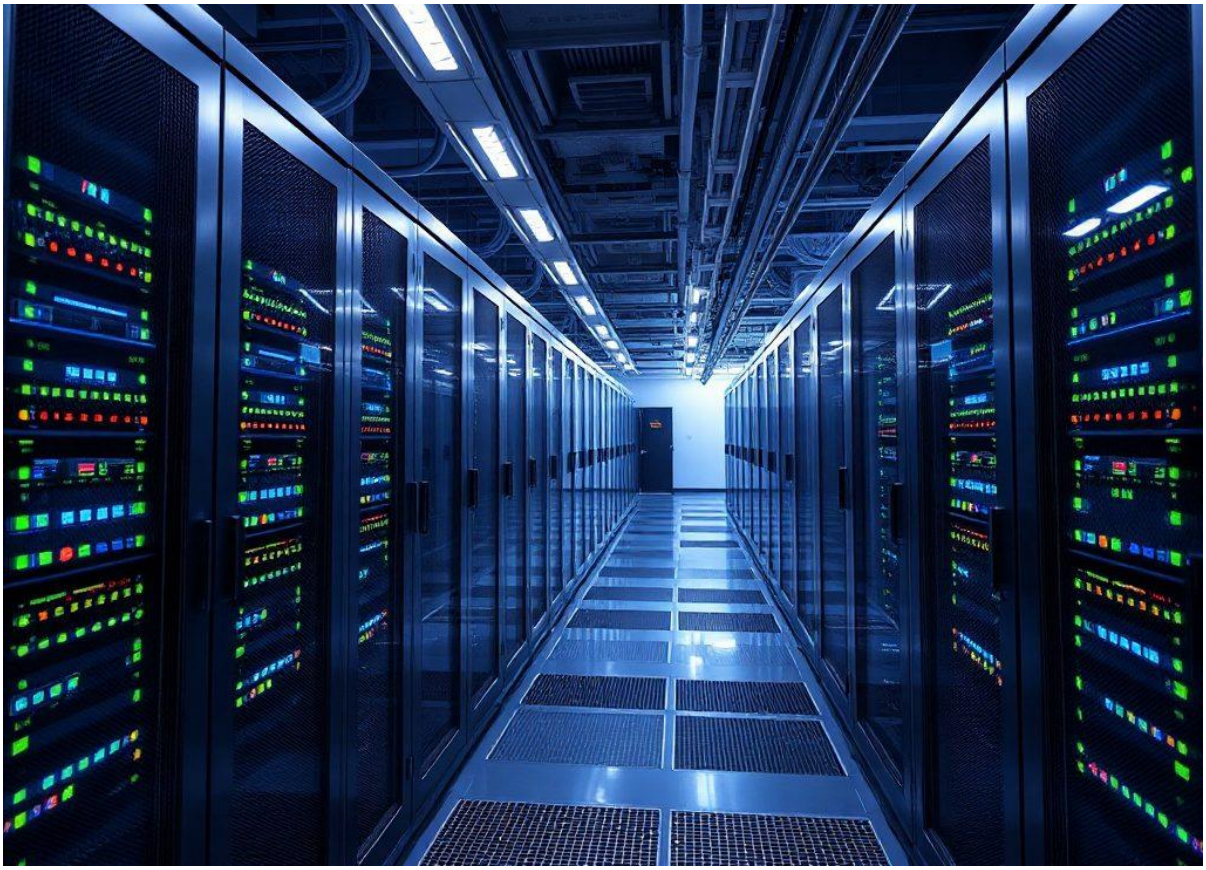

MARCH 23, 2026

# PUE in data centers: useful indicator, but only valuable in the right context

**Power Usage Effectiveness (PUE) is without a doubt the best-known efficiency metric in the data center industry. The simplicity of the metric has ensured that PUE is widely applied, cited and even used as a marketing tool. At the same time, that same simplicity is also the biggest pitfall. If you use PUE separately, without context or additional metrics, you can come to the wrong conclusions. In the worst case, incentives are even created that are at odds with actual energy savings.**

In this article, I outline the genesis of PUE, explain in simple terms what PUE measures, why a decreasing PUE does not automatically mean that a data center becomes more efficient, and why differences between hyperscalers and co-location data centers are essential. In addition, I introduce the Server Idle Coefficient (SIC) as a necessary complement and show why current PUE thinking sometimes creates an undesirable economic incentive — especially in co-location environments.

## From PUE 3.0 to 1.1: a brief history

When PUE was introduced by The Green Grid in 2007, the average efficiency of data centers was low. Many enterprise and early co-location data centers were running with heavy oversizing, inefficient UPS systems, and little attention to airflow and thermal design. PUE values around 2.5 to 3.0 were no exception.

The sector has made enormous strides in the past fifteen years. Hot and cold separation separation, containment, free cooling, higher flow temperatures, more efficient UPS topologies and advanced control technology have drastically reduced facility overhead. Hyperscalers, in particular, have been making these optimizations on an industrial scale, leading to PUE values around 1.1 and sometimes even below at specific locations or time periods.

Importantly, however, these values are especially true for best-in-class environments. Global surveys show that the average PUE is much higher, partly due to older data centers, multi-tenant environments and variable occupancy rates.

## What is PUE, in simple words?

PUE answers one key question:

*How much energy does a data center need to get 1 kWh of energy to the IT equipment?*

The formula is simple:

*PUE = total energy of the data center / energy consumption of the IT equipment*

- The numerator: the total energy consumption of the data center (IT, cooling, UPS losses, power distribution, lighting, security).
- The denominator: the energy consumption of the IT equipment itself (servers, storage and network).

A PUE of 2.0 means that for every kWh used by IT, an additional 1 kWh is needed for the supporting infrastructure. A PUE of 1.1 means that the overhead is only 0.1 kWh.

## The main misconception: PUE is not a measure of total efficiency

PUE only measures the efficiency of the facility shell in relation to IT consumption. The metric says nothing about:

- how much useful work the IT actually does,
- how efficiently servers are deployed,
- how much $CO_2$ emissions are associated with that energy,
- whether the absolute energy demand increases or decreases.

And here is a crucial point:

*A decreasing PUE does not automatically mean that a data center saves energy*

Because PUE is a ratio, it can improve because:

1. the facility overhead decreases (e.g. better cooling), or

2. IT consumption increases while overhead remains the same. In the second case, the PUE appears "better", while the absolute energy consumption actually increases.

## Hyperscalers vs. Colocation: Apples and Pears

Hyperscalers have structural advantages:

- full control over IT and facilities,
- high and stable occupancy rates,
- homogene hardware en airflow,
- optimization on cost per workload.

For them, idle IT is immediately a cost item. This explains why they combine low PUE values with active IT efficiency.

This is different with co-location, especially in multi-tenant environments:

- heterogeneous racks and power densities,
- fluctuating occupancy rate,
- limited control over customer IT architecture,
- additional overhead for flexibility and redundancy. A higher PUE does not necessarily mean worse design, but often a different business reality.

## The missing link: the Server Idle Coefficient (SIC)

Where PUE stops, the Server Idle Coefficient begins. The SIC provides insight into the part of the IT energy consumption used by servers that do little or no work. In many environments, servers run in a power-setting mode where with little or no workload, the energy consumption does not decrease but remains almost constant. As a result, a large part of IT consumption is structurally idle energy.

## The perverse incentive: more efficient IT can lead to a worse PUE

And this is where a fundamental field of tension arises. When customers optimize their IT environment — through consolidation, virtualization or by enabling power-management settings — IT energy consumption decreases. That means:

- the denominator of the PUE becomes smaller,
- While a large part of the facility energy remains relatively constant.

The result:

- the PUE increases, despite actual energy savings.

In concrete terms, this means that the more efficiently the IT is used, the worse the PUE score can be.

## Why this is an unwanted stimulus

1. **PUE rewards volume, not IT efficiency** – A higher IT load improves PUE, even if that load is mostly idle. PUE thus indirectly stimulates more IT consumption instead of smarter IT use.
2. Colocations **often earn from electricity** – In many colocation models, electricity is passed on to customers, often with storage. Less IT consumption at customers means:• lower revenue,• while the fixed costs of the data center remain largely the same.

This makes it clear that: the current KPI thinking (PUE) and the revenue model are not automatically aligned with IT efficiency on the customer side. Hyperscalers hardly have this problem, because they own both IT and facility and optimize on total cost per workload.

## Transparency and absolute figures: the competitive argument

Data centers often indicate that they do not share absolute energy figures due to competitive sensitivity. Although understandable, this argument is becoming less and less tenable. Regulation, grid congestion and social accountability require more than just a rationale.

Transparency is possible without commercial damage, for example by:

- aggregate figures,
- bandwidths,
- Standardization per rack or per kW,
- independent assurance.

## PUE is necessary, but not sufficient

PUE has helped the sector to make enormous efficiency gains in the facility infrastructure. But in a world of grid congestion, AI workloads, and societal pressure, PUE alone is insufficient. Without additional metrics such as the Server Idle Coefficient, a paradox arises in which:

- IT efficiency is "punished"
- absolute energy savings are not visible
- and wrong economic incentives arise

If you want to seriously assess data centers, you have to look beyond PUE alone and include the entire chain: from workload to watts.