmental Footprints of Query Processing: A Vision for Sustainable Database Architectures. PVLDB, 18(11): 4064 - 4072, 2025.
doi:10.14778/3749646.3749676

1 INTRODUCTION

The exponential growth of digital data has made database systems a critical foundation of modern computing infrastructure [20], yet their environmental impact remains a significant blind spot in both industry practice and academic research. Database systems mediate between users and an ever-expanding digital corpus, powering everything from financial transactions to scientific discoveries. As the global "datasphere" expands from 33 zettabytes in 2018 to a projected 394 zettabytes by 2028 [40], the environmental consequences of database operations are becoming increasingly significant. The data centers hosting these systems already consume approximately 1.5% of the world's electricity [83], a figure expected to more than double by 2030, reaching roughly 945 TWh annually, comparable to Japan's entire power consumption in 2024 [83].

This trajectory places database systems on a collision course with global climate objectives. While governing bodies have established legally binding net-zero emissions targets to be reached by 2050 [21, 26, 27], the expanding energy footprint of data centers and database systems threatens to undermine these goals [13, 41, 64, 90]. Unlike other computing domains, the database community has been slow to recognize sustainability as a fundamental design concern. In contrast, the artificial intelligence community has begun accounting for the carbon footprint of model training and inference, calling for "Green AI" practices that prioritize efficiency [19, 48, 61, 68, 94]. Hardware architects routinely consider environmental impact in their designs [15, 31, 85], and cloud providers increasingly publish sustainability metrics [24, 39, 70]. Yet database systems, which often serve as the computational foundation for these other technologies [9, 95], have not received proportional attention in sustainable computing research and practice.

Traditional database performance engineering focuses primarily on query execution time, throughput, and resource utilization, treating energy as merely an operational cost rather than a constrained resource with environmental implications. Benchmark standards in the database community rarely incorporate environmental metrics, and premier database conferences feature few papers explicitly addressing sustainability. Though every database operation, from queries to index updates, has environmental consequences through CPU, memory, and storage hardware usage, these impacts remain largely unaccounted in system evaluations.

In this context, a handful of pioneering works in "green database" research have primarily focused on energy-efficient query processing [28, 49, 88]. While these efforts show potential energy savings, they address just one environmental dimension. Truly sustainable approaches must consider the multidimensional nature of environmental footprints, where complex interactions between factors defy simple optimizations and demand holistic solutions.

We posit that environmental efficiency must be elevated to a first-class design and evaluation criterion for database systems, on par with traditional metrics like performance and scalability. Rather than superficial "green" optimizations layered onto existing designs; this calls for a fundamental reimagining of how database systems are architected, deployed, and evaluated. This paper informs this conversation with the following contributions:

(1) We provide a comprehensive analysis of the multidimensional environmental impact of database systems, revealing complex interdependencies between operational energy consumption, carbon emissions, water footprint, and hardware manufacturing costs.
(2) We propose a vision for environmentally-conscious database architectures that integrate sustainability considerations at every level of system design, from storage management to query processing.
(3) We identify key research challenges and opportunities for database systems to minimize their environmental footprint, highlighting the technical innovations needed to realize sustainable database systems.

Proceedings of the VLDB Endowment, Vol. 18, No. 11 ISSN 2150-8097.
doi:10.14778/3749646.3749676

2 THE MULTIDIMENSIONAL NATURE OF DATABASE ENVIRONMENTAL IMPACT

Database systems' environmental impact extends beyond energy consumption, encompassing complex, interconnected dimensions that current research and engineering practices have yet to fully address. Understanding this multidimensional impact is essential for reimagining database technologies in an increasingly data-driven but resource-constrained world.

When database systems process queries, they generate operational carbon emissions through electricity consumption that can vary by location and timing. A query executed in a coal-dependent region may produce orders of magnitude more carbon dioxide than the identical query in a solar-powered region. This temporal and spatial variability in grid carbon intensity creates both challenges and opportunities for environmental optimization [31].

Perhaps more surprising than operational impact is the significant environmental footprint embedded in database hardware manufacturing and disposal, called embodied carbon or Scope 3 emissions [93]. Research by Gupta et al. [32] reveals that the role of manufacturing carbon emissions in the overall carbon footprint of data centers is becoming more and more significant as the energy grid decarbonizes. Microsoft's data corroborates this finding [54], showing that with 70-75% renewable energy powering operations, nearly half of a data center's carbon emissions stem from hardware manufacturing.

This embodied impact becomes increasingly significant as hardware performance improvements plateau and newer server generations yield diminishing performance gains, leading organizations that replace hardware on traditional 3-4 year cycles to incur substantial environmental costs for relatively modest performance improvements [10, 42]. Bodner et al. [10] demonstrate that for common database workloads, hardware upgrades based solely on raw performance metrics provide minimal benefit once technology improvements slow, yet still incur the full manufacturing carbon penalty.

Database systems often require extensive storage resources, and the embodied carbon differences between storage technologies are substantial [31, 86]. Storage devices represent a particularly important consideration for database systems. SSD-based storage racks emit approximately ten times more embodied carbon per terabyte than older HDD racks, with storage devices accounting for approximately 81% of the total embodied emissions in SSD-based infrastructure [58]. This means that high-performance storage systems in data centers may generate most of their environmental impact through hardware manufacturing rather than operation.

While carbon emissions have received some attention, the water footprint of database operations represents another dimension of environmental impact that receives insufficient attention in database research. Similar to carbon footprints, the water impact of database systems can be divided into two fundamental categories [37]: operational water footprint and manufacturing water footprint.

Operational water footprint derives primarily from electricity consumption and system cooling, particularly in cloud computing and data centers, requiring distinction between water withdrawal and consumption [55, 59]. Water withdrawal refers to the total volume of water removed from a source, much of which may be returned, while water consumption specifically measures the volume of water that is not returned to the original source due to evaporation, transpiration, or incorporation into byproducts. For database environmental assessment, consumption is the more critical metric as it represents permanent removal from the local water cycle. This focus on consumption, particularly of blue water (fresh surface and groundwater resources from rivers, lakes, and aquifers) [77], provides a more accurate picture of long-term environmental impact than withdrawal figures alone. Table 1 presents the consumptive water footprint per unit of power for different energy sources to highlight the differences in water intensity for different variations of the energy grid [44].

Table 1: Water Footprint of Various Energy Sources

Energy Source	(L/MWh)	Energy Source	(L/MWh)
Biomass	1,817	Geothermal	1,363
Hydropower	51,480	Natural Gas	700
Nuclear	2,200	Solar	45
Oil	1,746	Wind	1.85
Coal	1,817		

The geographical variation in water availability introduces another layer of complexity. Regions differ in their water stress levels, making the environmental impact of water consumption location-dependent. An environmentally-conscious database system might deliberately schedule water-intensive computations in regions with abundant non-potable water resources, while minimizing water-dependent operations in drought-prone areas. Innovative infrastructure approaches are already exploring unconventional solutions to these challenges. Microsoft's Project Natick [14] demonstrates an alternative approach by submerging data centers underwater, using the ocean itself for cooling and eliminating freshwater consumption.

Beyond these operational requirements, the manufacturing water footprint constitutes the second major category of database water impact, encompassing the total volume of freshwater used during the production of hardware components. This includes both direct water use in manufacturing processes and indirect consumption through the supply chain. Semiconductor fabrication facilities are particularly water-intensive, requiring ultra-pure water for chip production as a single facility can consume millions of gallons daily [91]. Beyond chip fabrication, water is consumed throughout the hardware lifecycle: in metal mining [65], silicon wafer production [91], component assembly [25], and wastewater treatment [81].

These environmental dimensions interact in complex ways requiring holistic analysis. Optimizing for one dimension often creates unintended consequences elsewhere. For example, pursuing operational carbon efficiency through specialized hardware acceleration [15, 78, 84] may reduce energy use but increase embodied carbon from manufacturing specialized components. Conversely, extending hardware lifespans reduces manufacturing impacts while potentially increasing operational emissions [58].

Water and carbon footprints also exhibit complex tradeoffs. Hydroelectric power offers low carbon emissions but can have substantial water impact through reservoir evaporation. Conversely, wind power has minimal water requirements but may face intermittency challenges that affect long-term database environmental impact. These tradeoffs can be observed in Figure 1, which shows the carbon and water contributions of the energy sources in Ontario, Canada, depicting the different environmental impact of these dimensions.

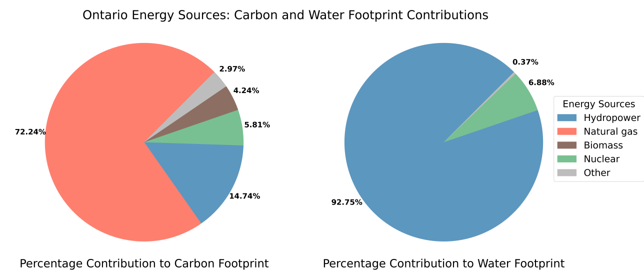


Figure 1: Carbon and Water Footprint of Energy Sources in Ontario, Canada

Understanding these interactions is essential for developing truly sustainable database systems. Environmental efficiency is multifaceted rather than one-dimensional, requiring sophisticated context-aware approaches. This holistic understanding forms the foundation for database architectures that balance complex environmental considerations alongside traditional performance requirements.

3 HARDWARE-AWARE SUSTAINABLE DATABASE ARCHITECTURES

Database systems rely heavily on storage infrastructure, making storage operations' environmental impact a critical consideration in sustainable design. The database-storage relationship presents significant environmental optimization opportunities that research has largely overlooked.

The storage technologies underpinning database systems represent a critical intersection of operational and embodied environmental impacts. Storage decisions affect not only energy consumption during operation but also the long-term manufacturing footprint of the database server and the replacement frequency of hardware components. While previous research has explored the performance implications of storage technologies for database workloads [18, 36], the comprehensive environmental impact spanning operational carbon emissions and water consumption, as well as manufacturing impacts, remains largely unquantified and underexplored. We envision future database systems that fundamentally rethink their relationship with storage media, incorporating environmental awareness throughout the storage stack to maximize component longevity while minimizing both operational and embodied environmental impacts.

Storage Media Environmental Considerations. Modern database servers predominantly use SSDs for their superior speed, but these storage media suffer from wear effects that limit their endurance based on the number of write operations performed over their lifetime and suffers from the limitation of small-granularity overwrites. As highlighted in recent studies [58, 86], SSD-based storage racks emit significantly more embodied carbon per terabyte than HDD alternatives. This environmental cost necessitates maximizing storage component lifespans through architectures that balance endurance with performance considerations.

The environmental tradeoffs between HDDs and SSDs present a nuanced challenge for database architects. While SSDs deliver superior performance and energy efficiency, their limited write endurance and higher manufacturing footprint create sustainability concerns. HDDs, conversely, offer nearly unlimited write endurance and lower embodied environmental impact per terabyte, but at the cost of lower performance, and higher energy consumption due to mechanical operations and increased query execution time.

Despite the critical importance of storage technologies in database systems, comprehensive research comparing the environmental impact of HDDs versus SSDs in database contexts remains sparse, posing a significant gap in sustainable database design understanding. A truly environmentally-conscious database architecture requires quantification of how different storage technologies affect the environmental footprint across multiple dimensions: operational energy use and resulting carbon/water footprints during workloads; manufacturing carbon/water footprint considering replacement rates from SSD wear versus HDD mechanical failures; and secondary performance effects, such as increased memory requirements for slower storage and their impact on overall system environmental footprint, including energy consumption from extended query execution times.

Beyond the HDD vs. SSD dichotomy, modern storage technologies come with unique sustainability implications that warrant examination. NVMe storage offers high throughput and has different performance characteristics than traditional SATA SSDs [33], potentially affecting its environmental footprint over time. Emerging technologies like persistent memory (PMEM) and storage-class memory (SCM) [8, 22, 67] promise to bridge the gap between memory and storage, creating new possibilities for performance optimization in database systems. Architectural innovations like Compute Express Link (CXL) and disaggregated memory [1, 29, 51] represent another frontier in storage evolution which promise to substantially improve resource utilization and mitigate memory stranding by fundamentally transforming how database systems access resources through dynamic allocation.

Despite the growing research on these technologies, their environmental implications in database contexts remain largely unexamined. The database research community lacks systematic studies measuring how these technologies affect the energy consumption across diverse workloads, including how any performance tradeoffs might influence operational environmental impact through carbon emissions and water consumption. Manufacturing impact represents another critical dimension as no research has quantified the embodied carbon and water implications of these technologies compared to traditional architectures. For instance, while researchers used an FPGA-based CXL prototype to enable an in-memory database system to access remote memory [51], they did not quantify the operational and manufacturing environmental benefits or overheads. Developing such environmental assessment frameworks, tailored to evaluate how architectural choices affect overall system sustainability, will be essential for informed decision-making when adopting these technologies.

Workload-Specific Storage Challenges. Different database workloads affect storage media in profoundly different ways, presenting unique environmental challenges. Analytical (OLAP) workloads are primarily read-intensive, requiring the processing of large amounts of data to obtain results and conclusions. The processing of this kind of workload puts substantial pressure on memory and requires numerous I/O operations to bring data into memory and store intermediate results. When memory cannot hold all the data and intermediate results, memory pressure increases, and much of the processed data and intermediate results are written to and retrieved from the storage medium. While SSDs provide superior speed for these operations, each write contributes to the eventual wear-out of the device, accelerating the need for replacement and thus increasing the embodied environmental impact of the system.

OLTP workloads present different but equally significant challenges. These transactional workloads involve frequent inserts, updates, and deletes of small amounts of data corresponding to individual transactions. OLTP traffic is characterized by millions of point updates scattered across the logical address space, preventing the flash-translation-layer (FTL) from grouping writes efficiently. Each update triggers expensive read-modify-write garbage-collection cycles that inflate the write amplification factor, reducing the endurance and lifetime of the storage component.

To address these challenges, we envision database architectures that include storage-conscious design principles explicitly considering the characteristics and environmental implications of underlying storage media. Storage-conscious design offers several sustainability opportunities. First, database architectures should prioritize in-memory processing of queries and data, exploiting available memory to minimize storage interactions. Techniques like column-oriented storage, cache-aligned vectors [11, 73], hot/cold splitting [80], lightweight compression [23], SIMD-friendly query operators [23, 46, 73], and variable-size pages [63] can optimize main memory query processing, especially for analytical workloads, while minimizing I/O operations. While these techniques are often employed for performance reasons, their environmental benefits through reduced storage wear remain largely unexplored and unquantified. Second, database query optimizers and storage managers should be aware of and adapt to the specific characteristics of the underlying storage technologies. These advanced optimizers would distinguish write operations not only by I/O cost but also by their aging effect on storage media, potentially accepting slight performance tradeoffs when they yield significant endurance benefits. This approach would require developing new cost models that quantify the "aging effect" of different storage operations, enabling optimizers to balance performance and storage longevity. While Pelley et al. [69] found that for read-dominant analytical workloads, traditional optimizers may be sufficient in terms of performance, storage-aware query optimization for write-intensive workloads has shown to significantly improve performance [7] and implicitly reduce the write-amplification effect. Moreover, studies demonstrated how understanding and actively managing the underlying storage behaviour can significantly reduce write-amplification [16, 35]. Despite these advances, the environmental implications of these optimizations remain largely unquantified. By incorporating wear-leveling awareness into database design, systems can extend the useful life of storage components and reduce the environmental impact associated with premature hardware replacement.

4 SOFTWARE-LEVEL SUSTAINABLE DATABASE ARCHITECTURES

Energy proportionality. A fundamental goal for energy-efficient and green database systems is to achieve energy proportionality, where a system's energy consumption scales linearly with its computational workload. This principle was formally articulated and popularized by Barroso and Holzle [6], where they found that the energy efficiency of Google servers in the 20-30% utilization range, where most systems operate, drops to less than half the energy efficiency at peak performance levels. Today's database servers remain far from this ideal often consuming over half their maximum power even when idle [52], due to fixed overheads in processors, memory, and other components. This inefficiency stems from two interrelated challenges: hardware limitations and database software design that fails to effectively engage with available power management features.

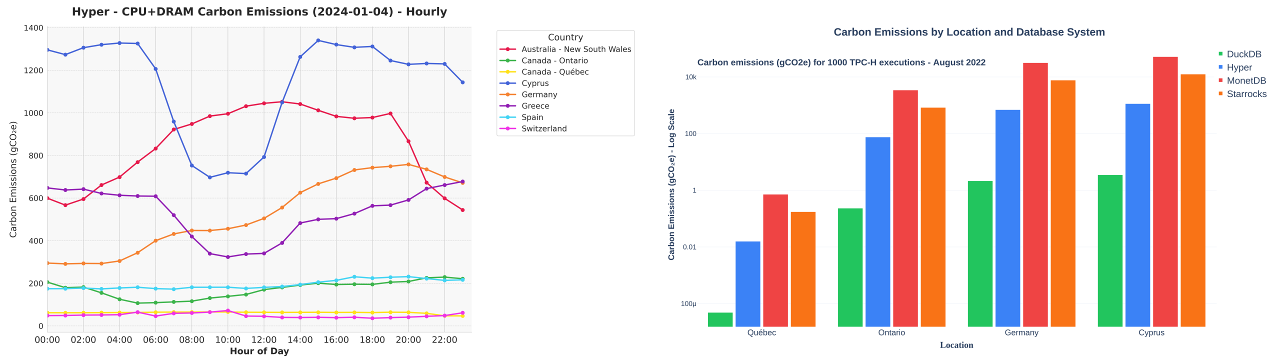
On the hardware side, database workloads present unique energy management challenges compared to general computing. The highly variable resource utilization patterns of database operations [88, 97], which shifts between CPU-intensive query execution, memory-bound joins, and storage-intensive scans, require more sophisticated power management capabilities than currently available. Database servers need hardware components with finer-grained power states, lower idle power consumption, and faster state transitions to accommodate these patterns and minimize performance penalties.

From the database software perspective, current systems largely operate without awareness of their energy consumption, treating hardware resources as always-on entities, without any distinction in the underlying workload. Database management systems must evolve to actively coordinate with underlying hardware power management features. This involves developing frameworks that can dynamically adjust resources and power states, while maintaining query performance guarantees, at least within specified bounds.

Prior research has explored various approaches in energy management: dynamic voltage/frequency scaling [6, 12, 89, 96], intelligent buffer pool management to enable memory power-down [45], and energy-aware query optimization [72, 88, 92] that considers power alongside traditional performance metrics. However, these techniques typically address only isolated aspects of the database stack rather than providing comprehensive solutions.

Achieving energy-proportional database systems requires coordinated innovation between hardware manufacturers and database architects. Hardware vendors need to design software-aware power management capabilities with appropriate granularity and transition speeds, while database developers must reimagine system architectures to incorporate energy awareness throughout the query lifecycle, from query parsing to execution to result delivery. In this context, several open research questions remain on how to integrate power-management primitives into database kernels, how to maintain performance SLAs while dynamically power-capping or power-gating components, and how to measure and optimize "work per Joule" at the query level. Progress on these questions would enable future database systems to provide the same computational work with reduced energy consumption, translating to lower carbon emissions, reduced water footprints, and extended hardware lifespans, aligning database technology with broader sustainability goals.

Environmentally-Aware Workload Scheduling. A critical opportunity for sustainable database systems lies in intelligently timing and placing query execution to minimize environmental impact. The carbon intensity of electricity grids can fluctuate significantly within a single day as renewable generation varies, while water requirements



(a) Hyper carbon emissions during 04/01/2024

(b) Carbon Emissions by Location and DBMS (August 2022)

Figure 2: Temporal and Regional Carbon Emissions Analysis

for electricity generation and cooling differ dramatically across regions and seasons. Figure 2a shows the carbon emissions of the Hyper [46] database for the TPC-H workload with scale factor 300 per 1000 executions, as if it was executing at different times of the day, while Figure 2b shows the carbon emissions of 4 different databases across different locations. These figures demonstrate how timing and location significantly influence carbon footprint, revealing that environmental impact extends beyond simple energy consumption metrics. Carbon emissions, however, represent just one of the environmental dimensions that must be considered. Carbon intensity and water usage interact in complex ways. For example, aggressively reducing carbon by shifting load to a "greener" region might increase water consumed for electricity generation at that location and vice versa.

Database systems are uniquely positioned to navigate these environmental tradeoffs due to their varied workload characteristics. Unlike many computational systems that require immediate execution of all tasks, database workloads naturally contain a spectrum of time-sensitivity requirements. This spectrum ranges from mission-critical transactions that must execute immediately (OLTP workload) to numerous deferrable operations, like batch queries, index rebuilds, statistics collection, maintenance tasks, RAID rebuilds [38, 87], data scrubbing [79], metadata backups, and ETL processes [17, 82], all of which offer significant flexibility in execution timing.

This inherent flexibility makes database systems ideal candidates for applying carbon and water-aware query scheduling approaches. While few database-specific implementations exist [50], general-purpose solutions offer valuable insights and foundational techniques that can be adapted. Recent years have seen initial strides in carbon-aware scheduling in cloud and distributed systems. Notably, Google has deployed a "carbon-intelligent" computing platform [47, 74] that shifts flexible jobs to times and locations with cleaner electricity grid. Using day-ahead grid carbon intensity predictions from Electricity Maps [57], the global scheduler defers non-urgent jobs to align with renewable energy availability without disrupting core services. Beyond industry, the research community has proposed carbon-aware schedulers for cloud clusters that balance performance and cost with emissions; for example, GAIA [34] optimizes batch job placement by considering execution latency, monetary cost, and carbon emissions together. Similarly, Osnes et al. [66] demonstrate wind-aware scheduling that defers latency-insensitive tasks to periods with renewable energy availability. These efforts show that by when and where computations run, we can cut a system's carbon footprint without wholly sacrificing performance.

Water footprint, though receiving less attention than carbon until recently, is equally vital for sustainable computing. Data centers consume enormous amounts of water for cooling and indirectly through electricity generation [62]. Pioneering research is now highlighting how scheduling can reduce water usage [30]. For instance, water-aware geo-distributed scheduling can exploit temporal and regional differences in Water Usage Effectiveness [4] of data centers by moving workloads to locations or times where the water footprint is smaller [4, 43]. Many batch workloads (like machine learning training or periodic analytics) have flexibility in when and where they execute, which creates an opportunity to time their execution during "water-efficient" hours. Importantly, it's not just the volume of water that matters but also the scarcity of local water resources. Running a job in an area experiencing drought or high water stress has a bigger environmental impact than using the same amount of water in a water-abundant area. Simply minimizing water consumption is not sufficient, database systems should account for the regional preciousness of water by, for example, preferring to schedule in areas with low water stress even if it means slightly higher water use overall.

The database community has yet to fully explore and implement these techniques, particularly for distributed and cloud-native database architectures where the flexibility in execution timing and location is greatest. A key challenge in implementing environmentally-aware database scheduling is respecting the complex interdependencies, consistency requirements, and performance SLAs that distinguish database operations from general batch jobs.

Building on innovations in distributed query processing, we can envision a new class of environmentally-conscious database systems that dynamically schedule queries by jointly optimizing carbon and water metrics in real time. A compelling example of this architectural direction is MotherDuck [3], which extends DuckDB for hybrid client-cloud processing. MotherDuck's remote-local optimizer partitions query plans into fragments, executing them either locally or remotely based on data locality and estimated data transfer cost, with bridge operators inserted to move intermediate data between client and cloud. This flexible execution model presents a natural opportunity for incorporating environmental awareness beyond traditional metrics. In such an enhanced system, query optimizers would incorporate carbon intensity and water impact as first-class cost factors alongside computational costs, I/O overhead, and network transfer times. When planning query execution, the optimizer could weigh environmental costs alongside performance considerations, potentially routing computation-intensive fragments to regions with lower carbon emissions or more sustainable cooling capabilities, while still respecting the same execution constraints that systems like MotherDuck [3] already handle. This approach would enable databases to make intelligent tradeoffs between performance, data locality, and environmental impact within a single optimization framework.

By integrating these environmental considerations into database query optimization and scheduling, we can create systems that significantly reduce their ecological footprint while maintaining the performance guarantees and consistency requirements that applications depend on. This approach leverages the inherent flexibility of database workloads and the distributed nature of modern database architectures to make meaningful contributions to sustainability goals without compromising functionality.

5 RESEARCH CHALLENGES

Database systems face challenges that expose current knowledge and methodological limitations, hindering environmentally efficient data management development. Below we outline key challenge areas revealing practical gaps and research opportunities:

Limited Transparency in Hardware Embodied Data. A fundamental obstacle is the difficulty of obtaining accurate data on the embodied carbon footprint of hardware components. Manufacturers often lack transparency in disclosing the carbon impact of producing servers and components, leaving researchers with coarse estimates [31, 75]. In practice, for most hardware components, no vetted emission factor is publicly available, forcing reliance on approximate models. This data gap makes it hard to quantify how much "hidden" carbon is associated with database infrastructure, and thus hinders optimizations.

An even less-explored metric is the water footprint associated with hardware manufacturing. Quantifying the water used in producing computing equipment is extremely challenging due to a lack of industry disclosure and modeling frameworks. Whereas carbon Life Cycle Assessment reports provide at least some basis for estimating embodied CO_2, analogous methods for water are underdeveloped. A recent study on AI infrastructure [53], for instance, had to exclude the manufacturing water footprint entirely due to lack of data.

Challenges in Accounting for Operational Water Footprint. Calculating operational water footprints is complex due to widely varying water intensity across electricity sources, locations, and water categories. Water footprint accounting distinguishes between different categories, which introduces methodological complexity in defining what "water footprint" should be counted. To make matters more difficult, location-specific data on grid water usage are often missing or outdated. While some studies [44, 56, 60, 62] and databases [55] publish average water-use factors for electricity generation, these factors show large variance across regions and technologies and may not capture real-time or regional specifics.

Inconsistent Methodologies for Electricity Carbon Intensity Across Regions. Measuring the carbon footprint of electricity consumed by a database system is crucial for sustainability, but current methods of calculating grid carbon intensity are inconsistent and non-standardized [2]. Different countries and providers use varying methodologies to estimate the carbon intensity of an energy source, yielding to significantly different carbon intensity values for the same mix of energy grid [76]. As a result, this lack of standardization poses a challenge for database systems research: a "carbon-aware" query scheduler or storage layout might perform well under one carbon accounting scheme but appear suboptimal under another. Addressing this issue will require harmonizing carbon intensity methodologies at least for evaluation purposes.

Lack of Standardized Benchmarks for Environmental Efficiency. Unlike performance and scalability, the environmental efficiency of database systems currently lacks standardized benchmarks and evaluation methodologies. The database community has long relied on benchmarks like TPC-C, TPC-H, JOB, YCSB to compare performance metrics. However, while some energy measurement methods exist for existing benchmarks [71], there is no widely adopted, standardized approach for sustainability metrics such as carbon and water footprint per query. This absence of standardized evaluation methods makes it difficult to quantitatively compare systems or to track progress across more than one sustainability dimension, and it also means that researchers may overlook trade-offs. Recent work [5] begins to address this gap by introducing EcoQuery[1], a sustainability-focused benchmark that extends traditional performance metrics with environmental impact measurements. Developing these benchmarks and agreeing on evaluation methodologies will be crucial for driving research in sustainable database systems as it will enable objective comparisons and help identify best-in-class designs that minimize environmental impact across the board.

6 CONCLUSIONS

Database systems stand at a critical environmental crossroads. As the digital datasphere expands exponentially, the environmental footprint of database operations, spanning energy consumption, carbon emissions, and water usage, demands urgent attention from both researchers and practitioners. In this vision paper, we have demonstrated the multidimensional nature of database environmental impact and proposed architectural approaches that elevate sustainability to a first-class design consideration. Our vision extends beyond energy efficiency optimizations to advocate for fundamentally reimagined database architectures that incorporate environmental awareness at both hardware and software levels. By developing storage-aware designs that maximize component longevity, implementing energy-proportional operations through hardware-software co-design, enabling carbon-aware and water-conscious workload scheduling, and creating standardized environmental benchmarks, we can transform how database systems interact with our increasingly resource-constrained world.

ACKNOWLEDGEMENTS

This work was supported by NSERC CREATE (SDSDS) and ORF.

[1] EcoQuery framework: https://github.com/MSRG/EcoQuery

REFERENCES

[1] Minseon Ahn, Andrew Chang, Donghun Lee, Jongmin Gim, Jungmin Kim, Jaemin Jung, Oliver Rebholz, Vincent Pham, Krishna Malladi, and Yang Seok Ki. 2022. Enabling CXL memory expansion for in-memory database management systems. In *Proceedings of the 18th International Workshop on Data Management on New Hardware*. 1–5.

[2] Robbie M Andrew. 2020. A comparison of estimates of global carbon dioxide emissions from fossil carbon sources. *Earth System Science Data Discussions* (2020), 1–51.

[3] RJ Atwal, Peter A Boncz, Ryan Boyd, Antony Courtney, Till Döhmen, Florian Gerlinghoff, Jeff Huang, Joseph Hwang, Raphael Hyde, Elena Felder, et al. 2024. MotherDuck: DuckDB in the cloud and in the client.. In *CIDR*.

[4] Dan Azevedo, Symantec Christian Belady, and Jack Pouchet. 2011. Water usage effectiveness (WUE): A green grid datacenter sustainability metric. *The Green Grid* 32 (2011).

[5] Michail Bachras and Hans-Arno Jacobsen. 2025. Beyond Performance: Measuring the Environmental Impact of Analytical Databases. (2025).

[6] Luiz André Barroso and Urs Hölzle. 2007. The case for energy-proportional computing. *Computer* 40, 12 (2007), 33–37.

[7] Daniel Bausch, Ilia Petrov, and Alejandro Buchmann. 2012. Making cost-based query optimization asymmetry-aware. In *Proceedings of the Eighth International Workshop on Data Management on New Hardware*. 24–32.

[8] Bishwaranjan Bhattacharjee, Mustafa Canim, Christian A Lang, George A Mihaila, and Kenneth A Ross. 2010. Storage Class Memory Aware Data Management. *IEEE Data Eng. Bull.* 33, 4 (2010), 35–40.

[9] Harrison John Bhatti and Babak Bashari Rad. 2017. Databases in Cloud Computing: A literature review. *International Journal of Information Technology and Computer Science* 9, 4 (2017), 9–17.

[10] Thomas Bodner, Martin Boissier, Tilmann Rabl, Ricardo Salazar-Díaz, Florian Schmeller, Nils Strassenburg, Ilin Tolovski, Marcel Weisgut, and Wang Yue. 2025. A Case for Ecological Efficiency in Database Server Lifecycles. In *Proc. 15th Annual Conference on Innovative Data Systems Research (CIDR 2025)*.

[11] PA Boncz and M Zukowski. 2012. Vectorwise: Beyond column stores. *IEEE Data Engineering Bulletin* 35, 1 (2012), 21–27.

[12] Lixia Chen, Jian Li, Ruhui Ma, Haibing Guan, and Hans-Arno Jacobsen. 2020. Balancing power and performance in HPC clouds. *Comput. J.* 63, 6 (2020), 880–899.

[13] Henry Cook, Ramke Ramakrishnan, Xingyu Gu, Aaron Rosenbaum, and Masud Miraz. 2024. *Magic Quadrant for Cloud Database Management Systems*. Technical Report MQ. Gartner, Inc. Gartner Research.

[14] Ben Cutler, Spencer Fowers, Jeffrey Kramer, and Eric Peterson. 2017. Dunking the data center. *IEEE Spectrum* 54, 3 (2017), 26–31.

[15] Pranav Dangi, Thilini Kaushalya Bandara, Saeideh Sheikhpour, Tulika Mitra, and Lieven Eeckhout. 2024. Sustainable Hardware Specialization. In *Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design*. 1–9.

[16] Niv Dayan, Luc Bouganim, and Philippe Bonnet. 2015. Modelling and managing SSD write-amplification. *arXiv preprint arXiv:1504.00229* (2015).

[17] Lina Dinesh and K Gayathri Devi. 2024. An efficient hybrid optimization of ETL process in data warehouse of cloud architecture. *Journal of Cloud Computing* 13, 1 (2024), 12.

[18] Jaeyoung Do, Yang-Suk Kee, Jignesh M Patel, Chanik Park, Kwanghyun Park, and David J DeWitt. 2013. Query processing on smart ssds: Opportunities and challenges. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. 1221–1230.

[19] Jesse Dodge, Taylor Prewitt, Remi Tachet des Combes, Erika Odmark, Roy Schwartz, Emma Strubell, Alexandra Sasha Luccioni, Noah A Smith, Nicole DeCario, and Will Buchanan. 2022. Measuring the carbon intensity of ai in cloud instances. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*. 1877–1894.

[20] S. Emirler. 2023. World Tour of Data: 2023 - 120 Zettabytes and Counting. https://www.linkedin.com/pulse/world-tour-data-2023120-zettabytes-counting-sena-emirler-apsrf/. Accessed: 2025-05-22.

[21] European Commission. 2018. *A Clean Planet for all: A European strategic long-term vision for a prosperous, modern, competitive and climate neutral economy*. Technical Report. European Commission. https://climate.ec.europa.eu/eu-action/climate-strategies-targets/2050-long-term-strategy_en Accessed: 2025-05-19.

[22] Ru Fang, Hui-I Hsiao, Bin He, Chandrasekaran Mohan, and Yun Wang. 2011. High performance database logging using storage class memory. In *2011 IEEE 27th International Conference on Data Engineering*. IEEE, 1221–1231.

[23] Florian Funke, Alfons Kemper, Tobias Mühlbauer, Thomas Neumann, and Viktor Leis. 2014. Hyper beyond software: Exploiting modern hardware for main-memory database systems. *Datenbank-Spektrum* 14 (2014), 173–181.

[24] Google Cloud. 2025. *Carbon Footprint reporting methodology*. https://cloud.google.com/carbon-footprint/docs/methodology Accessed: 2025-05-17.

[25] Martin Goosey. 2005. Water use in the printed circuit board manufacturing process and approaches for reducing consumption. *Circuit World* 31, 2 (2005), 22–25. https://doi.org/10.1108/03056120510571815

[26] Government of Australia. 2022. Climate Change Act 2022. https://faolex.fao.org/docs/pdf/aus213481.pdf Accessed: 2025-05-24.

[27] Government of Canada. 2021. Canadian Net-Zero Emissions Accountability Act. https://www.canada.ca/en/services/environment/weather/climatechange/climate-plan/net-zero-emissions-2050/canadian-net-zero-emissions-accountability-act.html Accessed: 2025-05-24.

[28] Binglei Guo, Jiong Yu, Bin Liao, Dexian Yang, and Liang Lu. 2017. A green framework for DBMS based on energy-aware query optimization and energy-efficient query processing. *Journal of Network and Computer Applications* 84 (2017), 118–130.

[29] Yunyan Guo and Guoliang Li. 2024. A CXL-powered database system: Opportunities and challenges. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*. IEEE, 5593–5604.

[30] Pranjol Sen Gupta, Md Rajib Hossen, Pengfei Li, Shaolei Ren, and Mohammad A Islam. 2024. A dataset for research on water sustainability. In *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*. 442–446.

[31] Udit Gupta, Mariam Elgamal, Gage Hills, Gu-Yeon Wei, Hsien-Hsin S Lee, David Brooks, and Carole-Jean Wu. 2022. ACT: Designing sustainable computer systems with an architectural carbon modeling tool. In *Proceedings of the 49th Annual International Symposium on Computer Architecture*. 784–799.

[32] Udit Gupta, Young Geun Kim, Sylvia Lee, Jordan Tse, Hsien-Hsin S Lee, Gu-Yeon Wei, David Brooks, and Carole-Jean Wu. 2021. Chasing carbon: The elusive environmental footprint of computing. In *2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*. IEEE, 854–867.

[33] Gabriel Haas and Viktor Leis. 2023. What modern nvme storage can do, and how to exploit it: High-performance I/O for high-performance storage engines. *Proceedings of the VLDB Endowment* 16, 9, 2090–2102.

[34] Walid A Hanafy, Qianlin Liang, Noman Bashir, Abel Souza, David Irwin, and Prashant Shenoy. 2024. Going green for less green: Optimizing the cost of reducing cloud carbon emissions. In *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3*. 479–496.

[35] Sergej Hardock, Ilia Petrov, Robert Gottstein, and Alejandro Buchmann. 2017. In-place appends for real: Dbms overwrites on flash without erase. In *Advances in database technology-EDBT 2017: 20th International Conference on Extending Database Technology, Venice, Italy, March 21-24, 2017: proceedings*. Universität Konstanz, 586–589.

[36] Haochen He, Erci Xu, Shanshan Li, Zhouyang Jia, Si Zheng, Yue Yu, Jun Ma, and Xiangke Liao. 2023. When database meets new storage devices: understanding and exposing performance mismatches via configurations. *Proceedings of the VLDB Endowment* 16, 7, 1712–1725.

[37] Arjen Y Hoekstra. 2011. *The water footprint assessment manual: Setting the global standard*. Routledge.

[38] Duwon Hong, Keonsoo Ha, Minseok Ko, Myoungjun Chun, Yoona Kim, Sungjin Lee, and Jihong Kim. 2021. Reparo: A fast raid recovery scheme for ultra-large ssds. *ACM Transactions on Storage (TOS)* 17, 3 (2021), 1–24.

[39] IBM. 2025. *Cloud Sustainability Reporting*. https://www.ibm.com/docs/en/cloudability-commercial/cloudability-premium/saas?topic=insights-cloud-sustainability-reporting Accessed: 2025-05-17.

[40] IDC and Statista. 2024. Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2023, with forecasts from 2024 to 2028 (in zettabytes). Online graph. https://www-statista-com.myaccess.library.utoronto.ca/statistics/871513/worldwide-data-created/ Accessed: 2025-06-28.

[41] International Energy Agency. 2023. Data Centres and Data Transmission Networks. https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks Accessed: 2025-05-22.

[42] Shixin Ji, Zhuoping Yang, Alex K Jones, and Peipei Zhou. 2024. Advancing Environmental Sustainability in Data Centers by Proposing Carbon Depreciation Models. *arXiv preprint arXiv:2403.04976* (2024).

[43] Yankai Jiang, Rohan Basu Roy, Raghavendra Kanakagiri, and Devesh Tiwari. 2025. WaterWise: Co-optimizing Carbon-and Water-Footprint Toward Environmentally Sustainable Cloud Computing. In *Proceedings of the 30th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming*. 297–311.

[44] Yi Jin, Paul Behrens, Arnold Tukker, and Laura Scherer. 2019. Water use of electricity technologies: A global meta-analysis. *Renewable and Sustainable Energy Reviews* 115 (2019), 109391.

[45] Alexey Karyakin and Kenneth Salem. 2017. An analysis of memory power consumption in database systems. In *Proceedings of the 13th International Workshop on Data Management on New Hardware*. 1–9.

[46] Alfons Kemper and Thomas Neumann. 2011. HyPer: A hybrid OLTP&OLAP main memory database system based on virtual memory snapshots. In *2011 IEEE 27th International Conference on Data Engineering*. IEEE, 195–206.

[47] R. Koningstein. 2021. *We now do more computing where there's cleaner energy*. https://blog.google/outreach-initiatives/sustainability/carbon-aware-computing-location/ Accessed: 2025-05-27.

[48] Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. 2019. Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700* (2019).

[49] Willis Lang, Ramakrishnan Kandhan, and Jignesh M Patel. 2011. Rethinking query processing for energy efficiency: Slowing down to win the race. *IEEE Data Eng. Bull.* 34, 1 (2011), 12–23.

[50] Adam Lechowicz, Rohan Shenoy, Noman Bashir, Mohammad Hajiesmaili, Adam Wierman, and Christina Delimitrou. 2025. Carbon-and Precedence-Aware Scheduling for Data Processing Clusters. *CoRR* (2025).

[51] Donghun Lee, Thomas Willhalm, Minseon Ahn, Suprasad Mutalik Desai, Daniel Booss, Navneet Singh, Daniel Ritter, Jungmin Kim, and Oliver Rebholz. 2023. Elastic use of far memory for in-memory database management systems. In *Proceedings of the 19th International Workshop on Data Management on New Hardware.* 35–43.

[52] Ricardo Lent. 2015. Analysis of an energy proportional data center. *Ad Hoc Networks* 25 (2015), 554–564.

[53] Pengfei Li, Jianyi Yang, Mohammad A Islam, and Shaolei Ren. 2025. Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. *Commun. ACM* (2025).

[54] Jialun Lyu, Jaylen Wang, Kali Frost, Chaojie Zhang, Celine Irvene, Esha Choukse, Rodrigo Fonseca, Ricardo Bianchini, Fiodar Kazhamiaka, and Daniel S Berger. 2023. Myths and misconceptions around reducing carbon embedded in cloud platforms. In *Proceedings of the 2nd Workshop on Sustainable Computer Systems.* 1–7.

[55] Jordan Macknick, Robin Newmark, Garvin Heath, and KC Hallett. 2011. *A Review of Operational Water Consumption and Withdrawal Factors for Electricity Generating Technologies.* Technical Report NREL/TP-6A20-50900. National Renewable Energy Laboratory, Golden, CO. https://doi.org/10.2172/1009674

[56] Jordan Macknick, Robin Newmark, Garvin Heath, and Kathleen C Hallett. 2012. Operational water consumption and withdrawal factors for electricity generating technologies: a review of existing literature. *Environmental Research Letters* 7, 4 (2012), 045802.

[57] Electricity Maps. 2024. *Live 24/7 CO2 emissions of electricity consumption.* http://electricitymap.tmrow.co Accessed: 2024-06-11.

[58] Sara McAllister, Fiodar Kazhamiaka, Daniel S Berger, Rodrigo Fonseca, Kali Frost, Aaron Ogus, Maneesh Sah, Ricardo Bianchini, George Amvrosiadis, Nathan Beckmann, et al. 2024. A call for research on storage emissions. *ACM SIGENERGY Energy Informatics Review* 4, 5 (2024), 67–75.

[59] Mesfin M Mekonnen, PW Gerbens-Leenes, and Arjen Y Hoekstra. 2015. The consumptive water footprint of electricity and heat: a global assessment. *Environmental Science: Water Research & Technology* 1, 3 (2015), 285–297.

[60] James Meldrum, Syndi Nettles-Anderson, Garvin Heath, and Jordan Macknick. 2013. Life cycle water use for electricity generation: a review and harmonization of literature estimates. *Environmental Research Letters* 8, 1 (2013), 015031.

[61] Jacob Morrison, Clara Na, Jared Fernandez, Tim Dettmers, Emma Strubell, and Jesse Dodge. [n.d.]. Holistically Evaluating the Environmental Impact of Creating Language Models. In *The Thirteenth International Conference on Learning Representations.*

[62] David Mytton. 2021. Data centre water consumption. *npj Clean Water* 4, 1 (2021), 11.

[63] Thomas Neumann and Michael J Freitag. 2020. Umbra: A Disk-Based System with In-Memory Performance.. In *CIDR*, Vol. 20. 29.

[64] NextWork. 2025. Cloud Computing Statistics 2025: Adoption, Trends, and Forecasts. https://www.nextwork.org/blog/cloud-computing-stats-2025 Accessed: 2025-05-22.

[65] TE Norgate and RR Lovel. 2006. Sustainable water use in minerals and metal production. *Proceedings of Green Processing* (2006), 133–44.

[66] Idun Osnes, Anis Yazidi, Hans-Arno Jacobsen, Frank Eliassen, and Sabrina Sartori. 2022. Harnessing Task Usage Prediction and Latency Sensitivity for Scheduling Workloads in Wind-Powered Data Centers. *Energies* 15, 12 (2022), 4469.

[67] Ismail Oukid, Robert Kettler, and Thomas Willhalm. 2017. Storage class memory and databases: Opportunities and challenges. *it-Information Technology* 59, 3 (2017), 109–115.

[68] David Patterson, Joseph Gonzalez, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. 2021. Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350* (2021).

[69] Steven Pelley, Kristen LeFevre, and Thomas F Wenisch. 2011. Do query optimizers need to be SSD-aware?. In *ADMS@ VLDB.* 44–51.

[70] Katja Philipp, Jonas Bürkel, and Steffen Grunwald. 2023. *Measure and Track Cloud Efficiency with Sustainability Proxy Metrics, Part I: What Are Proxy Metrics?* https://aws.amazon.com/blogs/aws-cloud-financial-management/measure-and-track-cloud-efficiency-with-sustainability-proxy-metrics-part-i-what-are-proxy-metrics/ Accessed: 2025-05-17.

[71] Meikel Poess, Da Qi Ren, Tilmann Rabl, and Hans-Arno Jacobsen. 2018. Methods for quantifying energy consumption in TPC-H. In *Proceedings of the 2018 ACM/SPEC International Conference on Performance Engineering.* 293–304.

[72] Iraklis Psaroudakis, Thomas Kissinger, Danica Porobic, Thomas Ilsche, Erietta Liarou, Pınar Tözün, Anastasia Ailamaki, and Wolfgang Lehner. 2014. Dynamic fine-grained scheduling for energy-efficient main-memory queries. In *Proceedings of the Tenth International Workshop on Data Management on New Hardware.* 1–7.

[73] Mark Raasveldt and Hannes Mühleisen. 2019. Duckdb: an embeddable analytical database. In *Proceedings of the 2019 International Conference on Management of Data.* 1981–1984.

[74] Ana Radovanović, Ross Koningstein, Ian Schneider, Bokan Chen, Alexandre Duarte, Binz Roy, Diyue Xiao, Maya Haridasan, Patrick Hung, Nick Care, et al. 2022. Carbon-aware computing for datacenters. *IEEE Transactions on Power Systems* 38, 2 (2022), 1270–1280.

[75] Lisa Rivalin, Lingyun Yi, Megan Diefenbach, Alex Bruefach, Frances Amatruda, and Tobias Tiecke. 2024. *Estimating embodied carbon in data center hardware, down to the individual screws.* https://sustainability.atmeta.com/blog/2024/09/10/estimating-embodied-carbon-in-data-center-hardware-down-to-the-individual-screws/ Accessed: 2025-05-15.

[76] Malte Schäfer, Felipe Cerdas, and Christoph Herrmann. 2024. Towards standardized grid emission factors: methodological insights and best practices. *Energy & Environmental Science* (2024).

[77] Caroline Schneider. 2013. Three shades of water: increasing water security with blue, green, and gray water. *CSA News* 58, 10 (2013), 4–9.

[78] Ian Schneider, Hui Xu, Stephan Benecke, David Patterson, Keguo Huang, Parthasarathy Ranganathan, and Cooper Elsworth. 2025. Life-Cycle Emissions of AI Hardware: A Cradle-To-Grave Approach and Generational Trends. *arXiv preprint arXiv:2502.01671* (2025). https://arxiv.org/abs/2502.01671

[79] Thomas JE Schwarz, Qin Xin, Ethan L Miller, Darrell DE Long, Andy Hospodor, and Spencer Ng. 2004. Disk scrubbing in large archival storage systems. In *The IEEE Computer Society's 12th Annual International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems, 2004.(MASCOTS 2004). Proceedings.* IEEE, 409–418.

[80] Reza Sherkat, Colin Florendo, Mihnea Andrei, Rolando Blanco, Adrian Dragusanu, Amit Pathak, Pushkar Khadilkar, Neeraj Kulkarni, Christian Lemke, Sebastian Seifert, et al. 2019. Native store extension for SAP HANA. *Proceedings of the VLDB Endowment* 12, 12, 2047–2058.

[81] Jeonghoo Sim, Jonghun Lee, Hojung Rho, Kwang-Duck Park, Youngkwon Choi, Deokhwan Kim, Hyeonbin Kim, and Yun Chul Woo. 2023. A Review of Semiconductor Wastewater Treatment Processes: Current Status, Challenges, and Future Trends. *Journal of Cleaner Production* 429 (2023), 139570. https://doi.org/10.1016/j.jclepro.2023.139570

[82] Manel Souibgui, Faten Atigui, Saloua Zammali, Samira Cherfi, and Sadok Ben Yahia. 2019. Data quality in ETL process: A preliminary study. *Procedia Computer Science* 159 (2019), 676–687.

[83] Thomas Spencer and Siddharth Singh. 2025. *Energy and AI.* Technical Report. International Energy Agency. https://www.iea.org/reports/energy-and-ai Accessed: 2025-05-15.

[84] Chetan Choppali Sudarshan, Aman Arora, and Vidya A. Chhabria. 2024. GreenFPGA: Evaluating FPGAs as Environmentally Sustainable Computing Solutions. In *Proceedings of the 61st ACM/IEEE Design Automation Conference (DAC).* Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1145/3649329.3657343

[85] Chetan Choppali Sudarshan, Nikhil Matkar, Sarma Vrudhula, Sachin S Sapatnekar, and Vidya A Chhabria. 2024. Eco-chip: Estimation of carbon footprint of chiplet-based architectures for sustainable vlsi. In *2024 IEEE International Symposium on High-Performance Computer Architecture (HPCA).* IEEE, 671–685.

[86] Swamit Tannu and Prashant J Nair. 2023. The dirty secret of ssds: Embodied carbon. *ACM SIGENERGY Energy Informatics Review* 3, 3 (2023), 4–9.

[87] Alexander Thomasian and Jai Menon. 1994. Performance analysis of RAIDS disk arrays with a vacationing server model for rebuild mode operation. In *Proceedings of 1994 IEEE 10th International Conference on Data Engineering.* IEEE, 111–119.

[88] Dimitris Tsirogiannis, Stavros Harizopoulos, and Mehul A Shah. 2010. Analyzing the energy efficiency of a database server. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data.* 231–242.

[89] Yi-Cheng Tu, Xiaorui Wang, Bo Zeng, and Zichen Xu. 2014. A system for energy-efficient data management. *ACM SIGMOD Record* 43, 1 (2014), 21–26.

[90] U.S. Department of Energy. 2024. DOE Releases New Report Evaluating Increase in Electricity Demand from Data Centers. https://www.energy.gov/articles/doe-releases-new-report-evaluating-increase-electricity-demand-data-centers Accessed: 2025-05-22.

[91] Qi Wang, Nan Huang, Zhuo Chen, Xiaowen Chen, Hanying Cai, and Yunpeng Wu. 2023. Environmental data and facts in the semiconductor manufacturing industry: An unexpected high water and energy consumption situation. *Water Cycle* 4 (2023), 47–54.

[92] Lang Willis. 2009. Towards eco-friendly database management systems. In *CIDR 2009, Fourth Biennial Conference on Innovative Data Systems Research.*

[93] World Resources Institute and World Business Council for Sustainable Development. 2004. *The Greenhouse Gas Protocol: A Corporate Accounting and Reporting Standard.* World Resources Institute, Washington, D.C. https://ghgprotocol.org/sites/default/files/standards/ghg-protocol-revised.pdf Accessed: 2025-05-22.

[94] Carole-Jean Wu, Ramya Raghavendra, Udit Gupta, Bilge Acun, Newsha Ardalani, Kiwan Maeng, Gloria Chang, Fiona Aga, Jinshi Huang, Charles Bai, et al. 2022. Sustainable ai: Environmental implications, challenges and opportunities. *Proceedings of Machine Learning and Systems* 4 (2022), 795–813.

[95] Yongjun Xu, Xin Liu, Xin Cao, Changping Huang, Enke Liu, Sen Qian, Xingchen Liu, Yanjun Wu, Fengliang Dong, Cheng-Wei Qiu, et al. 2021. Artificial intelligence: A powerful paradigm for scientific research. *The Innovation* 2, 4 (2021).

[96] Puyuan Yang, Peiquan Jin, and Lihua Yue. 2014. Exploiting the Performance-Energy Tradeoffs for Mobile Database Applications. *J. Univers. Comput. Sci.* 20, 10 (2014), 1488–1498.

[97] Yi Zhou, Mohammed Alghamdi, Shubbhi Taneja, Wei-Shinn Ku, and Xiao Qin. 2016. Towards energy-efficient multicore database systems. In *2016 Seventh International Green and Sustainable Computing Conference (IGSC)*. IEEE, 1–8.