# Green Data Centers

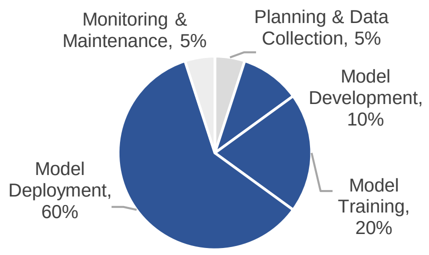
## Templewater Energy Transition & Decarbonization Insights

22 December 2025 | The 3rd Issue

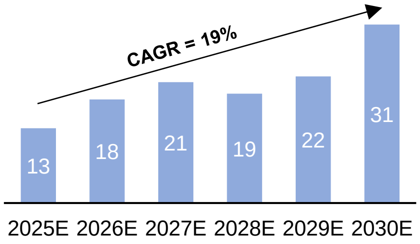
## Why are we looking at data centers?

| | |
|---|---|
| **Large Sector** | In 2024, data centers consumed 1.5% of global electricity demand, with 12% CAGR and a market size of over US$350bn. The market is expected to grow to **3% of global electricity demand** by 2030 at a **15% CAGR**. |
| **Technology Innovation** | Technology innovation emerged as data centers scaled across multiple layers of the stack, including **advanced semiconductors**, **thermal management & HVAC**, optical interconnects & photonics and power management. |
| **AI Boom** | Over **90% of AI energy demand** is within data centers; with AI-dedicated workloads projected to account for 70% of global data center capacity by 2030. McKinsey projects data center capacity cumulatively to exceed 200GW by 2030 with US$1.6tn of global spending required. |

**>90% AI Energy Demand in Data Centers**

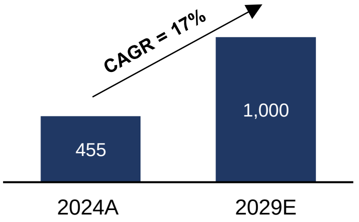



**New AI Data Center Capacity (GW)**




Source: McKinsey

**Annual Data Center CAPEX (US$bn)**




Source: Dell'Oro Group

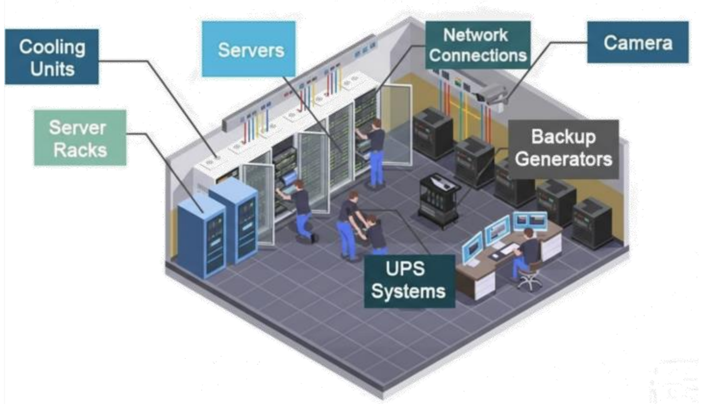
## What are data centers and why do they matter?

A data center ("DC") is a facility that **stores, processes and delivers digital data**. In the 1980s and 1990s, enterprises ran their own server rooms with on-premise storage, basic Central Processing Units ("CPUs") and limited scalability. The 2000s saw the rise of internet services and content delivery, followed by the growth of cloud and hyperscale campuses in the 2010s. Today, the **AI era is shaping data centers designed for massive computing, model training** and high-speed interconnects.

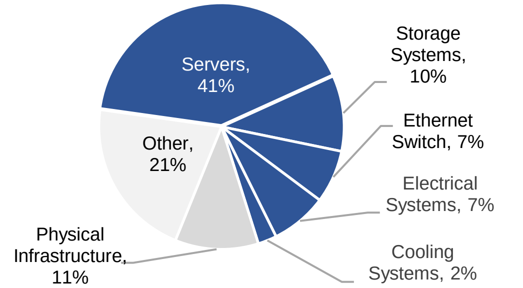
**User query in Google** → **Edge server determines optimal data center** → **Data centers search for information in the database** → **Ranking of the results in data center** → **Output transmitted to user**

| | Purpose | CAPEX | Electricity | Key Components | Key Suppliers |
|---|---|---|---|---|---|
| **IT Systems** | Core of the data center that process, store, and transmit digital data | ~60% | ~80% | **Compute:** CPU, GPU<br>**Networking:** Ethernet<br>**Storage:** SSD/HDD | Cisco, NVIDIA, AMD, Dell |
| **Non-IT Systems** | Supporting systems that enable operation of the IT systems | ~40% | ~20% | Power Infrastructure<br>Cooling Systems<br>Building infrastructure<br>Cybersecurity | Arista, Vertiv, Huawei, Schneider Electric, Western Digital, Envicool |

IT systems are the digital core of an AI data center while non-IT systems form the physical backbone. In this version of insights, we will be looking at growth areas and key market trends for both IT and non-IT systems.




**% of Data Center CAPEX**




**Innovations in Templewater's focus areas are mainly around chips and increasing PUE**

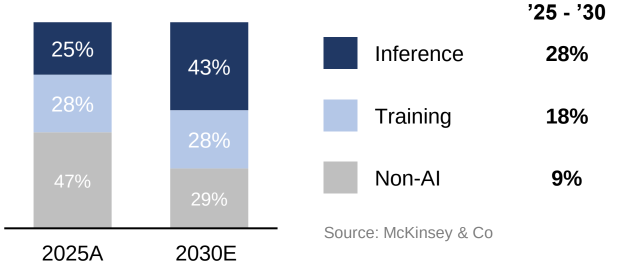
## What changes will be brought by AI?

| | | |
|---|---|---|
| **Requirement for high power ultra-dense racks** | → | Advanced AI chips and power semiconductors |
| **Increasing PUE pressure** | → | Next-gen cooling and power conversion efficiency |
| **High Reliability with Zero-Failure Uptime** | → | Battery back-up units and integrated energy storage |
| **Exponential Power Demand** | → | On-site generation and distributed power (renewables, storage, gas turbines) |
| **Decarbonization Mandates** | → | Low-carbon power solutions (gas turbines, hybrid systems) |

Data centers are typically categorized into edge, enterprise, colocation and hyperscale, as detailed below. We expect an increasing share of **colocation and hyperscale data centers** to dominate AIDC deployment, as the shift from training to inference favors **scale, capital intensity, and operational efficiency**, driving a more **consolidated market structure**. Inference workloads benefit from large, standardized platforms that can deliver reliable performance at **lower unit costs**, while supporting rapid capacity expansion and geographic distribution.

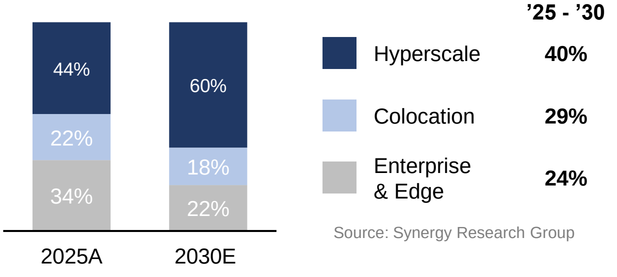
| | | | More prevalent given rise of AI | |
|---|---|---|---|---|
| **Data Center Type** | **Edge** | **Enterprise** | **Colocation** | **Hyperscale** |
| **Power Consumption** | 500kW to 2MW | 1-10MW | Tens of MW | Over 50MW, up to 650MW |
| **Operator** | Self or Third Party | Self or Third Party | Third Party | Self or Third Party |
| **Latency** | Very Low | Low | Low to Medium | Low to Medium |
| **Installed Capacity (2024)** | 3GW | 20.6GW | 8.9GW | 29.5GW |
| **Customer Examples** | AT&T, YouTube, 中国移动 China Mobile | J.P.Morgan, Goldman Sachs, Walmart | 世纪互联 VNET, DAYONE, Microsoft, amazon, Tencent, Equinix, Digital Realty, Alibaba.com | Tencent, Alibaba.com, Google, Meta, Digital Realty, Equinix |
| **Operator Examples** | Customers are generally operators. | Customers are generally operators. | Customers can be operators. | Customers can be operators. |

**% of Total Data Center Power Demand**



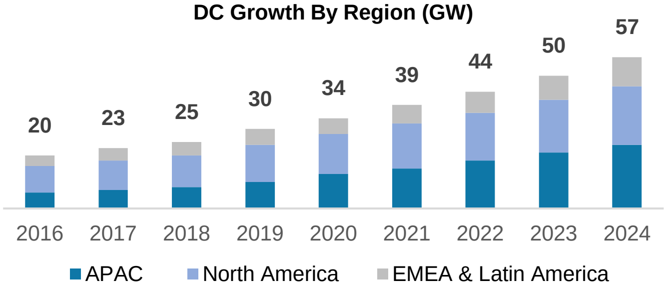

Source: McKinsey & Co

**% of Total Data Center Capacity**

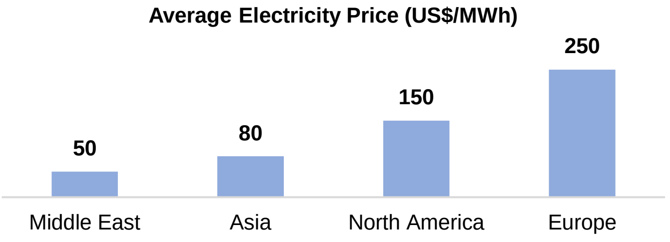



Source: Synergy Research Group

## Key Observation 1: U.S. and APAC are two key regions for growth and tech innovation




Source: Goldman Sachs




Source: World Population Review

The latest data center trends show that **North America, specifically the U.S. and APAC are the main engines of global growth** and technology innovation. Across regions, tech innovation is centered on **chip performance**, **power efficiency**, and **improving PUE** through advanced cooling, better energy management, and AI-driven operations.

APAC is the **fastest growing region for data center capacity**. China is accelerating capacity while leveraging domestic technology innovation in DC chips. Southeast Asia is also growing rapidly, with restrictions in Singapore's electricity supply pushing hyperscale development into Johor, Malaysia, where DC capacity is projected to reach **4.8GW by 2030**. Indonesia and Thailand are scaling with government support, and India is becoming one of the fastest-growing markets at **~21% CAGR**.

In EMEA, the **Middle East is emerging as a competitive hyperscale destination because it offers some of the lowest global electricity prices globally**. This cost advantage, combined with sovereign-backed mega-campuses in the UAE and Saudi Arabia, is driving rising interest from operators seeking scale and power-cost efficiency.

## Key Observation 2: Global bottleneck is on energy and power

The global bottleneck for data centers is now **power availability**, **power cost** and the ability to **scale clean energy** for AI. Different regions are taking distinct paths to secure cheaper and more reliable power.

**Oracle 1.4GW DC powered by on-site gas turbines**

**US$600m Gansu Qingyang DC Project**

**Project Stargate in UAE, a collaboration between OpenAI, SoftBank, and Oracle**

### United States

In the U.S., AI demand is outpacing grid capacity, so operators are adopting hybrid solutions that combine clean PPAs, on-site solar and batteries, **natural-gas microgrids** and **small modular reactor** projects to secure 24/7 clean energy base loads.

### China

China is addressing power constraints **by shifting hyperscale development inland to regions like Inner Mongolia, where renewable energy is cheaper** and more abundant. Current DC capacity and inference workloads can also be supported by ultra-high-voltage transmission that delivers this power to coastal demand centers.

### Middle East

The Middle East has a structural advantage with the lowest electricity prices among major markets. The UAE and GCC are pursuing off-grid and hybrid models using large solar projects, emerging green-hydrogen systems and clean backup power. Templewater's portfolio company **Marvel-Tech supports this shift with hydrogen-ready turbines**, while operators like Khazna and AirTrunk deploy advanced cooling suitable for high-density AI in hot climates.

### Technology Innovations to Reduce Power Consumption

At the frontier of innovation, several markets are experimenting with concepts that remove or reduce dependence on terrestrial grids entirely. **Space-based data centers powered by continuous orbital solar energy are being explored in the U.S.**, China and Japan as a long-term pathway to eliminate land and power constraints. **Underwater data centers** are also being tested to reduce cooling energy requirements.

## Key Observation 3: Constraints in China will drive long-term localization

U.S. export controls have created a shortage of **high-end GPUs in China**, slowing large-scale AI training. The Chinese government has further restricted access to NVIDIA chips, forcing operators to seek alternatives. Domestic firms such as Huawei and Biren are developing GPUs, but current models lag NVIDIA by one to two generations in performance and efficiency. This gap constrains near-term growth and raises questions about global competitiveness, prompting Chinese operators to focus on efficiency and software-level optimization. At the same time, industry momentum remains strong, underscored by two recent large IPOs of domestic GPU developers, Moore Threads and Muxi, which highlight sustained capital support and policy backing for local AI hardware. In the long run, China is targeting a self-sufficient AI hardware ecosystem, and given its rapid product iteration, **major tech firms and startups are expected to deliver GPUs or substitutes that can eventually compete more directly with NVIDIA.**

**Huawei**

Started 7nm shipments in 2025, 910C GPU which is equivalent to H100 by NVDIA, launched in Oct'22

**Biren Technology**

Launched BR100 chip in 2024 with 7nm; equivalent to A100 by NVIDIA, launched in May'20

**Cambricon Technologies**

Focused on ASIC chips, next launch MLU590 is expected to be closer to H100, though remains 2.5x slower

**Iluvatar CoreX**

Big Island II is the latest chip launched by Iluvatar on 7nm architecture, currently lagging behind Cambricon in speed

### Market Implication

Export controls are accelerating China's shift toward **localized AI chip manufacturing**, resulting in a **high-growth investment cycle** for Chinese AI chip manufacturers which will be bolstered by policy support and forced domestic substitution. As the domestic supply chain matures, and innovation deepens across the ecosystem, we expect more **high-quality investment opportunities** to emerge. Over time, scale-driven iteration and ecosystem build-out should narrow performance gaps and improve the competitiveness of Chinese players to expand their addressable markets.

## Key Observation 4: AI & software integration will be the next phase of development

As data centers scale in size, density, and complexity, software is becoming as critical as physical infrastructure in determining performance and economics. AI is starting to influence **how data centers are operated**, though most current applications remain narrow. The best-known examples are in cooling, where **reinforcement learning has adjusted set-points and chiller sequencing** to cut cooling energy by **10–40%**. The larger opportunity lies in managing complexity as facilities grow into hundreds of megawatts and add liquid cooling, on-site generation and more diverse workloads

Another key area of innovation comes from startups applying **artificial intelligence to data center operations**. One approach involves creating an AI-driven Digital Twin, a real-time virtual model of the facility. This allows the AI to analyze live sensor data and automatically optimize cooling and power systems for peak efficiency. A second trend is Predictive Maintenance, where AI learns the normal operating patterns of critical equipment. By detecting tiny deviations, it can forecast failures before they happen, allowing teams to fix issues proactively and prevent outages.

**phaidra**

AI-powered control systems that automatically learn, adapt, and get better over time

**VERDIGRIS**

Improve utilization individual devices in a data center using AI based analysis of power consumption data

**coolgradient**

AI software solutions designed to optimize operations and drive significant improvements in efficiency

**emeraldai**

Real-world AI power demand flexibility to relieve peak grid stress

### Market Implication

Rising data center scale and operational complexity are creating strong tailwinds for **AI-powered software**, making digital twins, predictive maintenance, and autonomous optimization platforms **crucial to drive competitive performance**. As these systems become deeply embedded in day-to-day operations, they are likely to command **increasing strategic value** for hyperscalers and colocation operators. We therefore see **attractive long-term growth opportunities** for software and integrated hardware-software providers related to data centers.

## POTENTIAL OPPORTUNITIES FOR INVESTMENT IN IT SYSTEMS

# Shift towards ASICs for training and inference in specific sectors

GPU performance has surged as transistor density increases, allowing thousands of cores to work in parallel. Modern GPUs such as NVIDIA's H100 and AMD's MI300 use 4–5 nm nodes with tens of billions of transistors, moving toward 3 nm and below. Techniques like **3D stacking and materials such as graphene, carbon nanotubes, and gallium nitride aim to sustain further gains.** Energy per compute is now the main constraint in AI data centers. To address this, ASICs designed for specific workloads reduce energy per operation by 30–50% and are increasingly used by hyperscalers.

**Moore Threads**

GPU and accelerated-computing solutions for AI, graphics, and data-center workloads

**Enflame Tech**

AI training and inference accelerator chips designed for large-scale data-center deployment

**Cerebras**

Wafer-scale AI processors enabling ultra-high-performance training for frontier-scale models

**SambaNova Systems**

Full-stack AI acceleration platform built on custom dataflow-optimized chips

## POTENTIAL OPPORTUNITIES FOR INVESTMENT IN NON-IT SYSTEMS

# Direct-to-Chip cooling expected to be the leading technology up to 2030

Cooling systems are becoming increasingly important as the rack densities increases. NVDIA plans to launch **600 kW rack in 2027, however the adaptation of it is expected to take longer** as going forward the focus would be more on the inference as compared to training. The overall market is shifting toward liquid cooling, with **direct-to-chip projected to take the lead in AI-era deployments till 2030**, post 2030 we may see increasing adaptation of immersive cooling.

Complementing this trend, microfluidics is emerging as a promising enhancement, using **microscopic channels etched directly into the chip to improve heat transfer** efficiency. Two-phase direct-to-chip cooling, which uses refrigerant boiling to remove significantly more heat at lower flow rates, is also being explored, although it may face longer commercialization timelines due to complexity. These developments indicate that while **direct-to-chip will remain the primary solution in the near term**, advanced liquid-cooling technologies are evolving quickly to support the next generation of high-density AI infrastructure.

**Yxindata**

Leading Chinese direct to chip startup with 10% market share in domestic market

**Zutacore**

Direct-to-chip, waterless liquid cooling systems for data centers

**Apheros**

Innovative metal foam to increase contact area for direct to chip technology

**Iceotope**

Immersive cooling only for servers to reduce weight of system by up to 50%

# Alternative Energy Solutions to Manage Increasing Electricity Demand

Data centers already consume 1.5% of global electricity and this is expected to rise to 5–6% by 2030. The challenge is not global supply of energy but the mismatch between demand and grid build timelines**. New connections often take 2-3 years in the U.S. and 1-2 years in Europe, making electricity the slowest input in the build cycle**. High energy costs and dependence on diesel backup further complicate operations.

Operators in the U.S. are responding with **near-site generation such as natural gas turbines to secure dedicated power** or natural gas-based fuel cells from providers like Bloom Energy to supply power during construction phase. Over the longer term, **small modular reactors (SMRs), nuclear fusion** and **space energy** are attracting billions in investment as potential baseload sources for hyperscale campuses, though meaningful deployment is unlikely before 2035 due to long development and regulatory timelines.

**GE Vernova**

DC demand has created a global shortage of gas turbines, GE Vernova lead times reaching 2029

**Marvel Tech**

Global shortages provide growing turbine developers like Marvel Tech an opportunity to expand globally

**Bloom Energy**

Natural gas-based fuel cells have gained significant interest from DCs, particularly in the U.S.

**China National Nuclear Corporation**

Linglong One is the most advanced SMR in the world which is under construction in China

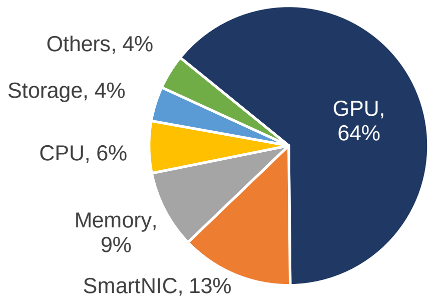
## Appendix – Introduction of DC key components

**IT SYSTEMS**

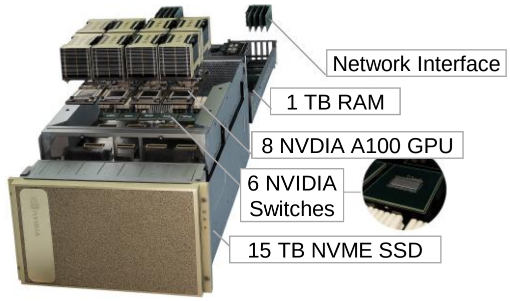
### IT Systems – Compute

Around 60% of a data center's investment goes into IT hardware, with compute at the core. **Compute refers to the processors and their supporting components** within servers. CPUs were once the backbone of data centers, handling most workloads. Today, their role is shifting toward system control, logic, and general applications. **GPUs, designed for massively parallel computing, are now the engines of AI** and high-performance workloads.

**Cost Breakdown for IT Systems**




**Components of a A100 NVDIA GPU Server**

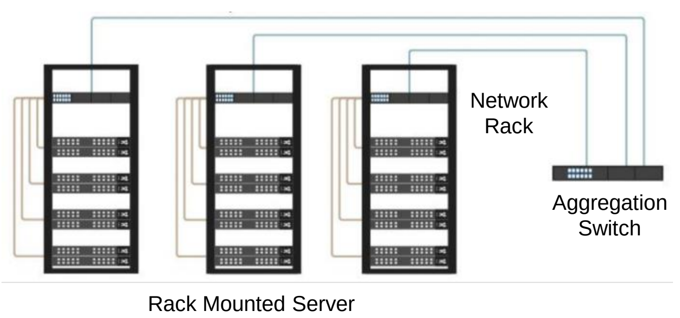



| | Tasks | Cores | Speed | Bandwidth | Power |
|---|---|---|---|---|---|
| **CPU** | Sequential general-purpose tasks | 8 to 64 cores | 1–2 TFLOPS (trillion floating-point operations per second) | ~200–300 GB/s | ~300 watts |
| **GPU** | Massively parallel computations | 1000s of small cores | 60–70 TFLOPS | ~900 GB/s or higher | ~700 watts |

### IT Systems – Networking & Storage

Networking and storage together **determine how efficiently data moves and persists in a data center**. While not the primary bottlenecks, both are becoming important levers for efficiency as AI clusters grow larger. As model training and inference scale across thousands of accelerators, the performance of interconnects increasingly affects utilization and throughput, while storage architecture shapes how quickly massive datasets can be accessed, reused, and managed.

**Networking System**




**Storage Devices**


#### IT Systems – Networking

Inside a server, **interconnects link CPUs, GPUs, and memory. Between servers, connections run over Ethernet or InfiniBand**, carried through cables and managed by switches. Ethernet is the general-purpose technology used across the internet and in most enterprise networks. It is cheaper and widely available but has higher delay. **InfiniBand is specialized for AI clusters**, providing faster response and better coordination across thousands of GPUs, which is why it is used for training the largest models.

#### IT Systems – Storage

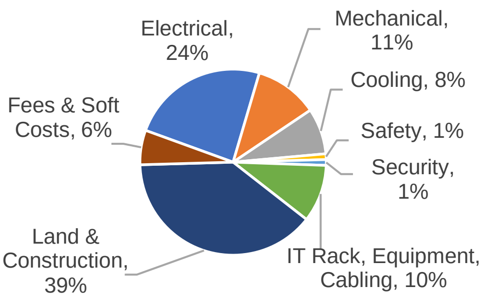
Data centers now manage billions of **gigabytes of information, stored mainly on Solid-State Drives (SSDs) and Hard Disk Drives (HDDs)**. Storage systems use about 5% of total power, and only modest improvements are expected in the underlying technology.

## Appendix – Introduction of DC key components

**NON-IT SYSTEMS**

### Non-IT Systems – Buildings, MEP & Prefabrication

**Cost Breakdown for Non-IT Systems**




Construction and Mechanical, Electrical and Plumbing (MEP) systems account for up to half of a data center's capital cost, with land being the key variable. Traditional builds take 2.5–3 years, while prefabricated modules now enable delivery in 9–12 months. **Faster builds accelerate revenue, lower financing risk, and improve IRRs**. Cybersecurity is now engineered into facilities from day one as operators harden against attacks that compromise uptime and customer trust.

*Prefabricated modular data center deployed by Vertiv in Barcelona, Spain for T-Systems*

### Non-IT Systems – Power Infrastructure

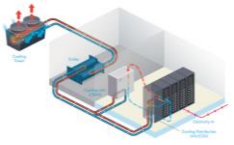
Non-IT systems consists of all the auxiliary systems required for reliable and continuous operation of data center. Power enters from the utility grid, is stepped down by transformers, stabilized by UPS batteries, and distributed to racks powering CPUs and GPUs. If the grid fails, the UPS bridges supply until generators or fuel cells take over.

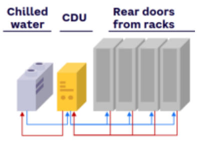
**Utility Grid** → **Transformers** → **UPS systems** → **Server Racks**

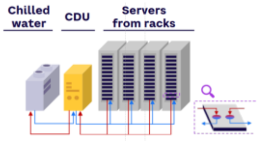
Power is the gating constraint and the biggest source of emissions. Many sites still rely on diesel backup, but operators are turning to on-site generation with gas turbines or fuel cells and contracting for 24/7 clean power to satisfy hyperscaler ESG mandates. For back up power systems, PEM fuel cells can be a potential back-up power solution for a small data centers considering the TCO.

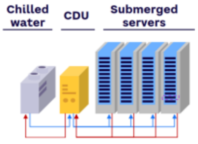
### Non-IT Systems – Cooling Systems

**Emerging Technologies**






| Particular | Air cooling | RDHx | Direct to chip | Immersive |
|---|---|---|---|---|
| **Cooling Medium** | Air | Water | Water or glycol | Dielectric liquid |
| **Pros** | Mature, easy to deploy | Effective at moderate density | Targeted cooling, scalable | Ultra-high density support |
| **Cons** | Poor efficiency at high density | Still constrained by air distribution | Complex retrofits | Hardware compatibility limitations |
| **Suitable use case** | Legacy workloads | Incremental retrofits for heat-intensive workloads | HPC, AI/ML, CPU/GPU intensive tasks | HPC, AI/ML, blockchain |
| **Adoption** | ~90% of legacy DCs | ~5% | ~3% | Still at pilot scale |

Servers convert most consumed power into heat, which must be **removed to keep chips within safe operating limits**. Data centers rely on two main cooling approaches: air and liquid. **Air cooling dominates over 90% of installations** and is suitable for lower rack densities, typically below ~25 kW per rack, using server fans and air handlers supplied by chilled water systems.

As rack densities rise, liquid cooling is gaining share. Rear-door heat exchangers (RDHx) account for roughly 5% of deployments, absorbing exhaust heat at the rack level. **Direct-to-chip liquid cooling is scaling more rapidly** as AI workloads push densities beyond 50 kW, supporting up to ~200 kW per rack and expected to become the primary solution for high-density AI environments despite representing only 3–4% of current installations. Immersion cooling enables densities above 200 kW per rack but remains niche, limited to the most extreme AI use cases.