

Structural Demand and the Rise of Scalable Data Centre Platforms in the U.S

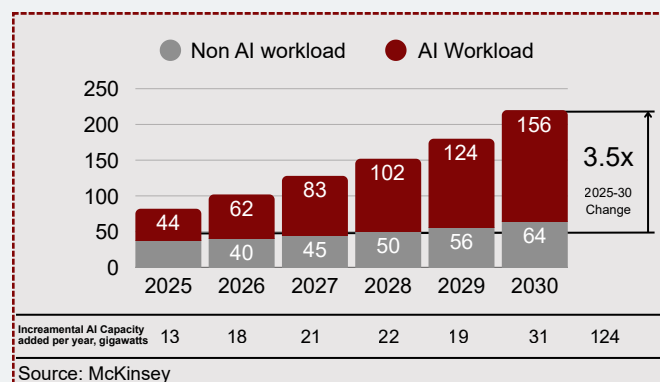
Despite ongoing debate around an “AI bubble” and short-term shifts in investor sentiment, U.S. data-centre demand continues to expand as a structural requirement of the digital economy. Generative AI is accelerating this need, but the underlying drivers, cloud migration, enterprise digitization, and rising compute and storage intensity, are sustaining long-duration demand for capacity beyond any near-term market cycle.

Market Context

Technology transitions often follow a repeatable pattern: early optimism accelerates capital deployment, near-term utilization sometimes lags, markets reprice risk, yet the underlying infrastructure frequently endures and becomes the platform for the next decade of growth. In the telecom-internet boom around the turn of the century, firms invested more than \$500 billion building fibre networks, switches, and capacity ahead of realized demand. When the cycle turned, the industry faced a period of overcapacity; Brookings observed that no more than ~2% of long-distance capacity in North America was being used at the time; an emblem of supply outrunning short-term demand. But the medium-term outcome was not “wasted infrastructure” it was lower connectivity costs and a deeper digital backbone that supported the subsequent waves of broadband expansion, enterprise digitization, and cloud, consistent with the OECD’s description of a boom-bust cycle followed by restructuring rather than disappearance. (OECD, Brookings) Building on past technology cycles, one conservative takeaway is that valuations and funding conditions can swing sharply, while the enabling infrastructure often remains useful and is progressively absorbed as adoption broadens; much as telecom and internet buildouts ultimately supported later waves of digitization after the dot-com era.

What Is Driving Data-Centre Demand Today

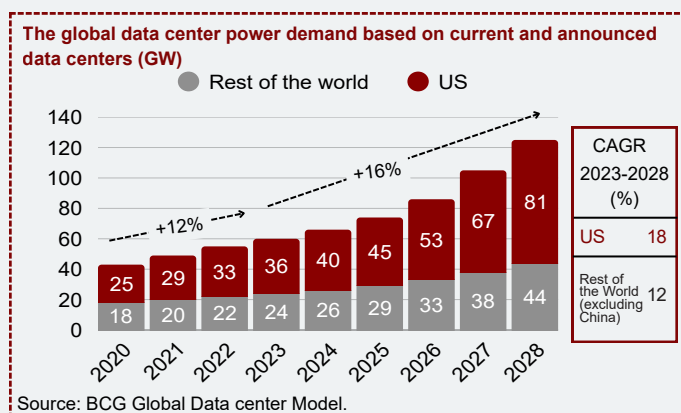
Data-centre demand continues to rise despite ongoing debate around an “AI bubble,” reflecting a broader and more durable set of underlying drivers. While AI workloads are accelerating capacity requirements, non-AI workloads; cloud migration, enterprise digitisation,



content delivery, and data growth, remain a substantial and growing component of total demand. As illustrated above, global data-centre capacity demand is projected to increase materially through 2030, with both AI and non-AI workloads contributing meaningfully to growth rather than demand being concentrated in a single use case. This expansion is increasingly anchored by hyperscalers and large cloud platforms, which are expected to account for approximately 60% of incremental global demand growth between 2023 and 2028, increasing their share of total capacity usage to around 45% by 2028 (BCG). Importantly, demand is translating into committed utilisation rather than speculative build-out, with over 90% of hyperscale capacity pre-leased on long-duration contracts prior to construction (Columbia Business School). U.S. market indicators reinforce this pattern: vacancy across primary hubs has fallen to historic lows of ~1.9%, and in markets such as Dallas–Fort Worth, the vast majority of new capacity is pre-leased before delivery (CBRE). Taken together, these dynamics indicate that data-centre demand is structural, multi-driver, and increasingly secured through long-term commitments, with AI accelerating growth but not solely defining it

The United States as the Centre of Gravity

Global demand for data-centre electricity is projected to expand at approximately ~16% CAGR between 2023 and 2028, reaching ~125–130 GW by 2028, reflecting a clear acceleration relative to the early-2020s baseline. Within this expansion, the United States is expected to remain the single largest data-centre market, with power demand growing at an estimated ~18% CAGR over the same period,



materially outpacing the rest of the world excluding China, indicating that while data-centre demand is increasing globally, the United States remains the dominant and fastest-growing market, driven by concentrated cloud adoption, AI deployment, and large-scale enterprise workloads.

How Large Platforms Are Translating Demand into Capacity

Rising demand is being reinforced by disclosed capital allocation and long-duration capacity commitments from large cloud, AI, and compute platforms, translating demand expectations into secured infrastructure usage. These commitments demonstrate that capacity is increasingly being locked in well ahead of delivery by creditworthy counterparties, reducing reliance on speculative leasing and reinforcing the durability of demand supporting new data-centre development. These dynamics indicate a broad and creditworthy tenant pool, from hyperscale/AI platforms seeking anchor capacity to enterprises procuring colocation, supported by low vacancy and forward capacity procurement. This indicates that demand is active and increasingly structured through long-duration capacity arrangements.

ORACLE

Oracle: ~\$248B in long-term lease commitments related to data centres and cloud capacity, with 15–19-year tenors commencing between FY2026–FY2028 (CoStar)

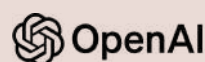
Alphabet (Google Cloud): ~\$75B of 2025 capex tied largely to AI and data-centre expansion, plus \$40B committed in Texas through 2027 for cloud and AI infrastructure (Reuters)



Microsoft (Azure): ~\$80B of planned FY2025 investment to expand AI-enabled data-centre capacity (Reuters)



Amazon (AWS): ~\$125B of expected 2025 capex, with spending described as “largely on AI projects” (Reuters)



OpenAI: Reported ~\$300B multi-year computing power agreement with Oracle, alongside its “Stargate” initiative targeting \$500B of AI infrastructure and 10 GW of capacity (Reuters; OpenAI)



NVIDIA: Announced plans to deploy at least 10 GW of AI systems for next-generation infrastructure, with potential investment of up to \$100B as capacity is rolled out (NVIDIA Newsroom)

As demand scales and capacity becomes increasingly committed in advance, outcomes for new data-centre developments are shaped less by demand uncertainty and more by how effectively location, power access, design choices, and operating constraints are managed. In this context, understanding the factors that influence performance, scalability, and resilience becomes central to successful execution.

Key Factors That Strengthen Data-Centre Performance and Scalability



Power availability and time-to-power:

In the U.S. data-centre market, power access is now the main determinant of delivery timelines, so the best-positioned developers treat “time-to-power” as a core planning advantage rather than a constraint. While some market outlooks flag a widening gap between projected data-centre demand and utility supply plans, this has triggered accelerated grid investment and more structured interconnection approaches in priority regions,

improving visibility on how capacity will be delivered and when. The commercial upside is clear: projects with a credible, well-sequenced power pathway can move faster, secure anchor tenants earlier, and reduce commissioning risk relative to peers.



Land, location, and network proximity:

Site selection remains a performance and monetization lever because proximity to dense fibre routes, cloud on-ramps, and large enterprise demand pools improves latency, interconnection choice, and commercial attractiveness. Established hubs (e.g., Northern Virginia and other primary markets) continue to demonstrate the “ecosystem flywheel,” where connectivity attracts customers and customers attract more connectivity, supporting durable absorption even as broader market sentiment shifts. This creates a practical advantage for developments that prioritize network adjacency and campus scalability from day one.



Design philosophy and tiering strategy:

Strong delivery platforms use design as a value tool, matching resiliency to workload needs rather than applying a single expensive standard everywhere. A fit-for-purpose approach (mixing Tier I/II zones for less critical loads with Tier III zones for mission-critical workloads) expands the addressable tenant base while preserving reliability where it is priced and contractually required. The benefit is faster scaling and capital efficiency; the key to capturing that benefit is disciplined governance and clear product definition, so customers understand what they are buying and operators avoid complexity creep.



Energy and thermal readiness for AI workloads:

As AI increases compute density, energy strategy and thermal design increasingly determine what workloads a facility can host and at what economics. The market is already reflecting this shift through rapid growth in grid power requirements for data centres, reinforcing why “energy-ready” design (efficient electrical topology, upgrade-ready cooling, and density-flexible halls) improves both performance and revenue optionality (S&P Global). In practical terms, developers that plan for a mixed portfolio (enterprise + AI) can protect utilization by accommodating higher-density deployments without disruptive retrofits, while also improving operating efficiency and scalability.



Regulatory, environmental, and community context:

In major hubs, permitting and stakeholder alignment can be a differentiator, well-managed projects treat these as solvable execution workstreams rather than external surprises. As local authorities refine zoning and review processes in response to development intensity, early engagement, design mitigation (noise/visual/traffic), and credible sustainability positioning can shorten timelines and improve “license to operate,” especially for multi-phase campuses.



Operational flexibility and scalability:

Operational flexibility also supports commercial durability: the ability to absorb event-driven surges (elections, major live events, peak commerce periods) without service degradation helps retain enterprise customers while meeting hyperscale expectations for consistency and scale. The opportunity is clear differentiation through reliability; the requirement is standardized procedures, strong governance, and deep operating talent.

Bringing Market Demand and Execution Discipline Together

As the U.S. data-centre market continues to scale, the interaction between demand dynamics and execution discipline becomes increasingly important. Structural demand, anchored by utility-scale users and long-duration commitments, provides the foundation for capacity expansion; however, the consistency and resilience of outcomes are ultimately shaped by how effectively projects navigate power availability, site selection, design choices, regulatory context, and operational execution. In this environment, data-centre developments that align market demand with disciplined planning and delivery are better positioned to translate growth into stable, repeatable performance. The convergence of strong demand signals and well-managed execution parameters increasingly defines the difference between projects that merely participate in the market and those that achieve durable operating outcomes.

Conclusion

The evidence indicates that U.S. data-centre demand is increasing on a structural basis, driven by cloud computing, AI workloads, and enterprise digitization, and anchored by utility-scale users with long-duration contracting behaviour. As capacity becomes increasingly committed in advance, outcomes for new developments depend less on demand uncertainty and more on disciplined execution across power availability, location and network proximity, fit-for-purpose design and tiering, thermal readiness, regulatory context, and operational scalability. Taken together, these dynamics support the conclusion that data-centre development in the United States can be positioned as a commercially reliable infrastructure business when demand strength is matched with deliberate planning, delivery discipline, and operating model maturity.


About Us

With over three decades of management consulting experience across telecom, cloud, and digital infrastructure, VTT Global advises governments, investors, and large enterprises on data-centre strategy and transformation. Drawing on hands-on delivery across end-to-end market strategy, commercial feasibility, and infrastructure planning in South Asia, the Middle East, and the United States, we bring a practical, execution-aware perspective to data-centre development and investment decisions.

Why VTT Global?

- Global footprint with local expertise
- 200+ subject matter experts
- 260+ professionals worldwide

Our Focus Industries

-  Technology, Media and Telecom (TMT)
-  Energy, Climate & Sustainability
-  Public Policy and Governance