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

Executive Summary & Background

Fueled by the proliferation of IoT devices and the deployment of 5G networks, the continued growth of cloud and the surge of AI, the demand for data centers is rapidly increasing! 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 cooling 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 cooling 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. Cooling infrastructure including Hi Temp Chillers, CDU's, Rack Manifolds, Rear Door HX and traditional Air Distributions Systems must now support significantly higher densities and more variable loads. Critical power (UPS support for CDU, Chilled Water Pumps. Thermal Storage (and perhaps critical control systems) to support cooling is now in play more than ever before.

- **Rack-Level Densification**
  The adoption of accelerated parallel compute architectures requires greater cooling 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 air-cooling distribution and requires high density cooling architecture, technology, and operational considerations to ensure reliability, resilience, compliance, and scalability within the rack.

- **Impact of Densification as it relates to Critical Cooling Architectures**
  The push to liquid cooling to support this accelerated compute workload requires a new approach to cooling architecture. This would require Hi-Temp Chillers, Cooling Distribution Units and pending the IT requirements…either "Rack Manifolds" (DTC Cooling) and or Active Rear Door Heat Exchangers. In addition, adding a bit of complexity is the fact that there is an 80-20% / 70-30% split between liquid cooling and traditional air cooling (the lower number is the air requirement). And then finally – special care should be considered as to how to architect the power requirements for the mechanical equipment (A & B power paths / UPS support for select pieces (i.e... CDU pumps, Thermal Storage, CHW Pumps and Rear Door Fans, etc.)

- Cooling demand is more granular and variable (low to high loads) and may require “ride through” power & thermal delivery.
- Necessitate the need for real-time telemetry and intelligent cooling management & 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.

Questions:

Current Cooling Capacity & Utilization

1. What is the design density (W/SF) of your data center?
2. What is your current average and peak THERMAL requirement - usage per rack?
3. What is the maximum thermal capacity available per rack in your white space?
4. Are there any known limitations to your existing cooling infrastructure?

Cooling Delivery & Distribution

1. What type of cooling distribution units (CDUs) are you currently or looking to deploy? (Liquid to Air or Liquid to Liquid)
2. Is your main cooling plant able to meet higher temp cooling requirements?
3. Are you planning on a centralized, POD or Rack level CDU deployment?
   - Do you plan on UPS or ride through power for CDU pumps or other critical cooling equipment?
   - What sizes you considering (Single MW or Multi-MW CDUs)

Scalability and Redundancy

1. Do you have scalability plans for your cooling to support future densification and/or growth?
2. What level of cooling redundancy do you have in place (N, N+1, 2N)?
3. Are your UPS systems sized to handle increased load from critical cooling equipment? Or do you plan on a separate UPS for cooling?
4. Have you evaluated the need for additional storage capacity (thermal or energy) to support higher power and cooling demands?

Monitoring and Management

1. Do you monitor real-time cooling usage at the CDU or rack level?
2. Are you leveraging DCIM or BMS systems to track and analyze cooling trends?
3. Do you need an enterprise-level monitoring and reporting system across your portfolio?
4. Would you consider leveraging your digital monitoring and controls systems data to develop predictive analytics / AI tools?

Compliance and Planning

1. Do you have a documented plan for upgrading your cooling infrastructure as part of your densification strategy? (Standard Design Concept for HD Cooling)
2. Have you assessed the impact of densification on mechanical maintenance procedures / standards?