MULTI-CHAPTER GUIDE

Data Center Infrastructure Management Software

Green Data Center: Tutorial & Best Practices

Chapter 3

www.graphicalnetworks.com

CHAPTER 3

Green Data Center: Tutorial & Best Practices

The term “green data center” is often associated with renewable energy, low power usage effectiveness (PUE), or closed-loop cooling systems. However, while those systems and factors are important, they do not necessarily describe the characteristics of an environmentally sustainable or operationally efficient data center. This is particularly true as computing demands continue to evolve rapidly with the exponential growth of AI computing.

Over the last several years, data center infrastructure has changed significantly. AI workloads, hyperscale deployments, and high-density compute environments are driving power and cooling demands far beyond what many facilities were originally designed to support. In many cases, the limiting factor for new development is no longer available floor space or cooling capacity but access to electrical infrastructure and utility interconnection capacity.

As a result, evaluating a data center using a single efficiency metric such as PUE no longer provides a complete picture of overall facility performance. A more complete approach requires looking at how energy is generated, how it is being used throughout the facility, how efficiently supporting infrastructure is operating, and whether portions of rejected energy can be recovered or reused elsewhere within the campus.

This article explains why modern AI-driven data centers require a broader, systems-level approach to sustainability that goes beyond traditional PUE metrics by integrating efficient cooling, renewable and on-site power generation, waste heat recovery, and DCIM-driven operational visibility to optimize overall facility performance and resiliency.

Summary of key Green Data Center concepts

Concept	Description
Looking beyond PUE	PUE remains a useful benchmarking metric, but increasing AI rack densities can artificially improve PUE values even when portions of supporting infrastructure become less efficient. Evaluating a modern green data center requires understanding how energy is generated, distributed, consumed, and potentially recovered throughout the facility.
Establishing a baseline	Developing an accurate facility baseline using utility analysis, EMCS/DCIM trend data, energy modeling, and equipment inventories helps identify where energy is being consumed and which systems have the greatest opportunity for efficiency improvement.
Optimizing energy efficiency across IT and facilities	Cooling systems are often the largest non-IT energy consumer within modern data centers. High-efficiency chilled water systems, optimized pumping strategies, free cooling approaches, airflow management, and operational optimization can significantly reduce overall facility energy consumption.
Powering data centers with renewable and on-site energy	Utility interconnection timelines are becoming a major constraint for hyperscale AI data center development. Renewable generation assets, battery energy storage systems, and distributed generation strategies are increasingly being evaluated to reduce grid dependence, improve resiliency, and support large-scale computational growth.

Heat reuse and circular resource strategies	As data center loads continue increasing, facilities are increasingly evaluating whether waste heat can be recovered and reused for adjacent buildings, campus heating systems, domestic hot water generation, or integration with distributed generation infrastructure.
Data center infrastructure management (DCIM) platforms	DCIM platforms help integrate IT infrastructure, cooling systems, electrical infrastructure, and operational controls into a centralized operational platform. These systems support real-time monitoring, capacity planning, thermal analysis, infrastructure visualization, and long-term operational optimization.
Common challenges	Most data centers no longer operate exactly as originally designed due to changing IT requirements, increasing rack densities, equipment additions, and evolving operational demands. Facilities often struggle with infrastructure coordination, operational visibility, and converting large amounts of monitoring data into actionable insight.

Looking beyond PUE

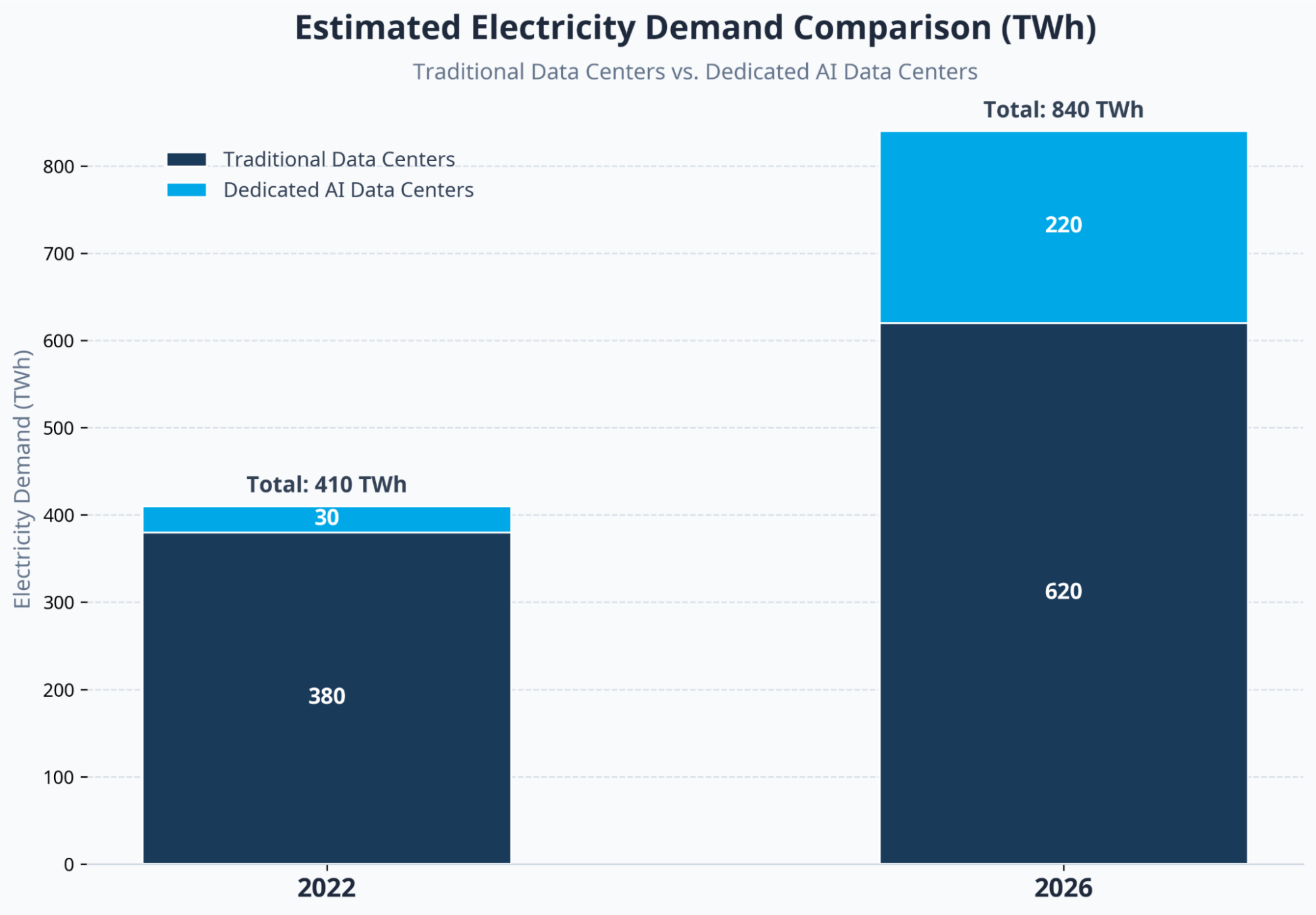
Power usage effectiveness (PUE) has long been one of the most widely used metrics for evaluating data center energy efficiency. In simple terms, PUE compares the total amount of energy consumed by a facility to the amount of energy delivered directly to IT equipment. Lower PUE values are generally associated with more efficient facilities because a larger percentage of the incoming power is being used for computing rather than cooling or support infrastructure.

This is the formula:

$$PUE = \frac{Total\ Facility\ Energy}{IT\ Equipment\ Energy}$$

While PUE remains useful for benchmarking facility performance, it has limitations, particularly when comparing facilities built for very different computing environments. For example, a data center constructed in 2017 may have been designed around rack densities near 5 kW per rack, while modern AI-focused facilities commonly operate in the 40–100+ kW per rack range. Some next-generation GPU cluster designs are reportedly targeting densities approaching 120–240 kW per rack. These increases fundamentally change cooling requirements, power distribution strategies, and the relationship between IT infrastructure and facility systems.

As rack densities continue to increase, power availability and interconnection timelines and grid power availability are becoming major constraints for hyperscale AI data center development.



The shift from traditional to dedicated AI data centers over four years.

Another limitation of PUE is that increasing rack density can artificially improve the metric even when portions of the supporting infrastructure become less efficient. As AI workloads increase rack densities from traditional enterprise levels near 5 kW per rack to modern deployments exceeding 100 kW per rack, the IT load within the PUE calculation increases dramatically. Cooling energy also increases, but many other facility loads—such as lighting, administrative spaces, security systems, and portions of the mechanical and electrical infrastructure—remain relatively constant on a per-square-foot basis. As a result, the ratio between total facility energy and IT energy may improve simply because the denominator grows faster than the supporting infrastructure loads.

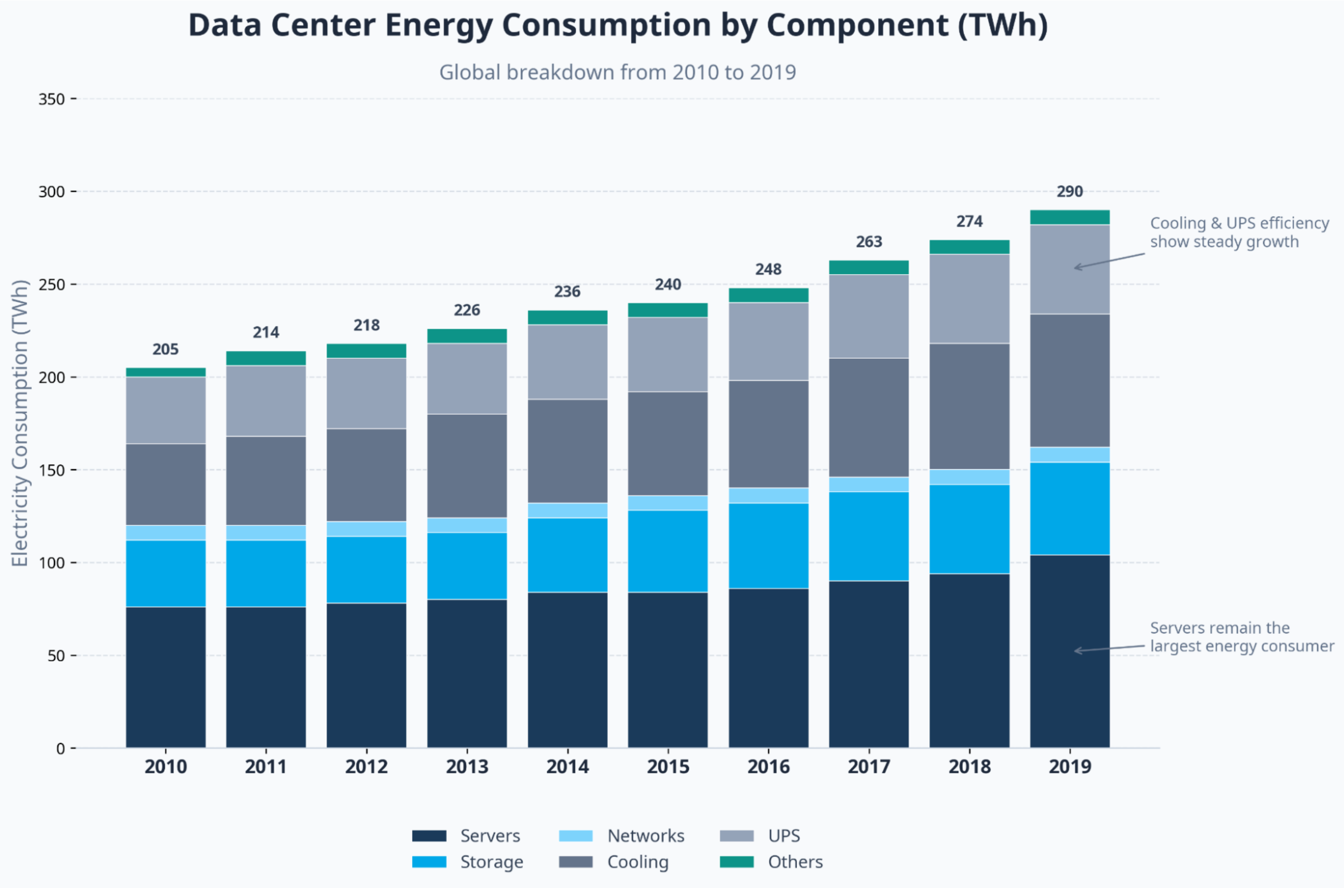


Chart showing the increase in data center annual power consumption over time, with data segments of servers and cooling driving the growth compared to other data segments. Total data center power consumption for 2026 is projected to be 840 TWh.

For example, a legacy data center operating at 5 kW per rack with a highly efficient chilled water cooling system operating near 0.3 kW/ton may ultimately report a higher PUE than a modern AI-focused facility operating at 100+ kW per rack using less efficient air-cooled chillers or DX cooling systems operating closer to 0.8 kW/ton. While the second facility may consume substantially more cooling energy per unit of heat removed, the extreme increase in computational density can still drive the overall PUE downward. This leads to an important distinction between operational efficiency and metric performance: A lower PUE does not always indicate that a facility is operating more efficiently in a broader engineering or environmental sense.

Another limitation of PUE is that it primarily measures site energy consumption rather than source energy consumption. It also does not necessarily measure how sustainably the electricity itself is generated or whether waste energy is recovered and reused elsewhere. The result is that two facilities with similar PUE values may have significantly different carbon footprints depending on their source of power generation, cooling type and design, or operational strategy.

For this reason, modern green data center strategies increasingly focus on system-level performance rather than a single benchmark. Understanding how and from where energy enters the facility, how it is consumed, and ultimately where it goes provides a more complete framework for evaluating overall sustainability.

Additional metrics that are available for use in defining the energy efficiency and sustainability of an AI data center include:

- Energy reuse effectiveness (ERE) evaluates how much energy is reclaimed and reused, often in the form of waste heat, and it subtracts that energy from the total facility energy before calculating PUE.
- Performance per watt (PPW) measures the actual computing power of IT connected devices against the energy they consume, allowing for the efficiency of the actual workloads being performed to be calculated rather than comparing the server loads to the facility as a whole.
- Water usage factor (WUF) measures facility water consumption against local water availability, which provides a metric to help understand the direct environmental impact of a data center on the local environment.

Water usage effectiveness (WUE) is also becoming an increasingly important metric in modern data center design, particularly as AI-driven computational densities continue increasing cooling demand. With this increase in cooling demand often comes an increase in water consumption, particularly true for water-side cooling systems and when liquid cooling is used directly for rack cooling. The higher water consumption can be prohibitive when trying to design a green data center, and needs to be balanced with energy savings and carbon reduction goals.

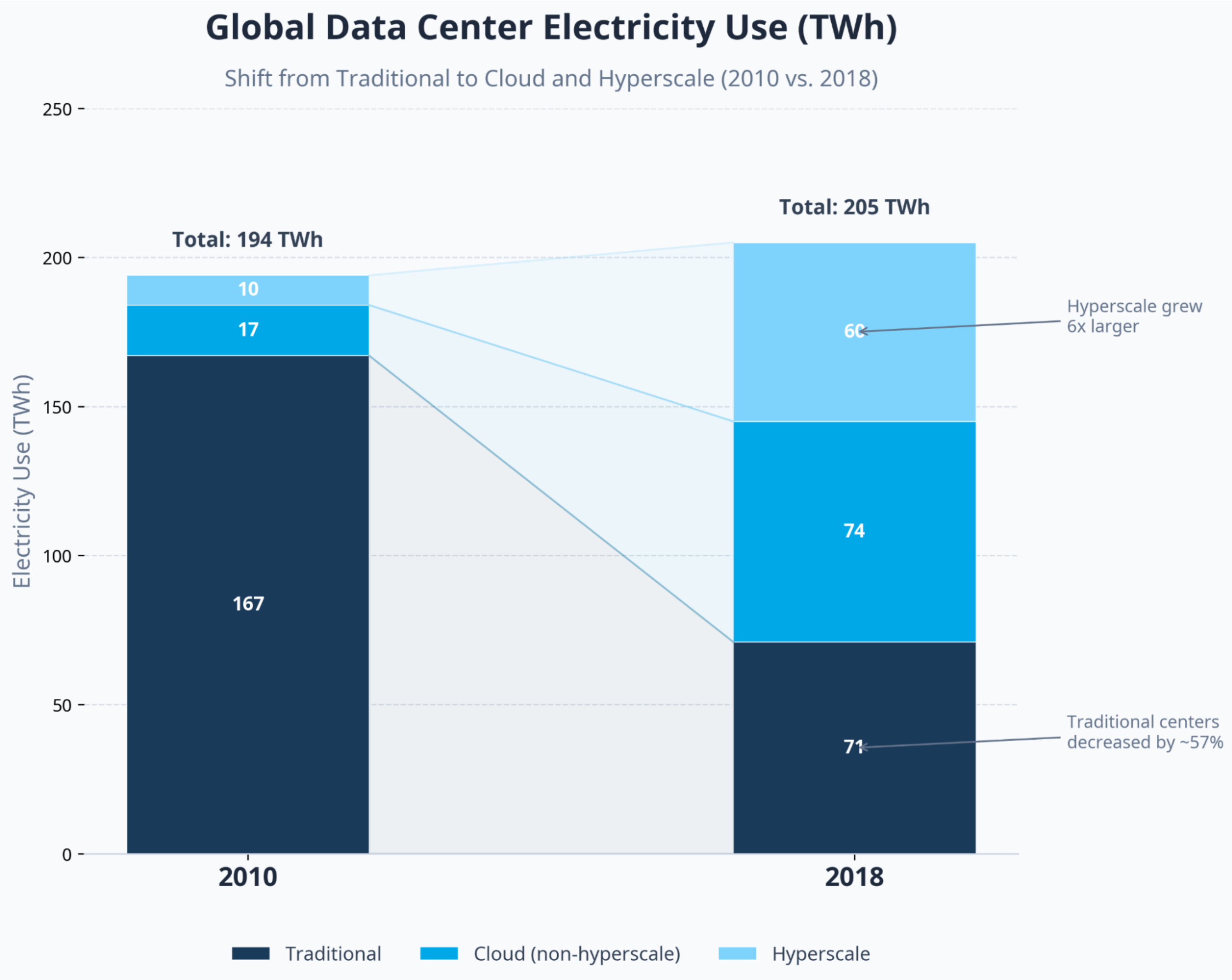
This becomes even more important in large AI-focused data centers where extremely high rack densities can require massive amounts of heat rejection and cooling infrastructure to maintain acceptable operating conditions, particularly when that increased water consumption is viewed against localized water availability.

Establishing a baseline

Before improving a data center, you first need to understand how the facility is actually using energy. Without that, it becomes difficult to identify which systems are worth targeting and whether implemented measures are actually improving performance.

For a new data center, this baseline is usually developed through energy modeling utilizing projected operational loads from equipment lists, occupancy schedules, and mechanical drawings. Existing facilities typically require a utility analysis utilizing at least 12 months of historical utility data, established trend logs from an energy management control systems (EMCS), and equipment inventories and spot measurements to understand how energy is being consumed throughout the building.

This becomes increasingly important with AI data centers because the overall facility load profile is changing rapidly. Traditional enterprise data centers may have operated at around 5 kW/rack, while modern AI facilities can exceed 100 kW/rack. As computational density increases, cooling loads increase with it, and in many cases, cooling becomes the dominant non-IT energy consumer within the facility.



The shift in energy consumption from traditional to cloud and hyperscale (AI) data centers. This shift translates through how electricity is being consumed as well, changing how metrics like PUE are viewed.

At the same time, many other building systems do not scale proportionally with rack density. Lighting systems, security systems, office spaces, and portions of the electrical and mechanical infrastructure may remain relatively constant on a per-square-foot basis even as server loads increase dramatically. This changes how overall facility energy use should be evaluated.

A meaningful baseline should be developed using data collected from the following major energy consumers within the facility:

- IT/server loads
- Lighting systems
- Chilled water systems
- DX/CRAC cooling systems
- Air-side HVAC systems
- Electrical infrastructure and associated equipment
- Distributed generation and resiliency infrastructure

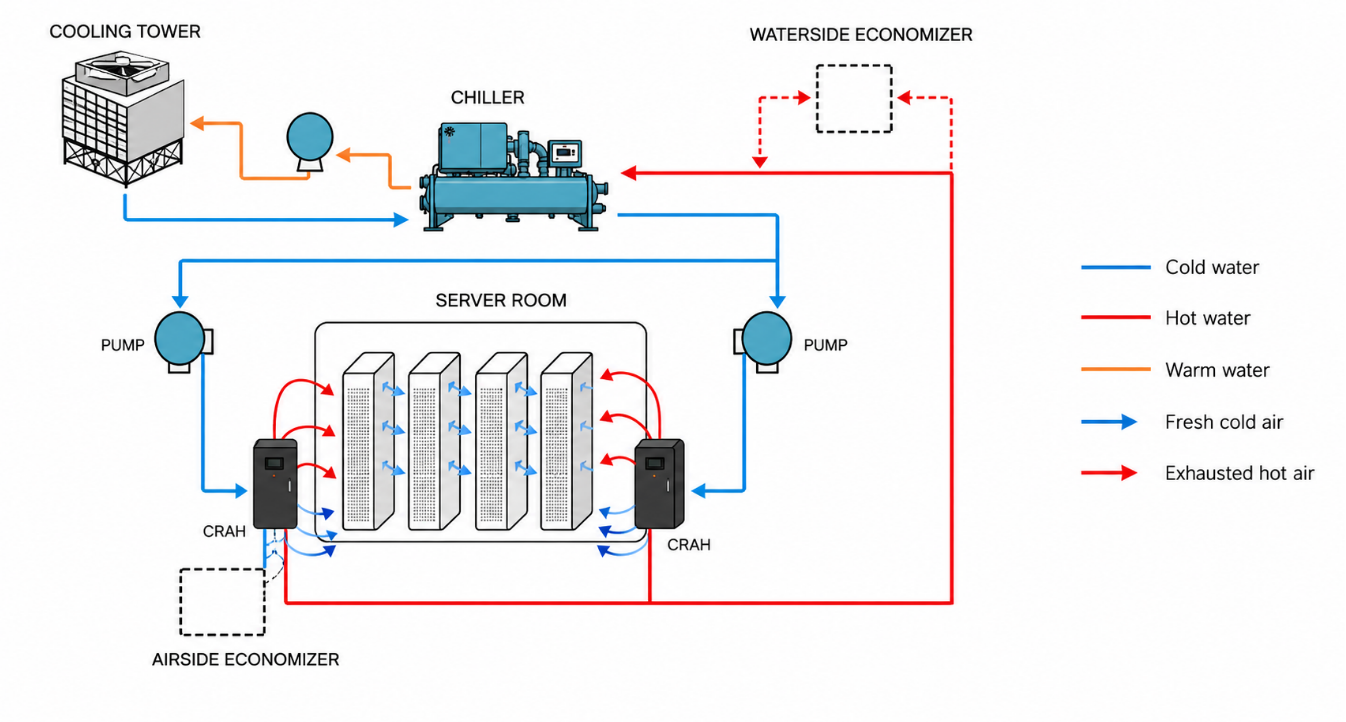
Once an accurate baseline has been established, the facility can begin identifying systems with the greatest opportunity for efficiency improvement. These opportunities may be identified through ongoing EMCS/DCIM monitoring, retrocommissioning activities, third-party energy studies, or engineering evaluations.

Optimizing energy efficiency across IT and facilities

Once facility energy use is understood and major loads have been identified, systems can begin to be evaluated for optimization opportunities. These opportunities may be identified through retrocommissioning activities, third-party engineering studies, EMCS/DCIM trend analysis, or ongoing operational reviews. In many modern data centers, cooling systems quickly become the primary focus because they represent the largest non-IT energy consumer within the facility, particularly as rack densities continue increasing.

For large AI-focused data centers, water-cooled chilled water systems should increasingly be viewed as the minimum standard for high-efficiency cooling design. As rack densities continue increasing, traditional air-cooled DX systems and air-cooled chillers become increasingly difficult to justify from both an operational efficiency and long-term energy cost standpoint.

In addition to the primary chilled water equipment, overall cooling system efficiency can be significantly increased by appropriate pumping strategies and chilled water and condenser water temperature reset strategies. Also valuable are "free" cooling approaches that allow for chilled water to be generated during favorable outdoor air conditions using only cooling towers, which lets chillers be bypassed. As computational densities continue increasing, cooling system optimization will increasingly become one of the primary drivers of overall facility energy performance.



A typical data center schematic utilizing a water-cooled system with optional energy efficient equipment in the form of water-and air-side economizers.

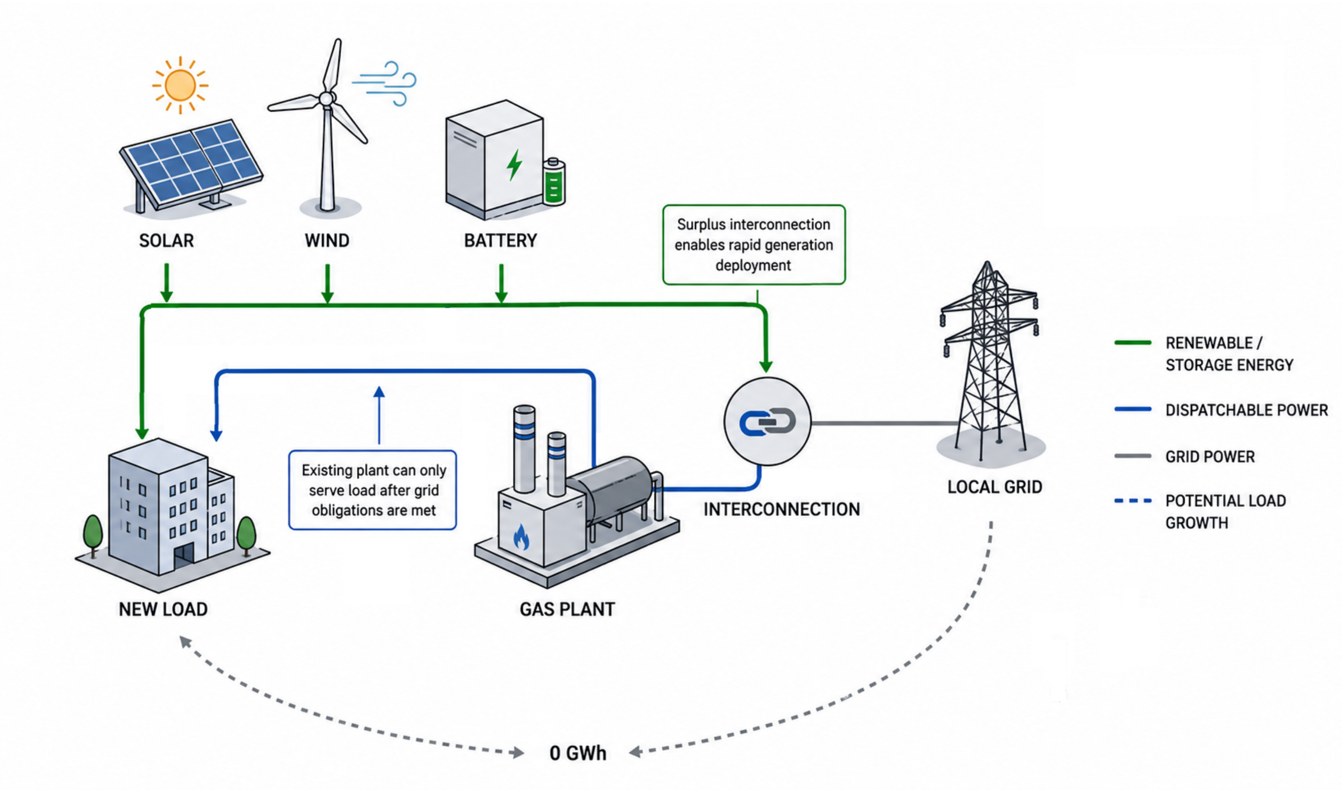
Airflow management is another area that can often be targeted for optimization because relatively minor changes can have a large impact on cooling performance. Poor containment, bypass airflow, recirculation, cable congestion, and improper rack layouts can all reduce cooling effectiveness and increase fan energy consumption. In many facilities, these issues can be corrected with operational or layout changes before major infrastructure replacement is required.

Energy efficiency improvements should also extend beyond the physical cooling plant and airflow management themselves. Lighting systems, control strategies, equipment scheduling, and overall operational sequencing can all affect energy load profiles. In many existing facilities, systems no longer operate the way they were originally designed due to years of equipment additions, operational changes, and evolving IT requirements. Continuous commissioning and operational review are often necessary to maintain efficiency as facility conditions change over time.

Powering data centers with renewable and on-site energy

Power availability is rapidly becoming one of the largest constraints facing new hyperscale and AI-focused data center development. In many regions, utility infrastructure expansion and interconnection timelines are now measured in years rather than months, particularly for facilities requiring hundreds of megawatts of capacity. As a result, many hyperscalers are increasingly evaluating strategies that partially or fully islanded data center campuses from the utility grid using combinations of on-site generation, renewable energy assets, and battery energy storage systems (BESS).

Renewable generation assets, especially utility-scale solar installations coupled with battery storage, are becoming more common in power generation planning for hyperscale AI data centers. Although it is difficult for renewables alone to satisfy the strict uptime and reliability demands of these facilities, when coupled with distributed generation via fossil fuel or other traditional power sources, they can help reduce reliance on the grid, offset a portion of overall electrical demand, and support broader resiliency objectives.



How renewable energy can be colocated and integrated into a data center to provide power that is not only resilient and reliable, but sustainable as well.

As these facilities continue increasing in size and density, power generation asset selection is becoming a major part of overall data center design. Developers are no longer just evaluating utility rates or backup generation strategies. They are evaluating how to provide large amounts of reliable power within realistic project schedules while balancing resiliency, operational cost, maintainability, scalability, and environmental impact. In many cases, the discussion of green data centers is now shifting from simply reducing facility energy consumption to fundamentally changing how these campuses are powered.

Heat reuse and circular resource strategies

Nearly all of the electrical energy consumed within a data center ultimately becomes heat. Historically, most facilities have treated this as a waste product that simply needed to be removed from the space as efficiently as possible. As data center loads continue increasing, particularly within large AI-focused facilities, there is increasing interest in whether portions of this waste heat can instead be recovered and utilized as part of a broader campus energy strategy.

One of the major challenges associated with heat recovery in traditional data centers is that many air-cooled systems produce relatively low-grade heat that is difficult to utilize economically. Large volumes of low-temperature air are generally much harder to recover and distribute than concentrated water-side heat transfer. As facilities continue moving toward liquid cooling systems, however, the feasibility of larger-scale heat recovery strategies may improve due to higher water temperatures and more concentrated heat accumulation.

Potential applications for recovered heat may include use in adjacent buildings, domestic hot water generation, industrial process loads, or broader campus heating systems. In large-scale applications such as hyperscale AI data centers, waste heat recovery may be utilized as part of a combined heat and power (CHP) plant or other distributed generation assets to generate on-site power.

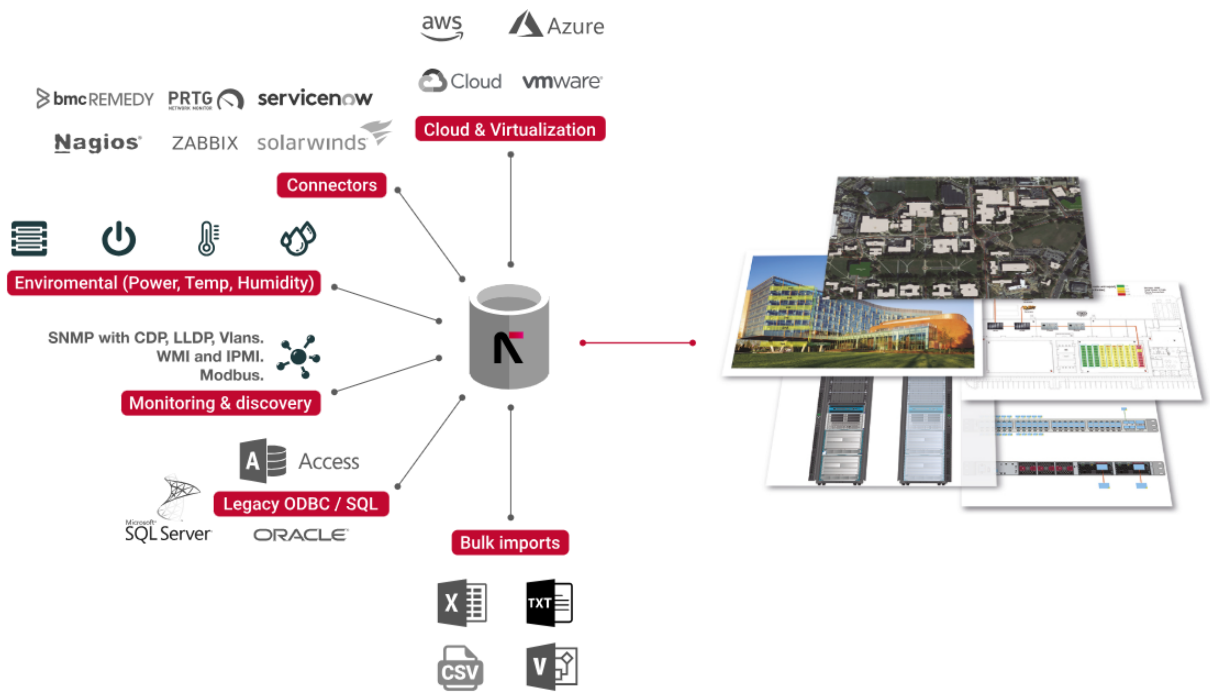
Data center infrastructure management (DCIM) platforms

DCIM platforms help connect IT infrastructure, cooling systems, power distribution equipment, and facility operations into a centralized operational view. Monitoring and controlling systems is a major part of maintaining long-term energy efficiency strategies within modern data centers, particularly as facilities continue increasing in density and operational complexity. DCIM platforms are intended to help connect IT/server loads, cooling systems, electrical infrastructure, and broader facility systems into a more centralized operational platform.

Cisco says that the key benefit of DCIM is unified operational visibility. DCIM systems provide the ability to track PUE and other parameters and matrices as well as to monitor in real time how changes in rack density affect cooling efficiency and overall facility performance. This is becoming increasingly important in AI data centers with high power densities, where relatively small changes in server deployment can lead to a significant impact on cooling infrastructure and power distribution systems.

DCIM platforms can also support rack-level power monitoring, thermal analysis, airflow optimization, capacity planning, and future infrastructure expansion planning. In many facilities, these systems are also being integrated with EMCS/BMS platforms and electrical power monitoring systems to provide broader visibility across both IT and facility infrastructure.

At the same time, simply collecting more data does not necessarily improve operational performance. One of the common issues with both DCIM and traditional building monitoring systems is that facilities often generate significantly more information than operations staff can realistically act on. Poor alarm management, passive monitoring strategies, and a lack of coordination between IT and facilities groups can result in systems generating “noise” rather than actionable operational insight.



How DCIM systems such as Graphical Networks netTerrain DCIM software can help you manage your infrastructure data and turn it into actionable information.

As facilities continue to increase server power densities and associated infrastructure complexity, DCIM systems will continue becoming more important for maintaining operational uptime, optimization system efficiencies, and supporting long-term infrastructure planning.

Common challenges

Most data centers operate differently than how they were originally designed. Changes in loads, equipment replacements, operational changes, evolving IT requirements, lack of proper maintenance, designed hysteresis, or even aging equipment all lead to building operations that shift away from the original design.

Coordination and communication between IT and facilities is a known gap in most data centers. Server deployments, rack densities, and associated cooling requirements often change quickly. These changes can often lead to mechanical and electrical systems that are not sized properly and not able to adequately meet the new loads.

Another issue is that many facilities collect large amounts of operational data but have difficulty turning that information into actionable insight. Monitoring systems can generate an overwhelming amount of information, making it harder for operators to isolate the conditions that are actually affecting reliability, cooling performance, or energy use.

As AI applications continue pushing rack densities and overall facility loads higher, these types of operational issues are becoming harder to ignore.

A practical path forward

Improving sustainability in data centers usually starts with gaining a clearer understanding of how the facility is actually operating today and where the biggest inefficiencies exist. In many facilities, operational improvements and better system visibility can provide more immediate value than large-scale infrastructure replacement projects.

As power densities and computing requirements continue to increase, data centers will need to be more flexible in how they operate. Facilities that can adjust cooling approaches, redistribute loads, and adapt supporting infrastructure as conditions change will likely have an easier time maintaining reliable and efficient operation over the long term. In order to do this effectively, equipment operational data will need to be measured, logged, analyzed, and utilized in an effective manner to ultimately optimize system performance.

Conclusion

Data centers continue to evolve quickly, particularly as AI workloads increase power consumption and cooling requirements. At the same time, the industry definition of a “green” facility is changing as well. Metrics such as PUE still provide useful information, but they no longer capture the full picture of overall sustainability or operational performance.

Metrics beyond PUE should be considered when defining a modern green AI data center, such as ERE, PPW, and WUF. These additional metrics, combined with adjusted PUE and WUE calculations, can provide a more holistic view of the overall efficiency of a data center and how “green” the design actually is.

A more complete approach requires looking at the facility as an integrated system rather than individual systems. DCIM allows the facility to integrate power generation, cooling systems, infrastructure utilization, operational controls, and opportunities for heat recovery or energy reuse into a coherent system and optimize that system from an energy standpoint.

As facilities continue moving toward higher rack densities and larger electrical demands, engineers and operators will need to balance efficiency, resiliency, scalability, and sustainability rather than treating them as separate objectives. This is one of the areas where DCIM and infrastructure visualization platforms can provide meaningful value by helping facilities better understand relationships among IT infrastructure, cooling systems, electrical distribution systems, and available capacity, all in real time. Facilities with strong operational visibility, integrated infrastructure management, and practical engineering oversight will likely be better positioned to support future