Smarter technology for all

Cooling the Future: Data Center Strategy Toward 2030.

Wil Wellington
wwellington@lenovo.com

Compute Paradigm Shift

Legacy Design (2010-2023)

Designed for CPU-centric servers with rack densities of 5kW - 15kW. Air cooling was the universal standard, leveraging standard hot/cold aisle containment.

Air Cooling Limit

Air cooling hits a physical wall at ~30kW per rack. Beyond this, the volume of air required (CFM) creates industrial noise and consumes 40%+ of server power for fans.

Future Design (2024-2030+)

GPU-centric architectures (NVIDIA Blackwell and beyond) demanding 100kW+ per rack. Liquid cooling is no longer optional; it is the fundamental utility.

Liquid Efficiency

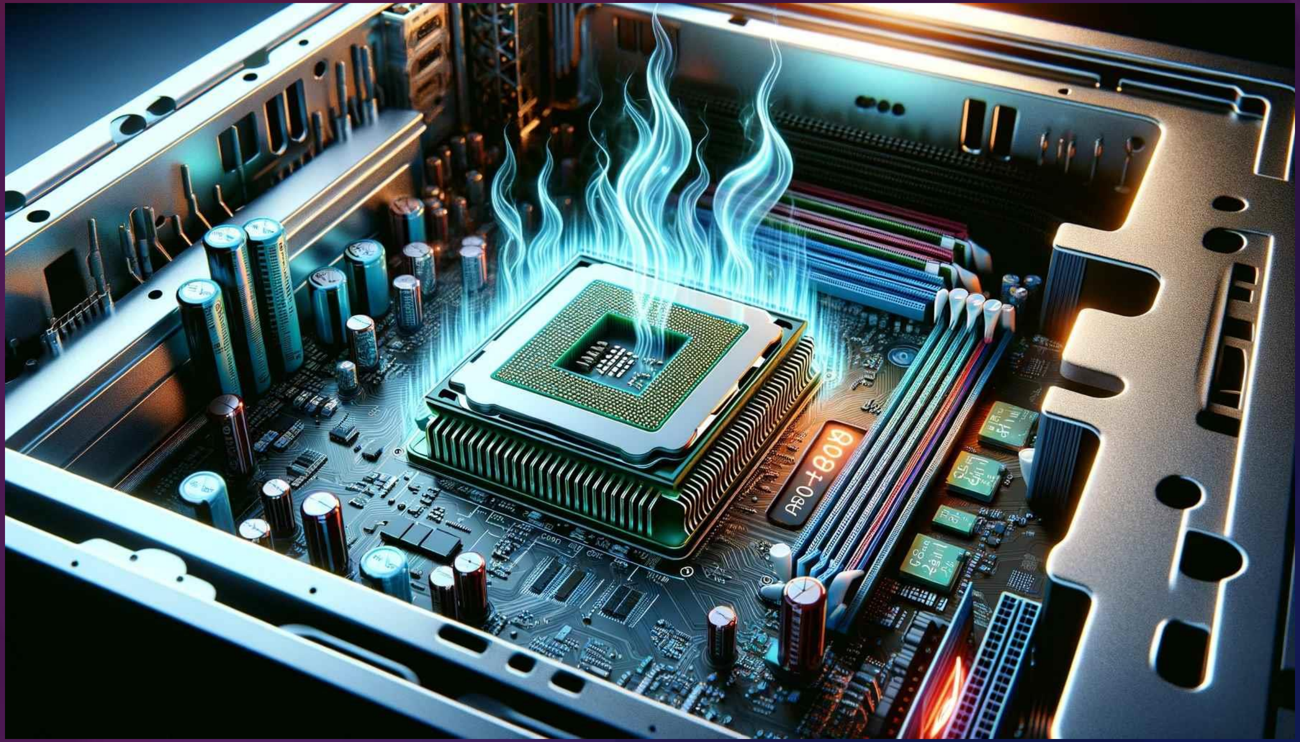
Water is 4,000x more effective at heat transfer than air. Transitioning to liquid reduces fan power to near-zero, enabling PUEs below 1.1.

The 1,500W Thermal Wall

Single-chip TDP is skyrocketing as semiconductor geometries shrink and power delivery increases.

- **H100 (2024):** ~700W TDP
- **B200 (2025):** ~1,000W - 1,200W TDP
- **2030+:** 1,500W+ per chip

Cooling these concentrated heat sources with air is physically impossible due to air's low heat capacity.



Flooring the Future

Pivoting from Raised Floors to Slab Architecture

Structural Load Capacity

AI server clusters (e.g., NVIDIA Blackwell GB200) can weigh upwards of 3,500 lbs per rack.

Slab-on-grade construction eliminates the point-load risks associated with raised floor pedestals, allowing for higher rack packing densities and significantly safer deployments of heavy GPU chassis.

Sustainability Impact

40%

Lower embodied carbon

Concrete Efficiency

By removing the aluminum and steel components of raised floor pedestals and tiles, data center builders can reduce the embodied carbon of their structural build-out by nearly half while improving operational PUE.

The Roadmap to 2030

2020 – 2024

Hybrid facilities; limited 20kW racks.

2025 – 2027

Shift to Liquid-Ready designs.

2028 – 2029

Slab becomes the standard for Hyperscale AI.

2030+

Total phase-out of traditional raised floors.

Energizing the AI Era

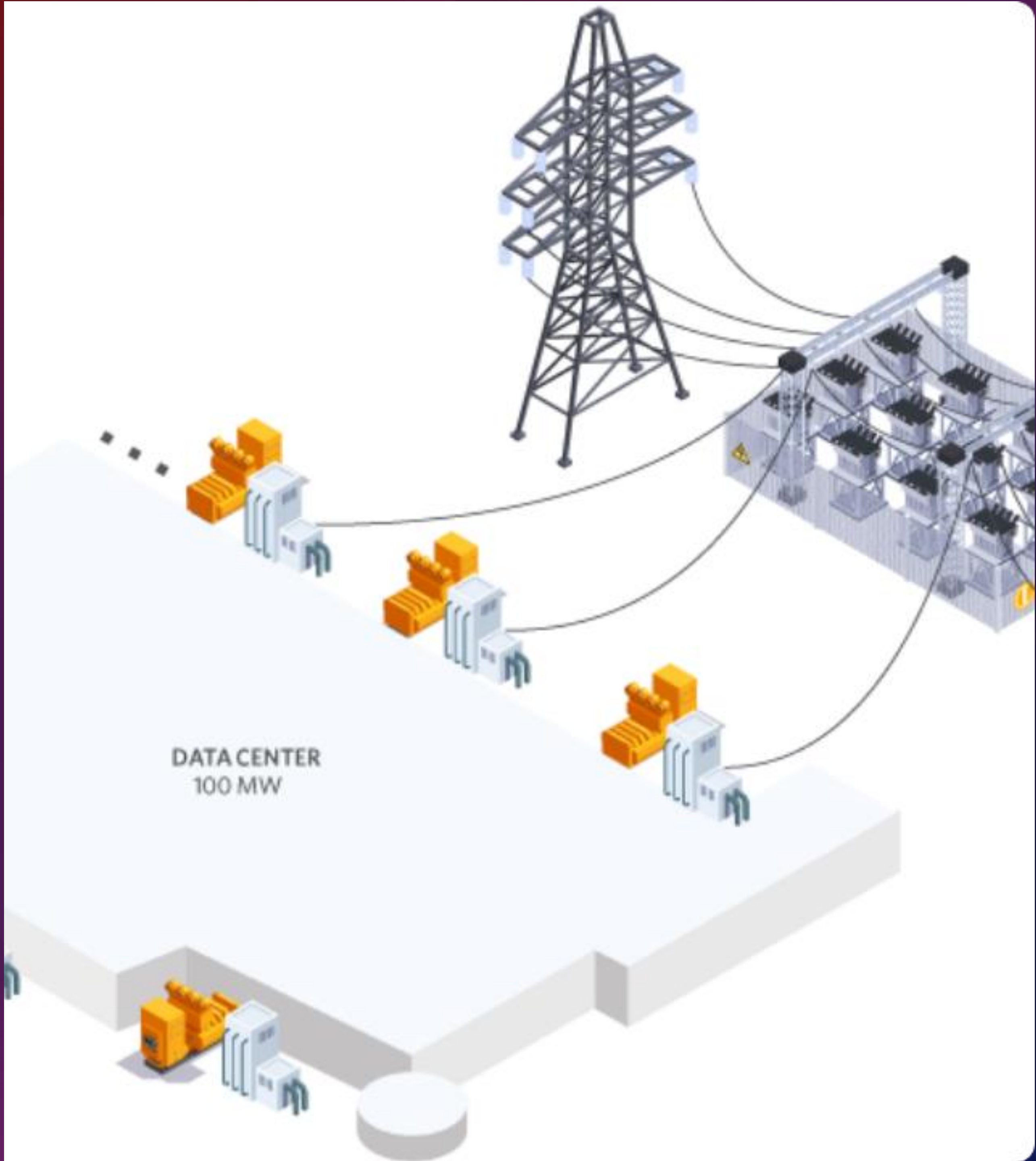
The Grid, The Society, and the Mandatory Shift to Renewables

THE POWER GRID BOTTLENECK

Grid Scarcity

The biggest challenge for 2030 is not the data center itself, but the grid infrastructure to power it. Utility lead times are reaching 5-8 years in key markets.

Future centers must integrate on-site microgrids, SMRs (Small Modular Reactors), and massive battery energy storage systems (BESS).



CLIMATE TIPPING POINT

10%

OF GLOBAL ENERGY BY 2030

800M

METRIC TONS CO_2 POTENTIAL

24/7

REQUIREMENT FOR GREEN BASELOAD

The CDU Bottleneck

Architectural Constraints and the Future of Liquid Cooling Scaling in the AI Data Center

Major Scaling Bottlenecks

Physical Footprint

CDUs consume up to 15% of rack-row real estate, reducing total GPU compute density per square foot.

Complexity & Cost

Massive pump systems and filtration nodes create high CAPEX and intricate secondary loop plumbing.

Pressure Limits

Maintaining uniform 80+ PSI across a row becomes exponentially harder as racks increase distance from the CDU.

Single Point of Failure Risks

Cascade Failure

A single pump failure or seal leak in a centralized CDU can force a thermal shutdown of 12-24 high-density racks simultaneously, leading to massive AI workload interruptions.

The Redundancy Trap

Adding N+1 redundancy at the CDU level further exacerbates the footprint and cost issues, creating a diminishing return on investment as power densities cross 150kW per rack.

Rise of CDU-less Solutions

Removing the intermediate heat exchange layer to achieve direct, facility-scale cooling efficiency

CDU SCALING BOTTLENECKS

THE MOBILITY CRISIS

CDUS are reaching 10mW capacities, causing critical friction:

- **Logistics:** Units cannot fit through standard data center doors.
- **Architecture:** Complexity driving a move toward CDU-less loops.
- **Decommissioning:** Massive units become "permanent" liabilities within the shell.

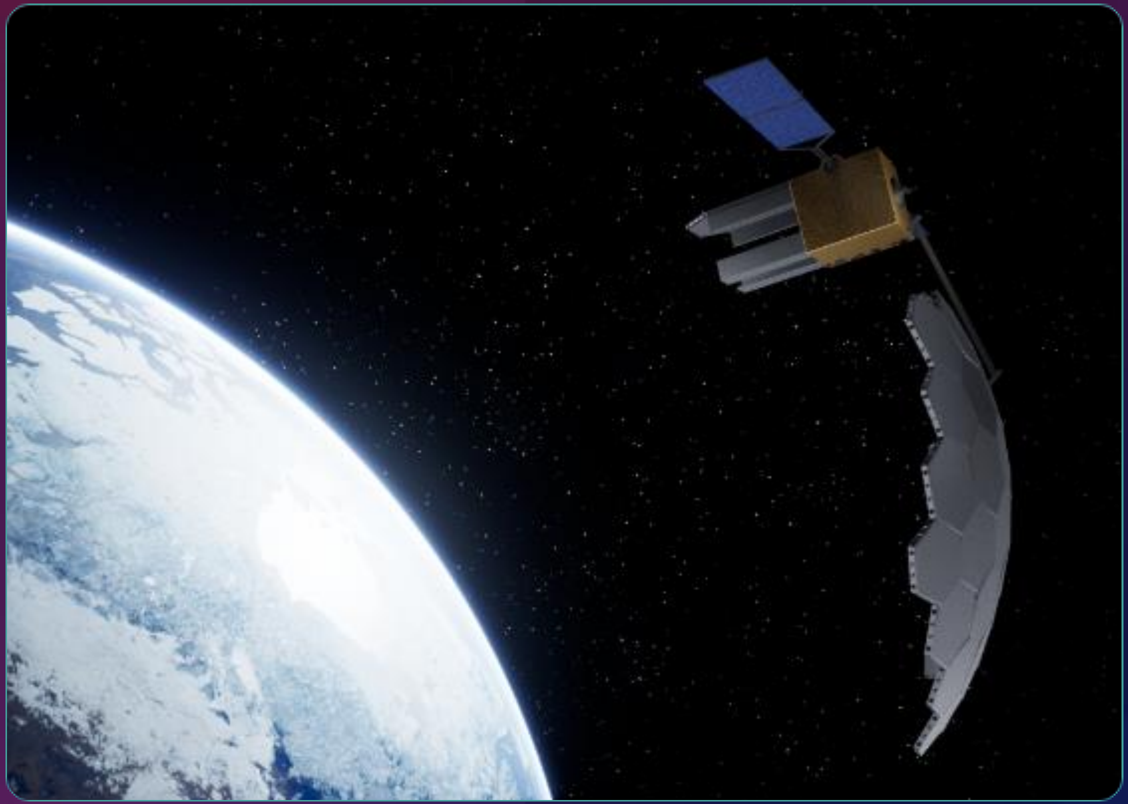
ORBIT: THE SPACE DATA CENTER

THE FAILED VACUUM THEORY

Concept: Infinite cold and solar energy in Earth orbit or Lunar orbit.

Why it won't work:

- **Heat Rejection:** Vacuum is an insulator. Without air/water, heat must move via radiation (massive solar sails).
- **Latency:** Speed of light is too slow for 2030+ edge processing needs.
- **Launch Costs:** $10k/kg makes hardware replacement economically impossible.



SECURITY & SURVIVABILITY

Hardening Infrastructure Against Geopolitical Targets

THE KINETIC THREAT

LESSONS FROM RECENT UNREST

Middle East conflicts have redefined multi-gigawatt facilities as **Giant Targets** for terrorism and kinetic warfare.

Concentrated "Mega-Regions" are now seen as sovereign vulnerabilities. Security must shift to "Hardened Geography."

THE RESILIENCE MANDATE

SUBTERRANEAN HARDENING

Mountain bunker facilities (NORAD/Iron Mountain style) survive kinetic strikes and EMP pulses.

DEPLOYABLE MOBILITY

Mobile data centers moving via railway networks for tactical risk dispersion during conflict.

BSI 200KM SEPARATION

Adhering to the German BSI rule for mandatory catastrophic disaster recovery distance.

Workloads Beyond 2030

The dawn of Physical AI, Real-time Global Digital Twins, and the transition to Quantum-Classical Hybrid computing.

FUTURE WORKLOAD EVOLUTION

Generative Physical AI

AI models that don't just generate text, but simulate physics for robotics and autonomous manufacturing, requiring continuous high-density GPU inference.

Global Digital Twins

Real-time simulations of entire cities, climate patterns, and industrial ecosystems (NVIDIA Omniverse at scale) consuming multi-MW per cluster.

COMPUTATIONAL SCIENCE 2.0

From Simulation to Discovery

Post-2030 computational science moves from "physics-based" (slow) to "AI-surrogate" models (millisecond discovery).

- **Accelerated Materials:** Finding 20 years of materials in 20 days.
- **Digital Biology**: Real-time protein folding and drug interaction simulation at scale.

CATERING FOR QUANTUM READINESS

Cryogenic Hosting

Dilution refrigerators require infrastructure cooled to 10-15 milliKelvin (colder than outer space), separate from GPU liquid loops.

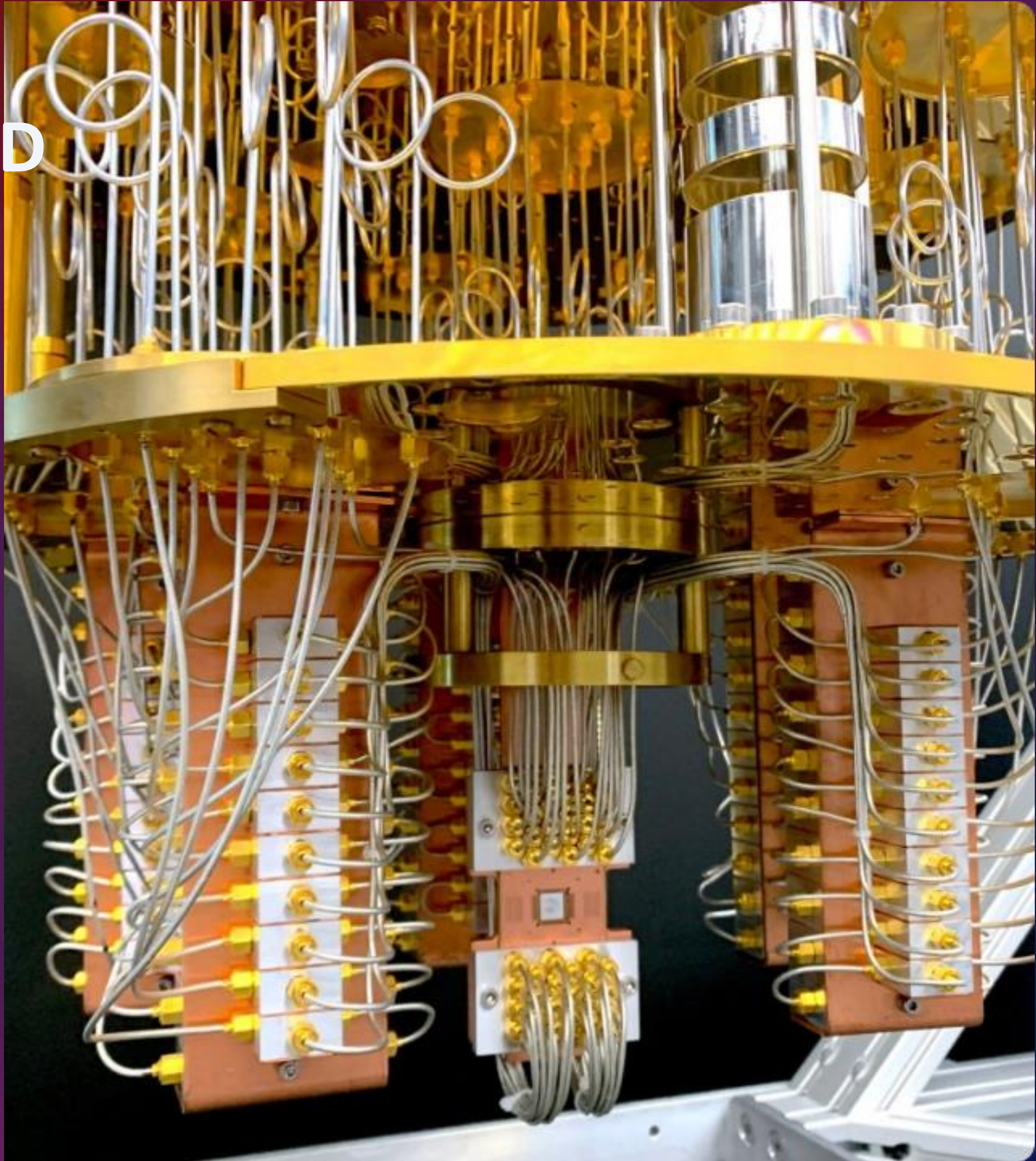
Vibration/EM Isolation

Quantum halls need electromagnetic shielding (Faraday cages) and acoustic/vibration dampening floors to prevent qubit decoherence.

Hybrid Interconnects

Ultra-low latency fiber to link Quantum Processors (QPUs) with GPU clusters for real-time error correction and hybrid algorithms.

THE CRYO-CLASSICAL HYBRID



Infrastructure for Absolute Zero

Quantum computers require a tiered cooling strategy: GPU heat rejected at 30°C, while Quantum QPUs are maintained at 0.015 Kelvin.

Today's facilities lack the specialized plumbing for liquid helium and the gas handling systems required for large-scale cryogenic cooling distribution.

" *"The data center of 2030 is not an evolution of the server room; it is a thermodynamic engineering feat that prioritizes fluid over air."*

Smarter
technology
for all

thanks.