IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Sustainable Data Centers: Advancing Energy Efficiency and Resource Optimization

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/2076071 since 2025-05-25T16:40:39Z

Publisher:

Institute of Electrical and Electronics Engineers Inc.

Published version:

DOI:10.1109/PDP66500.2025.00075

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

09 November 2026

Sustainable Data Centers: Advancing Energy Efficiency and Resource Optimization

Viviana Vaccaro
Polytechnic University of Milan
Milan, Italy
viviana.vaccaro@mail.polimi.it

Lavinia Chiara Tagliabue
Department of Computer Science
University of Turin
Turin, Italy
laviniachiara.tagliabue@unito.it

Marco Aldinucci
Department of Computer Science
University of Turin
Turin, Italy
marco.aldinucci@unito.it

***Abstract* — As global demand for digital services rises, the role of data centers has become increasingly pivotal, making it essential to adopt measures to curb energy consumption and carbon emissions. Data centers rank among the highest energy consumers and are therefore significant contributors to greenhouse gas emissions within the ICT sector. This context has driven the development of innovative solutions aimed at minimizing the environmental impact of data centers while maintaining high standards of performance, security, and reliability. This research presents an in-depth analysis of the current state of the sector. It offers a detailed understanding of the strategies that can be implemented to optimize performance, focusing on energy efficiency, sustainability, and technological innovation. It aims to highlight how the adoption of innovative strategies can significantly reduce energy consumption and CO2 emissions, while optimizing resource utilization. The study also highlights how these strategies can be integrated in a holistic approach to design and management, demonstrating the interdependence between different sustainability practices. Methodologically, this work combines an accurate analysis of existing literature and established industry practices. Energy efficiency strategies are examined, such as server consumption reduction and the adoption of more efficient cooling systems, with a focus on advanced technologies. Special attention is also given to emerging IT technologies, such as hyper-converged infrastructures (HCI) and artificial intelligence (AI), evaluating their potential to improve the efficiency and scalability of next-generation data centers, outlining a pathway toward a sustainable digital future that prioritizes responsible resource usage and environmental management.**

Keywords — data centers, energy efficiency, sustainability, cooling technologies, and emerging IT technologies.

I. Introduction

The digital age has radically transformed the way we live, work, and communicate. Data centers support the exponential growth of digital services, cloud computing, and the Internet of Things (IoT). According to the International Energy Agency (IEA), total worldwide data center consumption in 2022 was estimated around 460 terawatt-hours (TWh), and in 2026 it could reach more than 1,000 TWh [1]. Since many data centers still rely on electricity from fossil fuels, increased energy demand often results in higher carbon emissions. This has led to the development of innovative solutions, such as integrating renewable energy sources, optimizing cooling systems, and improving resource management, aimed at minimizing the environmental impact while maintaining high standards of performance, security, and reliability. However, despite recent advancements, significant opportunities for improvement remain. This paper provides a comprehensive analysis of the current state, exploring these opportunities in detail to offer actionable insights.

II. Research Context

A. Data Center: Components and Energy Impact

A data center is a physical infrastructure that hosts a collection of computational resources that work together to process, manage, store, and transmit large amounts of data. As the backbone of modern digital operations, it is composed of several key components:

- *Servers*, which run software applications and handle data management.
- *Storage Systems.*
- *Networking Equipment,* which facilitates secure, high-speed data transmission both within the data center and externally.
- *Power Supply Systems,* providing uninterrupted power to servers and other devices essential for continuous operation.
- *Cooling Systems,* which regulate temperature and humidity levels to ensure the optimal functioning.
- *Physical Security Infrastructure,* which includes systems to protect the data center from unauthorized access and potential threats.

Each of these components is vital to the seamless operation, security, and reliability of data centers, allowing them to support diverse digital activities across sectors. However, meeting the intense computational demands of modern applications requires massive amounts of energy. Power is used to operate IT equipment like servers, storage devices, and networking equipment; cooling systems, such as air conditioners, computer room air handler (CRAH) units, and chillers; and supporting infrastructure, such as power distribution networks, backup power systems, lighting, and additional equipment.

Among these components, servers and cooling systems stand out as the primary drivers of energy consumption, each accounting for an estimated 43% of the total energy usage, followed by storage units and network devices (11% and 3% respectively) [2]. Servers running continuously to handle high computing loads, generating significant heat in the process which must be managed effectively to prevent hardware failures and service disruptions. Historically, cooling has relied heavily on HVAC (heating, ventilation, and air conditioning) systems, which, despite their effectiveness, consume considerable amounts of energy due to their mechanical refrigeration cycles. Despite recent innovations, thermal management remains one of the most complex and energy-intensive challenges in data center operations.

B. Sustainability in the Data Center Industry

Sustainability in data centers refers to the adoption of practices, technologies, and strategies that minimize their

Postprint of

V. Vaccaro, L. C. Tagliabue, and M. Aldinucci. Sustainable data centers: Advancing energy efficiency and resource optimization. In 33rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2025, Turin, Italy, March 12-14, 2025, pages 486–493. IEEE, 2025.

environmental impact while maintaining or improving operational efficiency. This encompasses a wide range of activities, from reducing energy consumption, by optimizing energy use through advanced cooling technologies and efficient hardware, and integrating renewable energy sources to optimizing water usage and employing sustainable materials in construction and operation. It also involves responsibly disposing of outdated equipment and designing longer-lasting systems to reduce replacement frequency.

It involves a holistic approach that integrates environmental stewardship, economic efficiency, and social accountability into the core operations of data centers. This not only helps in reducing the overall carbon footprint but can also lead to cost savings, improved reliability, and enhanced corporate reputation in an increasingly eco-conscious world. In a market disrupted by price fluctuations and supply constraints, this reliability is economically beneficial as it reduces the risk of downtime, which can be costly for businesses.

C. Regulatory and Political Framework

The rapid growth of digital solutions has made data center optimization a political priority, driving the creation of evolving regulations to enhance energy efficiency, cut carbon emissions, and encourage sustainable resource use. Data center operators must navigate this regulatory environment, adopting internationally recognized standards and certifications to ensure compliance and support long-term sustainability goals.

In the **European Union**, among the most important are:

1) Energy Efficiency Directive (2012/27/EU): requires EU member states to set national energy efficiency targets and develop strategies to improve the energy efficiency of infrastructure, including data centers.

2) Energy Labeling Regulation (EU 2017/1369): provides a framework for energy labeling and information on energy-related products, such as servers and storage devices.

3) European Green Deal: is a comprehensive strategy to make Europe climate-neutral by 2050.

4) Delegated Regulation (EU) 2019/424: for the repairability, recyclability, and lifecycle management of products used in data centers.

International standards and certifications include:

1) ISO 50001 - Energy Management Systems: for establishing, implementing, maintaining, and improving energy management systems, reducing operational costs and environmental impact.

2) ISO 14001 - Environmental Management Systems: helps improve environmental performance through more efficient resource use and waste reduction.

3) LEED (Leadership in Energy and Environmental Design), developed by the U.S. Green Building Council (USGBC), and *BREEAM (The Building Research Establishment Environmental Assessment Method),* developed in the U.K. by the Building Research Establishment (BRE), are widely recognized sustainability assessment systems used to evaluate, rate and certify the environmental performance of buildings, including data centers. They require facilities to meet criteria related to energy savings, sustainable materials use, water management, and indoor environmental quality.

In addition to these international standards, national and local regulations may directly impact the design and operation of data centers. For instance, in regions where water is a scarce resource, regulations may impose restrictions on Water Usage for data center cooling, driving the adoption of alternative solutions like dry cooling or water recycling.

D. Sector Initiatives

Recognizing the need for sustainable practices, various global initiatives have emerged to guide the sector towards more responsible and energy-efficient operations. Among the most notable are:

1) Climate Neutral Data Center Pact: is a self-regulatory initiative launched in Europe in 2021 by 25 leading companies and 17 associations with the goal of making data centers climate neutral by 2030. The pact promotes the improvement of energy efficiency and sets specific targets such as achieving a Power Usage Effectiveness (PUE) of 1.3 or less in new data centers and 1.4 in existing ones by 2025, using 100% renewable energy, minimizing water use, reusing waste heat and adopting circular economy principles [3].

2) The Green Grid: is a global consortium established in 2007, composed of companies, governmental organizations, and academic institutions dedicated to advancing resource efficiency in data centers and business computing ecosystems. The organization aims to create a collaborative platform where industry leaders can share insights and develop best practices, as well as develop metrics, tools and standards to help measure and manage energy consumption more effectively, such as the Power Usage Effectiveness (PUE) metric, which has become a global standard for measuring data center energy efficiency [4].

III. Optimizing Data Centers: Metrics, Strategies, and Best Practices

It is essential to evaluate how various metrics [5] and strategic practices can contribute to achieving the mentioned goals in a coherent and effective manner.

This chapter integrates the different sections into a cohesive narrative, directly linking metrics with strategic practices. This approach facilitates a clear understanding of how each best practice contributes to improving evaluation criteria such as energy efficiency, technological innovation, environmental impact reduction, and resource management. It highlights how these practices interconnect and influence the overall optimization of data centers, offering an integrated view of the strategies that can be adopted to enhance both operational efficiency and sustainability.

A. Energy Consumption and Efficiency

The energy consumption of data centers has become a critical area and represents the total amount of electrical power required annually to operate the facility, typically measured in kilowatt-hours (kWh). When considering data center energy consumption, it is also crucial to account for energy efficiency. While energy consumption refers to the total amount of power used, energy efficiency measures how effectively that energy is utilized to perform the required computing tasks. In other words, energy efficiency determines how much of the energy consumed contributes to useful work

versus how much is wasted, often as excess heat. A data center with high energy consumption is not necessarily inefficient if it is using that energy to achieve a high level of computational output with minimal waste. Conversely, a data center with low energy consumption might still be inefficient if a significant portion of that energy is lost through ineffective cooling, outdated hardware, or poor design.

Understanding the distribution of energy consumption within a data center is key to identifying areas for improvement. Typically, energy use is divided into three main categories: power supply, cooling, and IT equipment. Electrical loads are used for data processing and transmission, while thermal loads are dedicated to maintaining optimal environmental conditions.

Electrical loads primarily include energy used by servers, storage devices, and network equipment. Optimizing this area involves increasing the efficiency of hardware components and employing intelligent workload management. The use of energy-efficient servers and power supplies can significantly reduce energy losses during power conversion and lead to substantial savings.

Thermal loads, which include the energy used for cooling, are another critical aspect of energy efficiency. Cooling systems, necessary to prevent server overheating and maintain temperatures within ASHRAE-recommended limits, account for up to 40% of total energy consumption in a data center. Specifically, Computer Room Air Conditioning (CRAC) and Computer Room Air Handler (CRAH) units require significant amounts of energy to cool the air and maintain optimal conditions for equipment. Enhancing cooling efficiency necessitates a comprehensive approach, considering both system design and airflow optimization.

Although electrical and thermal loads require distinct technological approaches, both are crucial and interconnected. For instance, improving the efficiency of computing systems not only reduces power consumption but also decreases the cooling load required to maintain optimal operating conditions. Similarly, more efficient cooling systems can enhance overall data center efficiency without compromising processing capacity.

1) Metrics:

a) Power Usage Effectiveness (PUE): is one of the key metrics used to assess the energy efficiency of data centers. It is defined as the ratio of the total energy consumed by the data center to the energy used by the IT equipment (servers, storage, and networking devices) alone:

$$PUE = \frac{Total\ Data\ Center\ Energy\ Consumption}{IT\ Equipment\ Energy} \quad (1)$$

A PUE of 1.0 would indicate perfect efficiency, where all energy consumed by the data center is used solely for computing purposes, with no waste. However, typical PUE values for data centers range from 1.2 to 2.0, depending on factors such as design, location, and operational practices. The goal for sustainable data centers is to balance high performance with low energy waste, achieving optimal PUE values and minimizing the environmental footprint.

b) Data Center Infrastructure Efficiency (DCIE): It is the reciprocal of PUE and is used to convert the PUE metric into a percentage. This metric is used to identify how efficient the data center IT equipment is in relation to supporting infrastructure.

$$DCIE = \left(\frac{1}{PUE}\right) \text{x } 100 \quad (2)$$

A higher DCIE percentage indicates greater efficiency, meaning a larger portion of the total energy consumption is being utilized by the IT equipment, with less energy lost to cooling, power conversion, and other supporting infrastructure.

2) Strategies:

By improving energy efficiency through advanced cooling technologies, optimized airflow management, and energy-efficient hardware, data centers can significantly reduce their overall energy consumption, even as they scale to meet growing demand. The International Energy Agency (IEA) reports that, despite the exponential growth in demand for data center services over the past decade, the energy use of the data industry has increased only moderately. This remarkable achievement underscores the ongoing advancements in energy efficiency within the sector, enabling data centers to accommodate ever-increasing workloads while minimizing their environmental impact.

a) Efficient Servers and Processing Loads

∞ **Efficient Servers**

Servers, being the backbone of data centers, are responsible for processing, storing, and managing vast amounts of data. However, they are also among the largest consumers of energy within these facilities, making the efficiency of these systems crucial to reducing energy consumption and improve overall efficiency. Modern servers incorporate advanced technologies that reduce power consumption without compromising performance by managing energy dynamically, such as *Dynamic Voltage and Frequency Scaling (DVFS),* which dynamically adjusts the operating voltage and frequency of a processor based on the current workload.

This means that during periods of lower demand, the server consumes less energy, which cumulatively leads to significant energy savings. In addition to power management, efficient servers also typically incorporate components that are optimized for low power consumption, such as *high-efficiency processors*, which enable considerable computing power with low power consumption, and *high- efficiency power supplies*, which convert power from the mains to the necessary voltages with minimal waste, often achieving efficiency values of over 90 per cent. The use of higher efficiency power supplies also provides secondary savings due to lower uninterruptible power supply (UPS) and cooling system loads.

Consolidation

Consolidation is a foundational strategy in modern data center management aimed at optimizing resources, such as server, storage and networking. Through advanced technologies such as virtualisation, consolidation not only enhances efficiency but also reduces the physical footprint of

data centers and simplifies operations, laying the foundation for more advanced integration and efficiency strategies.

Virtualisation is a process that involves abstracting the hardware of a single physical server and allowing it to be divided into several isolated environments, known as *Virtual Machines (VMs)*, each of which operates as an independent server, capable of running its own operating system and applications. This process effectively maximises the utilisation of computing resources, such as CPU, memory and storage, that might otherwise be underutilised if each application or service ran on separate physical servers.

By pooling resources and dynamically allocating them based on demand, data centers can scale operations more effectively. Instead of deploying additional physical servers, new VMs can be quickly created to handle surges in workload, offering the flexibility required in modern environments with fluctuating demands. Additionally, virtualization simplifies backup and disaster recovery processes. VMs can be easily migrated between physical servers, ensuring continuity during hardware maintenance or unexpected failures. This mobility contributes to maintaining high uptime and reliability. Beyond technical benefits, virtualization has a tangible impact on the physical and economic footprint of data centers. Consolidating workloads onto fewer servers reduces the need for expansive physical spaces, which is particularly advantageous in urban areas where real estate is expensive, and it also leads to lower costs associated with building and maintaining data center facilities.

b) Cooling Systems and Environmental Control

Environmental Condition

Managing temperature and humidity within data centers is essential for maintaining the optimal performance of IT equipment while improving the overall sustainability of operations. By optimizing these environmental conditions, data centers can achieve significant energy savings, thereby reducing their environmental impact.

The American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) has established widely recognized guidelines outlining recommended and allowable temperature and humidity ranges for various classes of IT equipment [6]. These guidelines evolve alongside advancements in technology, progressively relaxing thermal limits to lower cooling costs. This gradual adjustment has enabled data centers to adopt strategies such as raising temperature set points closer to ASHRAE's upper limits, reducing the energy demands of cooling systems. While adjusting operating conditions can yield substantial energy savings, it is crucial to carefully balance these benefits with the potential risks to equipment reliability. Deploying environmental sensors and monitoring software is a critical component of this approach. These tools provide real-time data and alerts, enabling operators to quickly detect and respond to deviations from optimal conditions.

Cooling Systems

Traditionally, data centers have relied on HVAC (heating, ventilation and air conditioning) systems. These systems, while effective in maintaining desired temperatures, are notorious for their energy consumption due to their reliance on mechanical cooling, which uses compressors and refrigerants to achieve cooling, often consuming as much energy as the servers themselves. By optimizing these systems, operators can achieve substantial energy savings, lower Power Usage Effectiveness (PUE) scores, and reduce operational costs. Each type of cooling system has its own set of benefits and challenges, and the choice of the most suitable system depends on factors such as data center size, power density and budget.

In-Room Cooling:

Data centers typically employ two types of air conditioning units: CRAC (Computer Room Air Conditioning) and CRAH (Computer Room Air Handler) systems. CRAC Units operate like conventional air conditioners, utilizing a refrigeration cycle with mediums such as refrigerants, glycol, or water, depending on the specific design and cooling requirements.

The core of a CRAC unit is its refrigeration cycle, which includes a compressor, evaporator, condenser, and expansion valve. Warm air from the data center is drawn into the unit, where it passes over the evaporator coil filled with refrigerant. The refrigerant absorbs the heat, cooling the air before it is recirculated into the server room. The refrigerant, now heated, is compressed and sent to the condenser, where the heat is released to the outside environment, and the refrigerant is cooled and cycled back.

CRAH Units differ from CRAC units in that they primarily use chilled water, often provided by an external chiller, to cool the air. These units rely on large fans to draw warm air from the data center and pass it over chilled water coils, effectively reducing the air temperature before it is recirculated. The heat from the air is absorbed by the water, cooling the air before it is circulated back into the data center. The heated water is then sent to a chiller or cooling tower to be cooled and recirculated.

In-Room Cooling solutions have certain limitations, particularly in modern, high- density data centers. These systems cool the entire room, which can be inefficient because the cold air must travel significant distances to reach the servers, often leading to uneven temperature distribution and the formation of hotspots. To address these challenges, new localized cooling alternatives, known as "close-coupled" devices, have been developed. Which can be used in addition to, or instead of, standard room-based cooling systems. These systems offer several advantages, including limited impact in the event of failures, rather than affecting the entire data center, lower operating costs as they reduce unnecessary airflow and direct cooling directly to the source. However, there are some strong cons, such as higher initial capital costs and redundancy issues. Unlike room-based systems that use an N+1 redundancy model with a single backup unit, close-coupled solutions require an additional cooling unit for each row or rack, increasing both costs and system complexity.

In-Row Cooling:

In-row cooling systems each CRAC unit is dedicated to cooling one row of server racks, minimizing the distance cold air must travel. The cooling units may be mounted among the racks, overhead, or under the floor. This setup enhances cooling precision, reduces energy loss, and eliminates the need for raised floors, making it more versatile for different building designs.

In-Rack Cooling:
In this setup, the CRAC units are dedicated to one server rack and are directly mounted to (or within) the racks. This method is particularly effective for high-density servers (up to 60 kW per rack), as cooling is applied directly to the heat-generating components, minimizing energy loss and improving temperature control.

This approach can be based on solutions aimed at improving airflow movement within the cabinet, such as fan trays, or on precision cooling solutions with refrigerant or chilled water for high-density racks, either as a side cabinet or as a rear rack door assembly.

While in-rack air cooling solutions provide effective thermal management for many scenarios, their efficiency can diminish in high-density configurations where heat loads exceed the capacity of air-based systems to dissipate heat effectively. To address these challenges, **liquid cooling** technologies have emerged as alternative, leveraging the higher thermal conductivity and specific heat capacity of liquids to manage extreme thermal loads efficiently.

Liquid Cooling uses a liquid coolant, typically water or a dielectric fluid, that absorbs heat from the server components and transfers it away for dissipation. It takes advantage of the higher specific heat of liquid to achieve superior cooling performance.

Depending on which components the fluids contact, it can be classified into two main types:

Direct-to-Chip Cooling: In this system dielectric fluid – commonly water, ethylene glycol, or a combination of the two – is delivered directly to the hottest components, usually the GPU and/or CPU via a cold plate placed directly on the chip. The electric components are never in direct contact with the coolant, which is circulated through channels within the cold plates. The fluid absorbs the heat produced by the device, converts it into vapor, and then condenses it back into liquid.

Immersion Cooling: This method involves immersing the servers in containers filled with a dielectric (non-conductive) liquid, which absorbs the heat, converts it to vapor and condenses it back to liquid, enabling very high cooling efficiency and reducing the need for air conditioning.

There are two basic methods of immersion cooling:

Single-phase immersion cooling (1-PIC), in which the dielectric fluid cools hot server parts by direct contact but remains in liquid phase while its temperature increases. Pumps are required to enable proper heat removal and to circulate the dielectric liquid to an external heat exchanger, where it is cooled and cycled back into the tank.

Two-phase immersion cooling (2-PIC), in which the dielectric liquid boils when in direct contact with hot server parts and turns into vapor. Vapor rises and condenses on a heat exchanger (condenser) above the surface, returning to a liquid state in a perpetual and passive cycle. The latent heat of vaporization, the energy required to convert a liquid into vapor without a temperature rise, allows for efficient heat absorption from server components, with lower mass flow rates and consequently lower pumping power use, compared to liquid cooling systems, which means less energy is required to remove heat from the system, leading to significant improvements in overall Power Usage Effectiveness (PUE).

Despite these significant advantages, liquid cooling systems present several challenges that must be addressed to ensure their reliability. The risk of coolant leaks is a critical issue, particularly with the use of dielectric fluids, which are non-conductive but can still pose a hazard if not handled correctly and could damage sensitive electronic components.

Another issue is the choice of refrigerants and their environmental impact, as some refrigerants have a high global warming potential (GWP) and may be subject to regulatory restrictions in regions with strict environmental laws.

Optimization Techniques

Cooling in data centers revolves around two key principles: removing the heat generated by the equipment and distributing cool air evenly throughout the server room. These objectives are achieved through a combination of air conditioning systems, airflow management, and containment solutions.

Airflow Management:
Effective Airflow Management is essential for optimizing the performance and energy efficiency of data centers. Properly directing and controlling the movement of air within the facility helps to prevent hotspots, minimize energy consumption, and enhance overall cooling efficiency. There are two primary approaches to air distribution: overhead and underfloor. Each approach offers advantages and challenges, and the selection between them often depends on the specific requirements and constraints of the data center.

Underfloor air distribution systems are a common feature in data centers and use a raised floor, typically 60 to 90 cm (24 to 36 inches) high, as a plenum for distributing cooled air from the cooling units to the server intakes. To ensure effective airflow management, it is important to use perforated floor tiles strategically in cold aisles, where the cold air is needed most. Additionally, the space under the raised floor should be kept free of obstructions, such as cables or equipment, that could disrupt the flow of air.

Overhead air distribution systems supply cool air from above, typically through ductwork or ceiling-mounted diffusers. This cool air is then directed downward into the cold aisles or directly into the intake of the server racks. The cold air tends to mix better with the rising warm air, promoting a less variable and more even temperature distribution. However, installation can be more complex and expensive, particularly in retrofitted spaces where ceiling heights are fixed, and existing infrastructure may need significant modification.

Hot and Cold Aisle Containment:
It is a fundamental approach to optimize temperature control and improve energy efficiency. The concept involves arranging server racks in alternating rows, with the front of the servers (cold aisles) facing one another and the backs (hot aisles) facing each other. This arrangement helps separate the cold air supplied to the servers from the hot air exhausted by them, preventing mixing and improving cooling efficiency. This means that the hot air can be collected separately and directed to the air handler.

In *Cold Aisle Containment (CAC)*, the cold aisles are enclosed, either by using doors and ceiling panels or by

employing vertical barriers at the ends of the aisle. So, the rest of the data center becomes a hot-air return plenum where the temperature can be high. This method reduces the amount of cold air required. It is easy to implement without the need for additional architecture to contain and return exhaust air such as a drop ceiling or air plenum.

Hot Aisle Containment (HAC), on the other hand, encloses the hot aisles where the heated air is expelled from the back of the servers. The containment system captures this hot air and directs it back to the cooling units through ductwork, preventing it from mixing with the cold air in the data center. This approach can be more effective in facilities where there is a high density of equipment and standalone equipment, such as storage cabinets, can be positioned outside the containment area without risk of overheating. However, can be more expensive.

Finally, effective *Rack-Level Containment* is also crucial for preventing hotspots and ensuring that servers receive an adequate supply of cold air. Solutions, such as blanking panels, can be used to fill empty spaces in server racks, preventing hot air from recirculating to the front of the rack.

Free Cooling Techniques:

Free cooling refers to the use of ambient environmental conditions to assist or fully achieve the cooling needs, significantly reducing or even eliminating the reliance on mechanical refrigeration systems. This approach leverages naturally available cool air or water sources to dissipate heat generated by IT equipment. Free cooling techniques can be integrated with traditional CRAC and CRAH systems to enhance overall cooling efficiency.

External Air Cooling: This technique can be highly effective in climates with significant seasonal temperature variations. The process generally involves drawing outside air through a set of filters and then through cooling coils, where it absorbs heat from the internal air before being expelled outside.

Adiabatic Cooling or evaporative cooling: utilizes the principle of water evaporation to cool the air. This technique involves passing warm air through a water-saturated medium, where the water evaporates, absorbing heat and lowering the air temperature. Adiabatic cooling systems are particularly effective in dry climates where the evaporation process is more efficient.

Natural Water Sources: such as lakes, rivers, or underground aquifers. The water is circulated through a heat exchanger or directly through cooling coils, where it absorbs heat and then returns to its source or is expelled after releasing the absorbed heat. Data centers located near large lakes or rivers can leverage this technique to reduce cooling costs significantly.

Hybrid Systems: that combine free cooling techniques with traditional mechanical refrigeration to create a versatile and efficient cooling solution. By leveraging free cooling when environmental conditions are favorable and relying on mechanical cooling when necessary. One common hybrid approach involves integrating adiabatic cooling with CRAC or CRAH units. During periods of moderate to high outdoor temperatures, the adiabatic system pre-cools the incoming air, reducing the workload on mechanical units. When temperatures rise beyond the effective range of adiabatic cooling, the mechanical systems take over to ensure that the data center remains within optimal temperature ranges.

The effectiveness of free cooling techniques is heavily influenced by **climatic conditions**. Because of this, there has been a noticeable trend in recent years toward locating data centers in countries with climates favorable, e.g. Nordic countries like Sweden, Norway, and Finland. These regions offer ideal climatic conditions for free cooling throughout much of the year, with low average annual temperatures and an abundance of natural water resources for cooling.

Data centers often employ sophisticated control systems and sensors that monitor outdoor temperatures and humidity levels, adjusting the cooling strategy dynamically to optimize performance and energy efficiency.

Heat Recovery:

In recent years, a growing trend has emerged to recover some of the waste heat, which would otherwise be dispersed into the environment, converting it into useful energy for other thermal processes, inside or outside the building.

The process involves the implementation of heat exchangers to transfer heat from the hot exhaust air or cooling liquids used to cool the servers to a secondary medium, such as water or air.

In scenarios where the temperature of the waste heat is not high enough for direct use, heat pumps can be used to increase the temperature.

The recovered heat can be distributed for various uses:

Heating of nearby residential or commercial buildings: This application is particularly advantageous in urban areas where data centers can supply heat to buildings through district heating networks.

Swimming Pool Heating: This application is ideal because pools require a continuous supply of heat throughout the year.

Algae Cultivation: Algae require stable temperatures to grow optimally, and heat recovered from data centers can provide the necessary warmth to support this process. Cultivated algae can be used in various sectors, including biogas production and pharmaceutical applications.

Fish Farming: In Nordic countries, recovered heat is also used to warm aquaculture facilities. Heated water is ideal for creating controlled environments, enhancing the productivity and sustainability of aquaculture operations.

Industrial Processes: especially those requiring constant heat, such as food production or chemical production.

Greenhouse Heating: to maintain appropriate temperatures in greenhouses throughout the year, especially in cold climates.

The **feasibility and efficiency** of heat recovery techniques depend on several key factors:

Temperature of Recovered Heat: is crucial in determining its potential uses. Higher temperatures are generally more versatile and can be applied to a wider range of applications. In contrast, lower temperatures may limit the possible uses and may require more advanced systems to be efficiently utilized [7].

- ∞ *Air-cooled Air Handling Units*: have limited capacity for heat recovery, the temperature of which is generally lower than 30°C. This generally restricts the applications. It may be suitable for low-temperature applications such as water preheating or air pre-conditioning assistance.
- ∞ *Water-Cooled Air Handling Unit*s: This cooling method is much more efficient and can recover waste heat up to around 30°C (86°F), suitable for applications such as greenhouse heating, space heaters, aquaculture, biomass drying, and district heating when combined with heat pumps.
- ∞ *Rear-Door Cooling Systems*: can capture waste heat at temperatures up to 40°C (104°F), which can also be employed in desalination processes to produce potable water.
- ∞ *Cold Plate Cooling Systems*: By capturing heat directly from IT chips, these systems can achieve temperatures up to 45°C (113°F). This higher temperature allows for additional applications, including wastewater treatment for irrigation purposes.
- ∞ *Immersion Cooling*: can achieve temperatures of up to 55°C (131°F). The heat recovered can be used in all previously mentioned applications, including direct district heating without the need for heat pumps.

Seasonal Demand: The demand for heat can fluctuate seasonally, especially for residential heating. This variability can make it challenging to reuse waste heat continuously.

Distance and Heat Loss: The distance between the data center and the heat utilization point impacts the efficiency of the heat recovery system. The greater the distance, the higher the likelihood of heat loss during transfer. Using well-designed insulation and piping systems can help reduce these losses.

Available Infrastructure: The presence of district heating networks or other heat distribution infrastructure is crucial. Without such networks, heat recovery may not be economically viable.

B. Innovative IT Technologies

As the demand for advanced digital services grows, data centers are increasingly integrating innovative IT technologies to optimize performance, reduce costs, and achieve sustainability goals. With IT equipment accounting for over half of a data center's energy consumption, adopting cutting-edge solutions is essential for addressing energy efficiency and operational challenges.

a) Hyper-Converged Infrastructure (HCI): merges computing, storage, and networking into a unified hardware and software solution, simplifying management and reducing complexity. By centralizing resource allocation, HCI optimizes operational efficiency and scalability, enabling seamless expansion without service interruptions. Its streamlined architecture minimizes physical hardware requirements, leading to significant energy and cooling savings while reducing the overall carbon footprint.

b) Cloud Integration: bridges on-premises data centers with cloud platforms, creating a flexible and cost-effective IT ecosystem. By offloading workloads to the cloud, data centers can dynamically scale resources, reduce capital expenditures, and improve performance.

c) Artificial Intelligence (AI): is transforming data center management by automating processes, optimizing energy usage, and enhancing cooling efficiency. AI-driven tools also enable predictive maintenance, real-time energy provider selection, and enhanced security measures, significantly improving operational sustainability and reliability.

C. Sustainability

The sustainability of data centers is increasingly being evaluated through the lens of their carbon footprint, which is directly tied to their energy consumption. Most of the greenhouse gas (GHG) emissions, particularly CO2, from data centers stem from the energy required to power their operations. Often, this energy is generated from fossil fuels, making the reduction of carbon emissions a critical focus for improving sustainability.

According to the International Energy Agency (IEA), global electricity demand is expected to grow by 3.4% annually between 2022 and 2026. In 2022, global data center electricity consumption was estimated at 460 terawatt-hours (TWh), and it could reach more than 1 000 TWh in 2026.

However, despite the growing demand for digital services, the carbon footprint of data centers has only grown modestly since 2010. In 2020, data centers and data transmission networks were responsible for approximately 1330 Mt CO2 equivalent, accounting for about 0.9% of energy-related GHG emissions.

This progress is largely due to improvements in energy efficiency, the adoption of renewable energy, and the decarbonization of electricity grids in many regions. This underscores the need for data centers not only to optimize energy use, but also to switch to cleaner and renewable energy sources to reduce the overall carbon footprint of data centers.

To quantify and manage these emissions, the industry relies on a key metric known as **Carbon Usage Effectiveness (CUE)**, developed by The Green Grid. CUE is calculated as the ratio of the total CO2 emissions caused by a data center's energy consumption to the energy consumed by IT equipment.

$$CUE = \frac{Total\ Carbon\ Emissions}{IT\ Equipment\ Energy} \quad (3)$$

This metric provides a clear measure of how effectively a data center is minimizing its carbon emissions in relation to its energy use.

To achieve the ambitious goals of the Net Zero Emissions (NZE) Scenario, data centers will need to halve their emissions by 2030. This calls for the widespread implementation of mitigation strategies that go beyond energy efficiency, including the increased use of renewable energy, water conservation, and effective waste management.

These strategies not only contribute to reducing the carbon footprint but also play a vital role in enhancing the overall sustainability of data centers, aligning with global efforts to combat climate change.

1) Increased use of Renewable Energy:

Globally, the energy mix remains heavily reliant on fossil fuels, driven by established infrastructure, economic factors, and the vast scale of global energy demand. According to the International Energy Agency (IEA), as of 2022, fossil fuels made up approximately 81% of the global energy supply, with oil alone accounting for about 31%, followed by coal at 27%, and natural gas at 23%, posing a significant barrier to the widespread adoption of renewables, especially in regions like Asia and Africa where economic constraints make these investments more challenging.

2) Use of Materials and Waste Management:

Data centers rely on a diverse array of materials for both construction and operation, including metals, plastics, and various electronic components. The lifecycle of these materials, from extraction and production to eventual disposal, can have significant environmental impacts. These include habitat destruction, pollution, and the depletion of natural resources. To mitigate these effects, sustainable practices in data centers are increasingly emphasizing the use of eco-friendly materials, responsible sourcing, and effective recycling programs. Such practices aim to minimize the material footprint of data centers by ensuring that outdated, broken, or end-of-life assets are either responsibly disposed of or, ideally, recycled.

According to UNU reports (United Nations University), 53.6 million metric tons (Mt) of e-waste were generated globally in 2019, which averages 7.3 kg per capita [8]. This represents a 21% increase in just five years, with projections suggesting that global e-waste will reach 74.7 Mt by 2030. Alarmingly, current estimates indicate that less than 20% of the world's discarded electrical items are recycled properly.

Recycling materials not only reduces waste but also helps mitigate the environmental impact associated with the extraction of virgin raw materials. For example, a typical 1 MW data center generates 33,000 lbs of CO2 emissions from plastics, 73,000 lbs from aluminum, and 377,000 lbs from steel. The production processes for these materials are highly energy-intensive: plastics are made from petroleum or natural gas, contributing to 1% of global greenhouse gas emissions. Aluminum production consumes a lot of electricity, but recycling aluminum saves 90-95% of the energy needed and can be recycled indefinitely. Steel can also be recycled multiple times without losing quality, using 74% less energy compared to producing it from iron ore, and significantly reducing the need for virgin materials, water use, water pollution, air pollution, and mining waste. In addition, the use of recycled steel cuts CO2 emissions by 58%.

3) Water Consumption

Water plays a crucial role in data center operations, particularly in cooling systems, such as chillers and evaporative systems, that require significant amounts of water, which can put additional pressure on areas prone to drought or with limited water supplies.

In addition, the extraction and use of water can result in the alteration of local ecosystems.

One way that data centers can become more sustainable through water is by measuring the Water Usage Effectiveness (WUE) of the facility. WUE is a metric that was developed by The Green Grid to measure sustainability in terms of water usage and its relation to energy consumption. It is the ratio between the use of water in data center systems (water loops, adiabatic towers, humidification, etc.) and the energy consumption of the IT equipment.

$$WUE = \frac{Data\ Center\ Water\ Consumption\ (l)}{IT\ Equipment\ Energy} \quad (4)$$

The lower a data center's WUE ratio is, the more efficient it's use of water resources is. The average data center has a WUE of 1.8L per 1kWh.

Strategies for improving WUE may include using water-efficient cooling methods, recycling and reusing water, implementing advanced cooling technologies and optimizing facility design for reduced water usage.

However, reductions in water use generally have negative trade-offs with increased energy. Therefore, a series of factors must be taken in consideration, such as location, the method of power generation, business strategy and financial considerations.

IV. Conclusion

The growing demand for digital services has placed data centers at the forefront of technological advancement, making them indispensable yet energy-intensive components of the modern infrastructure. This paper has explored innovative strategies and emerging technologies aimed at enhancing the energy efficiency and sustainability of these facilities. By adopting these practices, data centers can significantly reduce their environmental impact while maintaining operational excellence. Achieving these objectives requires a collaborative effort involving technology providers, policymakers, and operators to adopt and refine sustainable practices. As the industry evolves, data centers must embrace innovation and maintain a holistic approach to resource management, ensuring they are equipped to meet the challenges of a rapidly changing digital landscape. By fostering a commitment to sustainability, the data center sector can not only mitigate its environmental footprint but also play a pivotal role in supporting a greener, more responsible future for the digital economy.

Acnowledgment

This work has been supported by the DYMAN project funded by the European Union - European Innovation Council under G.A. n. 101161930.

References

[1] International Energy Agency (IEA), *Electricity 2024: Analysis and Forecast to 2026*. Paris: International Energy Agency, 2023.

[2] M. Dayarathna, Y. Wen and R. Fan, "Data Center Energy Consumption Modeling: A Survey," in *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 732-794, Firstquarter 2016.

[3] Climate Neutral Data Centre Pact, *Commitment to Sustainable Digital Infrastructure*.

[4] The Green Grid, *Advancing Resource Efficiency in Business Computing*.

[5] Sustainable Digital Infrastructure Alliance (SDIA), *Data Center Metrics Framework*.

[6] ASHRAE, *Thermal Guidelines for Data Processing Environments*, 5th ed. Atlanta, GA: ASHRAE, 2021.

[7] K. Ebrahimi, G. F. Jones, and A. S. Fleischer, "A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities" *Department of Mechanical Engineering, Villanova University*, Villanova, PA, USA.

[8] V. Forti, C. P. Baldé, R. Kuehr, and G. Bel, *The Global E-waste Monitor 2020: Quantities, Flows, and the Circular Economy Potential*.