The AI Buildout

HOW NEXT-GEN COMPUTE IS RESHAPING THE GLOBAL DATACENTER MARKET

EXPERTS WITH IMPACT™

Executive Summary

Artificial intelligence ("AI") is accelerating a structural shift in the global data center market, moving us forward into a new era of compute. The **rise of AI workloads is driving unprecedented demand for high-performance data center capacity, expected to surpass 150 GW globally by 2030**.[1] Higher densities and tighter efficiency thresholds are rapidly redefining data center design and economics.

To meet this demand, the industry is entering a cycle of outsized capital investment. AI-driven capacity expansion is expected to require over **$4 trillion in CAPEX over the next five years**, which we estimate will demand **over $1 trillion in economic value added** yearly by 2030 to satisfy investor expectations.[2] Funding for such wave exceeds hyperscaler's balance sheets, pushing the industry towards new financial models, including outsourcing and consortiums.

In that context, new players such as **GPU-as-a-Service ("GPUaaS")** or **Neoclouds** have emerged in the form of specialized platforms that deliver high performance compute at scale. They do so under two main business models; i) long-term leases to large capacity consumers in the form of bare metal and/or orchestrated clusters; ii) and "Pay As You Go" (PAYG) offerings, the traditional "Infrastructure as a Service" (IaaS) marketplace propositions billed by the hour.

At the same time, the market's geographic profile is evolving. Rising congestion in core hubs such as **Frankfurt, London, Amsterdam, Paris, and Dublin ("FLAP-D")** is redirecting investment toward new regions. **Nordic countries are emerging as preferred destinations for high-density compute** due to attractive climate conditions, affordable and abundant renewable power, competitive land availability and strong connectivity to Europe.

In a constrained supply environment, **speed-to-market and reliable RFS ("Ready-for-Service") delivery** are becoming critical differentiators. Operators that can build first are securing long-term customer commitments with limited risk of overbuild. In response, investors diligence four elements to ensure timely delivery: supply chain agreements, power availability, permitting and construction readiness. **Grid power availability remains a key struggle to delivering capacity on time, leading to alternative power sources such as on-site natural gas turbines** gaining traction as accelerators of deployment timelines.

Although AI dominates growth, **cloud and colocation remain resilient pillars in the market and are more relevant as models become complex and enterprises look to protect their data and IP locally**. Hybrid architectures are strengthening due to tighter data-sovereignty policies and sovereign cloud strategies are gaining prominence as governments seek control over national AI infrastructure. Together, these segments provide complementary growth alongside AI-driven data center expansion.

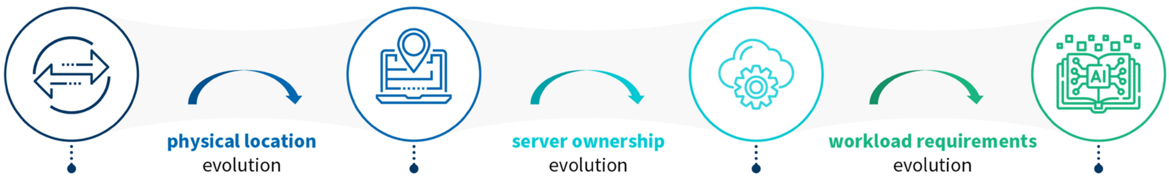
AI Triggers a Structural Break in the Datacenter Market

The global data center industry is entering a transformative phase driven by AI. Unlike previous shifts — such as off-premise migration, virtualization, and cloud — **AI does not simply change consumption models. It alters fundamental workload requirements, reshaping infrastructure design** and amplifying power and cooling demands beyond the tolerance of traditional facilities.

Prior computing waves gradually moved workloads off-premise and into scalable virtual environments. In contrast, AI introduces workloads that are both energy-intensive and, in some instances, latency-dependent. Training models require highly clustered, multi-gigawatt campuses where power density (+120 kW/rack) exceeds the thresholds feasible in cloud-oriented facilities. Inference, particularly for interactive or mission-critical applications, requires distributed deployments closer to users and cloud environments. The latter signals the emergence of new facility types, new deployment geographies and a reconfiguration of commercial models.

The market impact is structural, as cloud workloads are being enhanced rather than replaced. However, the pace and intensity of growth is led by AI, pushing data center infrastructure into an era in which density, latency and energy economics become core determinants of competitiveness. As a consequence, the buildout of AI-optimized data centers becomes one of the largest global infrastructure investment cycles in modern history.

Figure 1 – Evolution of enterprise and cloud computing



1990

MAINFRAME AND X86-BASED ON-PREM SERVERS

- Centralized on-premise servers
- Managed by internal IT teams

2000

VIRTUALIZED OFF-PREM INFRASTRUCTURE

- Enterprise servers shift off-premise DCs
- Remotely accessed and virtualized

2006 - 2023

CLOUD ERA

- Servers are no longer owned and managed by enterprises (opex for capex)
- Workloads become elastic, scalable and pay as you go
- **Hyperscalers** dominate

Now

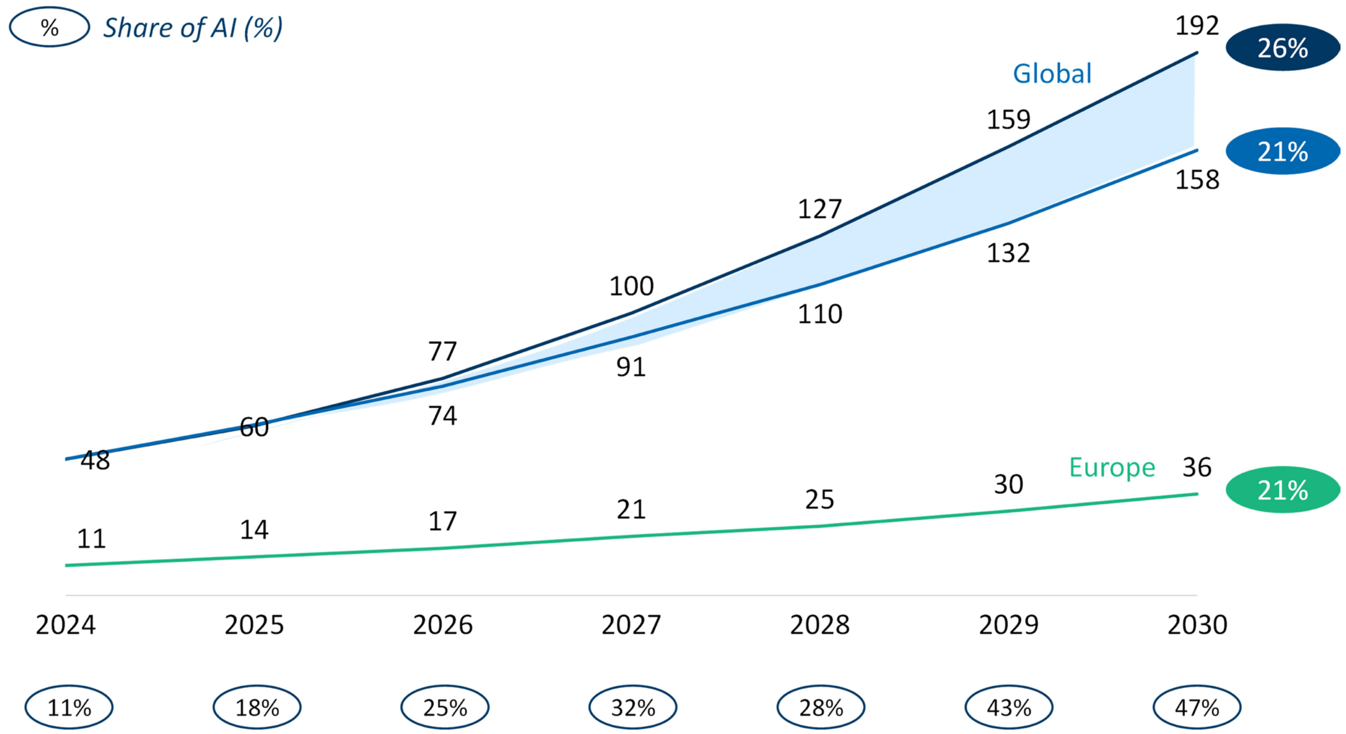
AI OPPORTUNITY

- AI democratization spreadheads global demand for DCs
- Cloud workloads being enhanced with AI

Market Trends: AI Workloads Drive a New Compute Era

Data center demand is forecast to grow at more than 20% CAGR through the end of the decade, with Europe alone expected to reach approximately 36 GW of installed capacity by 2030. Unlike historical growth, which was driven predominantly by cloud migration and enterprise off-premise adoption, future expansion is dominated by AI training and inference workloads.

Figure 2 – Datacenter global and Europe capacity projections



Source: FTI Consulting analysis

AI workloads create a **step-change in density requirements**. GPU racks supporting AI training can exceed 60–120 kW per rack — four to ten times more than conventional enterprise workloads. Inference also requires elevated density, although with more distributed placement, depending on latency constraints. In comparison, typical cloud and enterprise workloads operate at or below 15 kW per rack.

Latency requirements further differentiate facility needs. **Training workloads can tolerate delays and therefore operate in remote locations**. In contrast, **certain inference can be required to sit near the point of consumption, depending on the application**. While some use cases, including chat interactions or image/video generation, can tolerate delays of seconds to a minute, others, such as autonomous systems, surveillance and interactive AI, require sub-10 ms latency. These demands drive a **distributed topology where large, remote training campuses coexist with metro-proximate inference nodes, reshaping location strategies**.

Figure 3 – Workload requirements

	AI TRAINING	AI INFERENCE	COLOCATION / CLOUD
POWER DENSITY & TCO	GPUs 60-120 kW+ **high share of energy on costs**	GPUs 30–60 kW+ **high share of energy on costs**	CPUs 5-15 kW+ **lower share of energy on costs**
LATENCY SENSITIVITY	Offline/batch, tolerant to delay	**Depends on the type of workload**	Interactive apps, mix of batch workloads
SIZE & CONCENTRATION	Clustered, **up to GW facilities**	**Large facilities** driven by economies of scale	Organized in **cloud regions,** smaller facilities
	Highly energy-intense, concentrated workloads	**Energy-intense workloads – latency sensitive**	**Clustered, low-compute, hyperscale-driven**

The latter put pressure on data center designs which evolve from a standardized capacity product to a specialized infrastructure offering. AI-purpose sites adopt:

- Direct-to-chip liquid cooling
- High-power electrical distribution
- Clustered architectures supporting hundreds to thousands of GPUs
- Modular expansion enabling multi-GW horizontal scaling

The pace of GPU lifecycle progression further reinforces the need for facilities designed to accommodate rapid technological refresh without overhauling core mechanical and electrical infrastructure.

Figure 4 – AI flagship hardware evolution

			NVIDIA				
		Cloud	Ampere	Hopper	Blackwell		Rubin
Hardware characteristics	Release date	2012-20	2020	2022	2024	2025	2026	H2-2027
	System	1/2U server	DGX A100 8U with 8 GPUs (DGX-based systems)	DGX H100 8U with 8 GPUs (DGX-based systems)	DGX B200 10U with 8 GPUs (DGX-based systems)	GB200 NVL72 ("Oberon" rack)	Vera Rubin NVL144 ("Oberon" rack)	Rubin Ultra NVL576 ("Kyber" rack)
	GPUs per rack	-	32 (4 nodes per rack)	32 (4 nodes per rack)	32 chips[1] (4 nodes per rack)	72 chips[1]	144 GPUs	576 GPUs
	Power per rack	<20kW	**30 kW**	**40 kW**	**60 kW**	**120 kW**	**Not confirmed – higher than Blackwell**	**600 kW**
	Performance inference (FP4)	-	-	-	576 PF	1,440 PF	3.6 EF	15 EF
	Performance training (FP8)	-	20 PF	128 PF	288 PF	720 PF	1.2 EF	5 EF
DC infra	Cooling		Air cooling – containment	Air cooling – rear-door heat exchangers		Liquid cooling – direct to chip		
	Power		3 phase 440 V AC busbar					800 V DC

Source: NVIDIA datasheets and GTC (GPU Technology Conference)

Business Model Trends: Outsourcing and the Rise of GPUaaS

We estimate **meeting projected demand requires over $4 trillion in cumulative capital expenditure between 2025 and 2030**, spanning both chip procurement and data center infrastructure.[3] **By 2026–2027, annual infrastructure funding needs alone surpass the combined CAPEX spending of Amazon, Microsoft, Google, Meta and Oracle**.[4] This marks a structural limit wherein hyperscalers cannot absorb both chip acquisition and infrastructure expansion internally at the required pace.

The imbalance between demand and hyperscaler capital catalyzes a **shift toward outsourcing at both the chip and data center infrastructure layers**. This has produced a new class of compute providers — Neoclouds, often offering GPUaaS. These companies provide on-demand access to high-performance compute capacity without requiring customers to own GPUs or build data centers. We distinguish between four archetypes of neoclouds, depending upon the size and nature of their proposition.

Figure 5 – CapEx Projections (USD bn)

Required CapEx to reach forecast AI DC capacity vs. Hyperscalers' annual CapEx

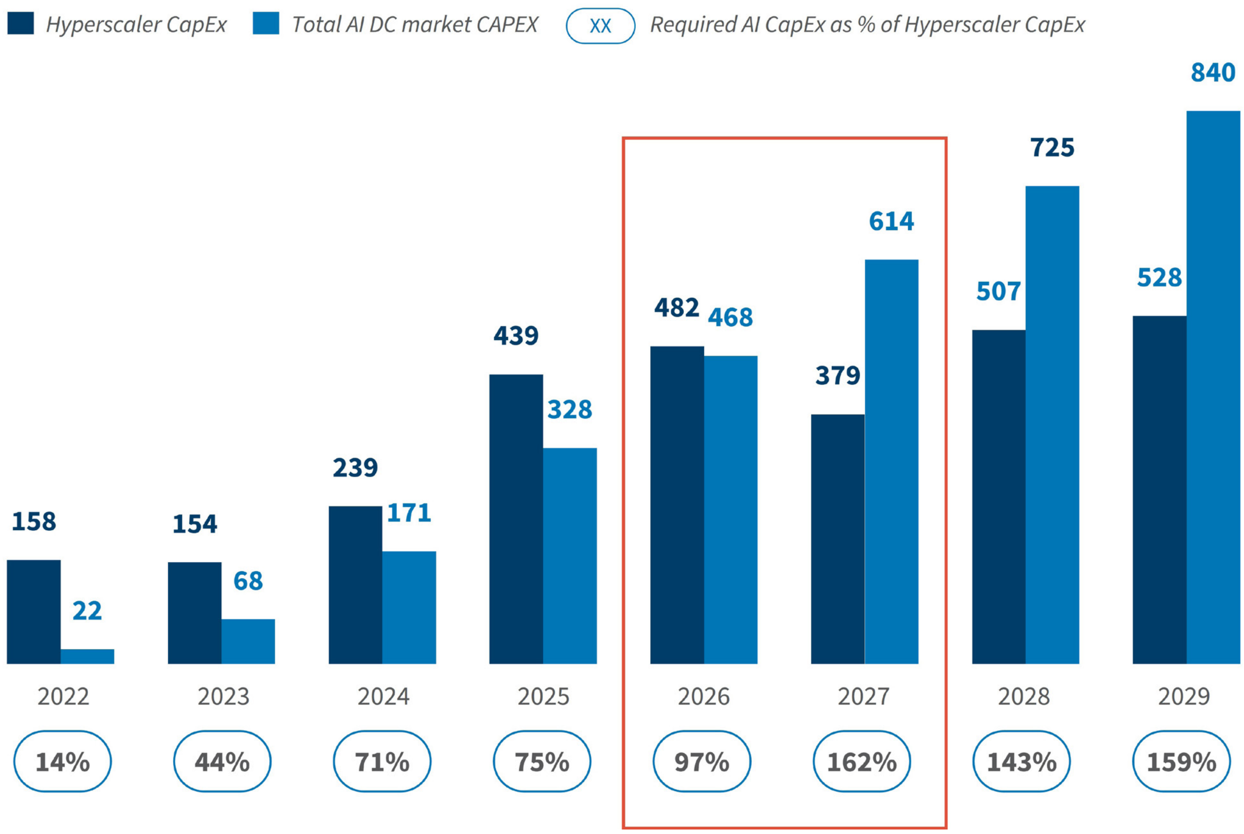


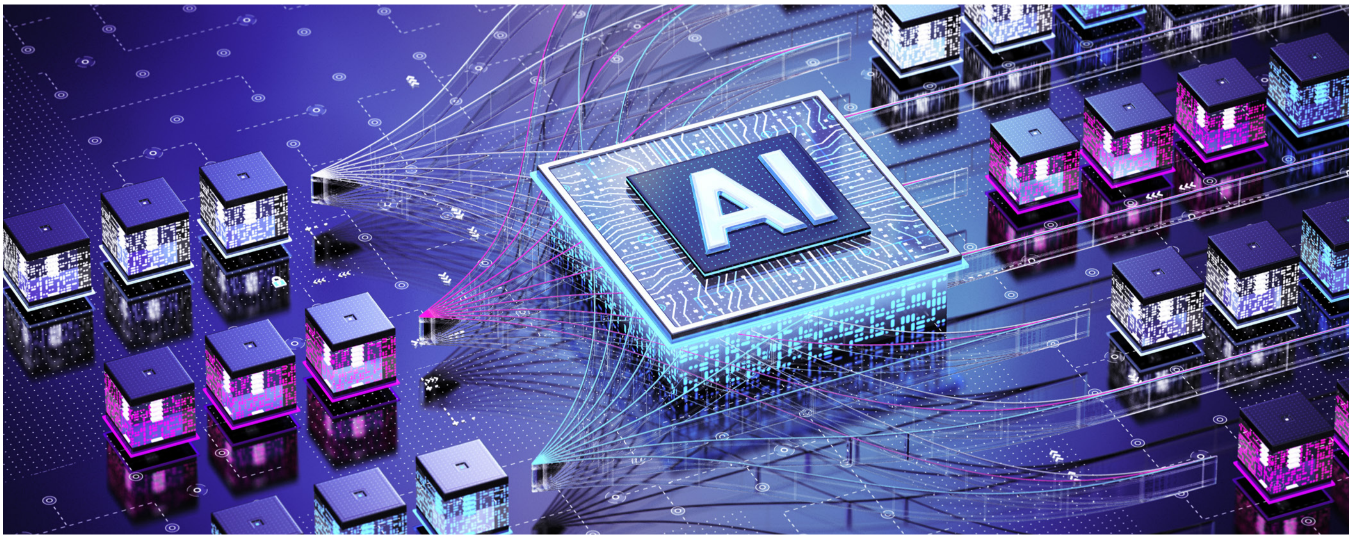
Figure 6 – Neocloud archetypes

Neocloud Giants	Emerging Neocloud	Sovereign AI Clouds	Legacy B2B Cloud Services
CoreWeave	Hyperstack	G42 Cloud	DigitalOcean
Crusoe	RunPod	Reliance Cloud Services	Linode
Lambda	Genesis Cloud	Singtel	Scaleway
Nebius	Cirrascale	FPT Smart Cloud	OVHcloud
	FluidStack	Taiga Cloud	
	Vast	NHN Cloud	
	Salad	Humain	
	Nscale	Jio	
	Akash	Mistrail AI	
AI-focused infrastructure clouds ultimately contracted by hyperscalers	*New entrants, often regional – can operate as marketplaces*	*National cloud providers – with local data residency, compliance, and domestic access*	*Enhanced offering with GPU services by hosting / enterprise clouds*

Neocloud unique selling points revolve around the **ability to source, finance and rapidly deploy large-scale clusters of specialized AI GPU hardware at unprecedented speed**. If we compare them to traditional hyperscalers, they bring:

- **Flexible commercial models:**
  - Long-term leases to large capacity consumers in the form of bare metal and/or orchestrated clusters.
  - PAYG offerings, such as traditional IaaS marketplace propositions billed by the hour.
- **Faster provisioning:** given their agile procurement approaches and streamlined infrastructure design requirements.
- **More affordable offering:** as they operate a **simplified software stack** and are **specialized in a narrow universe of hardware.**
- **Wider reach:** Locations that extend beyond hyperscalers' reach, eventually catering to sovereign requirements.

Long-term leases represent approximately 70–80% of capacity today, and have been seen by some industry players as a **way for large consumers to shift cost of the AI infrastructure push off-balance sheet**, shielding themselves from an eventual obsolescence.[5]



Location Trends: New Geographies Emerge, Led by the Nordics

Traditional data center hubs in FLAP-D have long served as the backbone of European cloud infrastructure. However, accelerated AI demand exposes structural constraints that limit their suitability for high-density deployments, including restricted development zoning, increased scrutiny on environmental permitting or power grid congestion.

At the same time, the Nordic region offers a rare combination of economic, environmental and regulatory advantages that uniquely fits AI workload demands.

Key differentiators include:

- **Renewable Energy Availability:** Over 60% renewable mix, with reliable baseload from hydro and geothermal sources.[6]
- **Lower Power Pricing:** Lower wholesale energy rates compared to primary markets in continental Europe.
- **Cool climate Benefits:** Naturally cold temperatures, which improve Power Usage Effectiveness ("PUE"), reducing total cost of ownership.
- **Land Availability:** Large contiguous plots at significantly lower prices, with agricultural land alone four times less than in FLAP-D hubs.
- **Connectivity:** Multiple subsea cables link Nordic hubs to key European metros (c.20 ms latency), allowing to AI training and non-latency sensitive AI inference workloads.
- **Regulatory Alignment:** Full compliance with EU data protection frameworks provides secure hosting for regulated workloads and sovereign compute.

As a result, the Nordics are positioned as both a cost-efficient alternative and a strategic location for next-generation compute clusters and sovereign infrastructure.

The willingness to AI train and fine tune models globally, also supports the rise of emerging data center geographies outside Europe. Investments continue to

expand in mature and developing markets where land and power can scale rapidly, such as across Mexico, Chile, India, Malaysia, Portugal and the Middle East , where multi-billion-dollar investments are underway to host cloud and AI-focused regions.[7]

These deployments are not substitutes for European hubs but are complementary additions driven by latency tolerance in AI training, enabling global capacity diversification to bypass regulatory, cooling and power limitations in saturated hubs.

Delivery Trends: Speed-to-Market as the New Competitive Moat

In an environment characterized by constrained supply rather than oversupply risk, operators compete on the ability to deliver capacity on time. **RFS deadlines have become a primary commercial differentiator**. Operators that establish presence early in new geographies develop competitive moats as hyperscalers and GPUaaS platforms tend to scale within existing partnerships rather than add new providers to their suppliers mix, if they can do so.

This dynamic creates two phases:

- **Short-term (in the next 2-3 years):** Limited competitive pressure exists for new entrants to secure customer commitments due to market tightness.
- **Mid-/Long-term:** Providers with delivered capacity, strong track record and available land-power pipeline two years down the line are positioned to secure multi-phase deployments.

For new facilities, securing long-term contracts requires demonstrating a credible path to power, permits and construction delivery, rather than available capacity alone.

Investors and customers increasingly evaluate projects based on four core levers:

Figure 7 – Key questions to stress test RFS delivery timelines

Lever	Key Questions
Customer Contracts	What proportion of pipeline capacity is contractually backed? Under which RFS terms?
Planning & Permitting	Does zoning allow data center construction? Are environmental and building permits secured?
Power	Is power contractually secured? Can the utility substation infrastructure deliver the power? Are MV/HV connections built? Are onsite generation solutions feasible?
Construction Strategy	What delivery model will be followed and what strategy is in place to secure long lead items?

Among these, **power availability is the most significant constraint**, particularly in congested core markets. To bypass grid delays, operators are increasingly exploring complementary or transitional power sources:

- **Natural-gas turbines:** Mature technology enabling rapid deployment, but utilize a non-renewable energy source.
- **Nuclear Small Modular Reactors ("SMRs"):** Long-term possibility, enabling baseload close to facilities, though commercial deployments remain beyond 2030.
- **Battery-Renewable Hybrids:** Energy storage combined with renewable "Power Purchase Agreements" (PPAs) to stabilize load and reduce utility dependence.

In parallel, modular data center construction is gaining prominence for AI/HPC ("High-Performance Computing") environments, delivering 10–15% cost savings, reduced on-site construction risk, predictable performance upon installation and faster scalability through phased deployments.[8]

This enables operators to deploy capacity faster with lower execution risk, supporting multi-stage expansion aligned with customer commitments.

Cloud and Colocation Remain Growth Pillars

Despite AI's rapid adoption, traditional retail colocation continues to play a critical role, particularly for industries requiring:

- Strict compliance and auditability
- Physical infrastructure control
- Ultra-low latency proximity to users
- Hardware customization and interoperability

Sectors such as **finance, healthcare, public services, manufacturing and gaming maintain a strong preference for colocation** because public cloud cannot fully address regulatory, latency or customization needs.

Enterprises increasingly adopt hybrid architectures, combining public cloud elasticity for unpredictable workloads, colocation-based infrastructure for predictable, stable loads and on-premise or sovereign cloud for sensitive or regulated data.

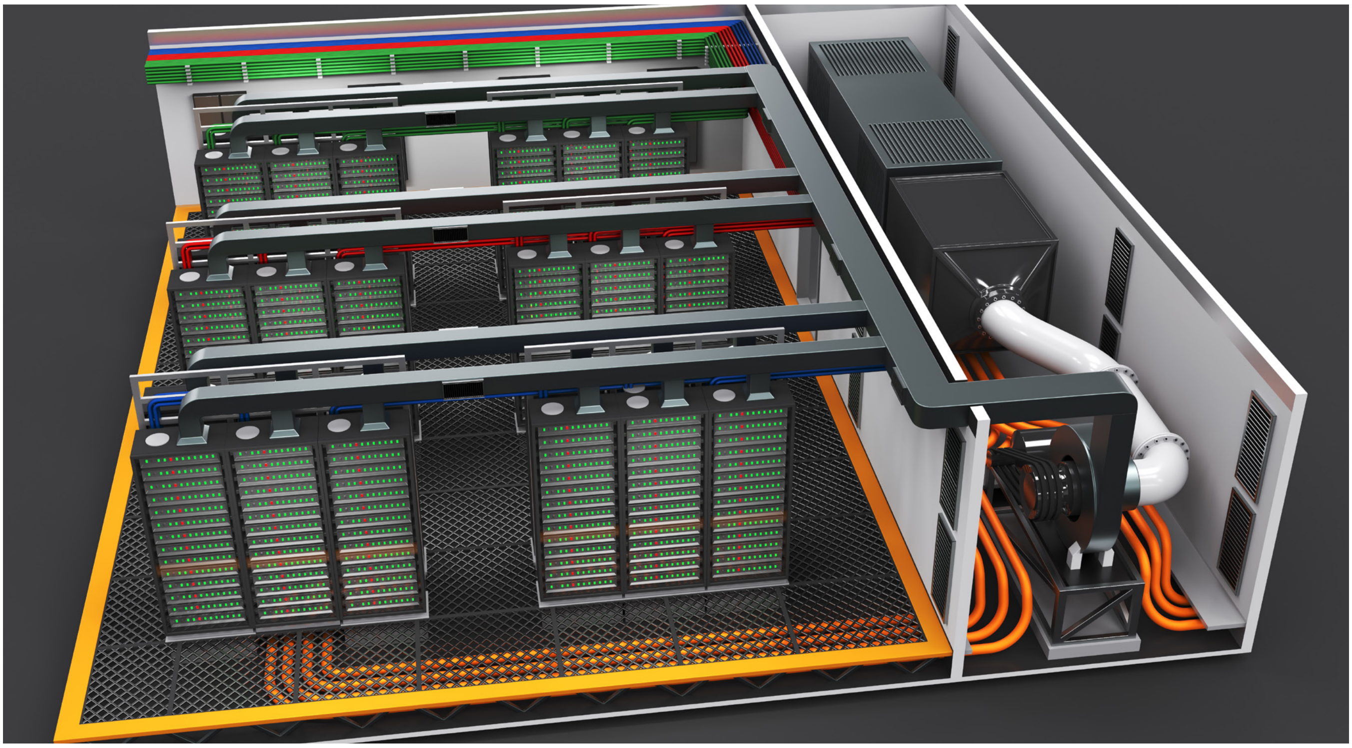
This approach balances cost efficiency, latency needs and regulatory compliance. Where workload patterns are predictable, colocation can be more cost-effective over multi-year horizons than reserved public cloud capacity. Hybrid models therefore reflect economic and operational optimization rather than a reversion from cloud adoption.

AI productivity gains are disproportionally dependent on large public cloud models that are currently incompatible with data sovereignty mandates. This creates an adoption gap for sensitive industries unable to legally or operationally use public cloud AI platforms.

Sovereign clouds solve this by providing:

- Localized data residency and encryption controls
- Compliance-aligned infrastructure
- Dedicated operational environments for government workloads
- Industry-specific AI toolkits hosted domestically
- Controlled off-boarding and workload portability

These platforms serve as the mechanism that allows strategic industries to adopt advanced AI without sacrificing regulatory or sovereignty requirements.



Conclusion – The Strategic Imperative

AI is not a cyclical uplift for the datacenter industry; it is a **structural reset that redefines how, where and by whom compute is built and financed**. The convergence of extreme density requirements, unprecedented capital intensity and constrained power availability is fragmenting the market into specialized facility types, new geographies and alternative ownership models. Hyperscalers alone cannot fund or deliver the required scale, accelerating the rise of neoclouds, outsourcing and long-term capacity partnerships. At the same time, location strategy is shifting decisively toward power-advantaged regions such as the Nordics, while speed-to-market and credible RFS delivery have become the primary sources of competitive advantage. **For investors and operators, value creation will hinge less on headline capacity growth and more on execution certainty, access to power and the ability to align infrastructure design with specific AI workloads**. Those that secure land, power and delivery capability early will shape the next decade of global compute; those that do not risk being structurally locked out of the AI value chain.

The views expressed herein are those of the author(s) and not necessarily the views of FTI Consulting, Inc., its management, its subsidiaries, its affiliates, or its other professionals. FTI Consulting, Inc., including its subsidiaries and affiliates, is a consulting firm and is not a certified public accounting firm or a law firm.

Endnotes

[1] FTI Consulting analysis using Kimberly Steele, "Global data center sector to nearly double to 200GW amid AI infrastructure boom," JLL (January 6, 2026), https://www.jll.com/en-in/newsroom/global-data-center-sector-to-nearly-double-to-200gw-amid-ai-infrastructure-boom

[2] FTI Consulting proprietary analysis

[3] FTI Consulting proprietary analysis using Turner and Townsend for cost benchmarks at the DC layer and HSBC for chip layer costs

[4] FTI Consulting proprietary analysis using Capital IQ Consensus Estimates

[5] FTI Consulting analysis using "Tech groups shift $120bn of AI data centre debt off balance sheets," Financial Times (December 24, 2025), https://www.ft.com/content/0ae9d6cd-6b94-4e22-a559-f047734bef83

[6] FTI Consulting analysis using "The Nordics: Two-thirds renewable," Nordic Energy Research (July 5, 2012), https://www.nordicenergy.org/figure/two-thirds-renewable/

[7] Press clippings, FTI Consulting analysis using:

- FTI Consulting analysis using "The future approach: Modular data centres," https://www.reuters.com/technology/brookfield-qatars-qai-form-20-billion-jv-ai-infrastructure-2025-12-09/
- "CloudHQ will invest $4.8 billion in the construction of data centers in Mexico," DCD (September 26, 2025); https://www.datacenterdynamics.com/es/noticias/cloudhq-invertir%C3%A1-4800-millones-de-d%C3%B3lares-construcci%C3%B3n-centros-de-datos-m%C3%A9xico/
- "Microsoft and Amazon's multibillion-dollar bets on India," Financial Times (December 11, 2025) ; https://www.ft.com/content/d564ba34-7514-43ca-90db-b4fa01737a10

[8] FTI Consulting analysis using "The future approach: Modular data centres," https://www.se.com/uk/en/work/solutions/data-centers-and-networks/modular-data-center/

AUDI POUS
Senior Managing Director
+34 696 050 980
audi.pous@fticonsulting.com

ARYA MOHANTY
Principal
+44 79 7410 8143
arya.mohanty@fticonsulting.com

POL CASTELLA
Senior Director
+34 637 945 856
pol.castella@fticonsulting.com

EXPERTS WITH IMPACT™

FTI Consulting is the leading global expert firm for organisations facing crisis and transformation, with more than 7,900 employees in 33 countries and territories. FTI Consulting is dedicated to helping organisations manage change, mitigate risk and resolve disputes: financial, legal, operational, political & regulatory, reputational and transactional. FTI Consulting professionals, located in all major business centres throughout the world, work closely with clients to anticipate, illuminate and overcome complex business challenges and opportunities.

FTI CONSULTING™