Life Is On | Schneider Electric

Topic A: Power Implications of White Space and Rack-Level Densification in the Era of Accelerated Parallel Compute

Executive Summary

Fueled by the proliferation of IoT devices and the deployment of 5G networks, the demand for data centers is rapidly increasing to accommodate the vast amounts of data generated and the need for ultra-low latency, real-time processing. All this to accommodate the vast amounts of data generated and the need for ultra-low latency, real-time processing. As the global demand for data and real-time processing escalates, driven by AI, machine learning, high-performance computing (HPC), and edge applications, data centers are undergoing a profound transformation.

A core aspect of this evolution is the densification occurring within the white space. This densification is fundamentally altering the power profile of modern data centers and is being driven largely by the proliferation of accelerated parallel compute technologies like GPUs, TPUs, and custom ASICs. This densification will impact not only AI factories and Hyperscale facilities but also Regional Colocation, On Prem Enterprise Data Centers and IDF's/MDF's all the way to the Network Edge.

This convergence of factors is reshaping the foundational characteristics of traditional data center topologies while densifying the white space. Key dynamics of this densification include:

- **White Space Densification**
  Data centers are moving from lower-density deployments (3–10 kW/rack) to high-density layouts exceeding 25 – +132 kW/rack. A single 24 Rack cluster, occupying 560 sq ft of whitespace (including setbacks) can consume more than 2,300kW. This shift allows operators to maximize compute capacity within fixed footprints but dramatically increases power concentration per square foot. Power infrastructure including UPS systems, switchgear, PDUs, busway, and cabling must now support significantly higher and more variable loads.

- **Rack-Level Densification**
  The adoption of accelerated parallel compute architectures requires greater power per node. A single rack populated with AI/ML workloads can now range from 25–132 kW, compared to the 3–10 kW range found in traditional enterprise deployments. This pushes the limits of existing power distribution and requires high density circuit design, busway systems, high density PDU's, RPP's, and integrated intelligent power monitoring to ensure reliability, resilience, compliance, and scalability.

- **Impact of Accelerated Parallel Compute**
  Accelerated compute workloads don't just increase average power draw; they create peak demands that fluctuate rapidly based on workload scheduling. These high-density workloads:
  - Demand more granular and responsive power delivery.
  - Necessitate real-time telemetry and intelligent power orchestration.
  - Introduce thermal loads that tightly couple power and cooling strategies, especially in liquid or hybrid-cooled systems.

For data center owners and operators, adapting to these changes is crucial. Embracing densification strategies will enable them to meet the growing demands for data processing and storage while ensuring operational efficiency, reliability, resilience, and sustainability.

Background

Densification of the white space in a data center refers to increasing the amount of computing power per square foot. Traditionally, data center rack densities have ranged from 3 to 10 kW per rack, which accommodates typical enterprise workloads such as web hosting, file servers, and moderate virtualization. These systems were relatively power-efficient and could be supported effectively using a variety of methods and architectures commonly found within the data center landscape.

The rise of accelerated computing and AI workloads is pushing these boundaries significantly. AI training, inferencing, and high-performance computing (HPC) workloads rely on dense GPU clusters, which can require 25 to +132 kW per rack in the near term and upwards of more than 600kW/rack in the future. These workloads demand high-speed interconnects, large memory bandwidth, and massive parallel processing, resulting in significantly higher power draw and heat output per unit of space. The shift to such high-density configurations forces data centers to rethink their traditional power protection and distribution strategies, exploring options ranging from higher AC voltage to DC distribution at the rack level.

White space and rack-level densification, driven by accelerated parallel compute, is rewriting the rules of power design in data centers. Operators must rethink power distribution, redundancy, and efficiency at every level, from breaker to chip, in order to stay competitive, sustainable, and resilient in the rapidly evolving, high-density, high-performance era.

Some key challenges and considerations include:

- **Power Availability & Provisioning:**
  Facilities must overprovision for peak load, often with 2N or N+1 redundancy, resulting in higher capital expense and greater energy losses if not actively managed.
- **Efficiency vs. Resilience Tradeoffs:**
  High-density zones increase risk of cascading failure if power is interrupted. Power blocks must be segmented and resilient to isolate faults and ensure uptime.
- **Grid and Sustainability Pressures:**
  Densification amplifies power draw per site, increasing stress on local grids. Operators are under pressure to integrate renewable sources, implement demand-response programs, and improve power usage effectiveness (PUE) through smart design.
- **Power Monitoring & Intelligence:**
  Continuous, rack-level power metering is now critical. AI-driven tools are being used to forecast load, balance phase draw, and optimize performance per watt.

Questions:

Current Power Capacity & Utilization

1. What is the design density (W/SF) of your data center?
2. What is your current average and peak power usage per rack? Are all systems able to handle the peak?
3. What is the maximum power capacity available per rack in your white space?
4. Are there any known limitations to your existing electrical infrastructure?

Power Delivery & Distribution

1. What type of power distribution units (PDUs) are you currently using?

2. Are your PDUs capable of supporting higher density configurations (>20 kW per rack)?
3. Are you using single-phase or three-phase power at the rack level?
4. How is power distributed across your white space (centralized or decentralized)?
5. Would a preference be for Busway or traditional “pipe & wire” for high dense applications (greater than 70kW/rack)?

Scalability and Redundancy

1. Do you have scalability plans for your power infrastructure to support future densification?
2. What level of power redundancy do you have in place (N, N+1, 2N)?
3. Are your UPS systems sized to handle increased load from high-density racks (IT load only or IT plus whatever cooling (CDU pumps) as well)?
4. Have you evaluated the need for additional generator capacity (to handle peak GPU spikes) and, or are you considering other energy storage solutions to support higher power demands and dynamics?

Monitoring and Management

1. Do you monitor real-time power usage at the rack or PDU level?
2. Are you leveraging DCIM or Power Management tools to track and analyze power trends and track capacity / constraints?
3. How are you planning for load balancing across circuits and phases to prevent over-subscription?

Compliance and Planning

1. Do you have a documented plan for upgrading your electrical infrastructure as part of your densification strategy? (Standard HD Power Design)
2. Have you assessed the impact of densification on power maintenance procedures?