ALERT MEMORANDUM

Financing the Data Center Boom: The Investment Opportunity, the Risks, and the Capital Solutions Taking Shape

September 9, 2026

This is the first in a series of articles examining the rapidly evolving landscape of data center financing. Future installments will take a closer look at specific capital solutions — including senior secured bonds, structured joint ventures, project finance, and other private credit structures — that are being deployed to meet the sector's extraordinary demand for capital. Today's article provides an overview of the data center financing landscape, the capital requirements needed to launch and operate data centers and accompanying power infrastructure, and the risks that sponsors should be thinking about when structuring capital solutions for data centers.

1. A Trillion-Dollar Question

The global buildout of AI data center infrastructure is generating an unprecedented wave of capital formation — what some have called the biggest peacetime investment project in human history, financed largely using off-balance sheet structures (including project financing, asset-backed financing, and structured JVs). Off-balance-sheet structures allow hyperscalers to fund this buildout without inflating their leverage ratios or impairing the balance-sheet metrics their credit ratings depend on.

If you have any questions concerning this memorandum, please reach out to your regular firm contact or the following authors

NEW YORK

Niqui Kohli
+1 212 225 2070
nkohli@cgsh.com

Yulia A. Solomakhina
+1 212 225 2848
ysolomakhina@cgsh.com

Albert Alsaygh
+1 212 225 2095
aalsaygh@cgsh.com

Brett A. Pearlman
+1 212 225 2649
bpearlman@cgsh.com

Dean Luo
+1 212 225 2077
dluo@cgsh.com

The global buildout of AI data center infrastructure is generating an unprecedented wave of capital formation — what some have called the biggest peacetime investment project in human history, financed largely using off-balance sheet structures (including project financing, asset-backed financing, and structured JVs). Off-balance-sheet structures allow hyperscalers to fund this buildout without inflating their leverage ratios or impairing the balance-sheet metrics their credit ratings depend on.

There are a lot of numbers flying around – all of them staggering. According to Goldman Sachs, capital spending by AI hyperscalers will exceed $525 billion in 2026. That figure approaches the $540 billion annual spend that the International Energy Agency says is needed to maintain current oil and gas output. McKinsey projects that nearly $7 trillion in capital spending on data centers will be required by 2030 to meet demand. Morgan Stanley forecasts $2.9 trillion of capex for global data centers through 2028, spanning hardware, real estate, power infrastructure, and build-out costs, with a $1.5 trillion financing gap — after accounting for hyperscaler cashflows — that needs to be met by external capital. Goldman Sachs analysts expect a combined $5.3 trillion in capital spending from 2025 through 2030. This estimate continues to climb: it stood at $4.5 trillion prior to the 2026 first-quarter earnings reports.

Five years ago, a single data center financing package of $30 billion would have been unprecedented. Today it is just one of many similar mega-construction projects announced over the past twelve months. On an annual basis, per Morgan Staley's estimate, financing needs will increase to just over $900 billion by 2028. For context, all companies in the S&P 500 combined spent approximately $950 billion on capex in 2024.

The challenge is stark and the opportunity is immense: traditional financing instruments alone cannot absorb this volume. For developers, investors, lenders, [power producers,] and hyperscale operators alike, understanding the current financing landscape and utilizing innovative capital structures is essential.

2. What Are Data Centers and Why Are They So Capital-Intensive?

In simple terms, data centers are the physical facilities that house the servers, networking equipment, and cooling infrastructure that power everything from cloud computing to generative AI. Data centers represent a new asset class that is expected to triple in size by 2034, from $256 billion to about $776 billion. The number of hyperscale data centers since 2018 has already tripled to 1,297 facilities. Data centers today also house greater capacity: advances in AI have precipitated a steep change from megawatt-level projects to gigawatt-scale developments.

Data center development is extraordinarily capital-intensive: a single large-scale campus requires billions of dollars in upfront investment for hardware, power infrastructure, fiber connectivity, cooling systems, land and the physical structures themselves. The cost of constructing a data center suitable for AI workload ranges from approximately $8 million to $14 million per megawatt for the shell alone, and $21 million to $31 million per megawatt once GPUs and servers are included. A critical feature distinguishing data centers from other asset classes is their unique blend of real estate (land, buildings, leases), infrastructure (power, cooling, connectivity), and technology (servers, GPUs, networking equipment). The physical and electrical infrastructure typically accounts for 30% to 50% of total data center capital costs, making the risks associated with this layer of primary importance in any data center financing, whether the project is a hyperscaler facility or a colocation operation.

3. Hyperscale vs. Colocation: A Crucial Distinction

Not all data centers are alike, and the distinction matters enormously when it comes to financing.

Hyperscalers — the major cloud platforms — operate vast facilities primarily for their own workloads. Colocation companies, by contrast, lease space to tenants, including a growing class of "neocloud" GPU-as-a-Service operators, and supply their tenants with "powered shell," i.e., with a powered structure to house computers for high-performance computing

(HPC) and AI tasks, while the tenants use their own servers and processors. These tenants, including neocloud companies like CoreWeave or Fluidstack, rent out processing power to startups, research labs, and enterprises.

Because hyperscale facilities often have greater credit quality tenants (the hyperscalers themselves), longer leases, and lower churn rates compared to colocation data centers, hyperscale deals are easier to underwrite compared to projects by neocloud tenants and sub-investment-grade operators. In contrast, many "powered shell" colocation operators, like TeraWulf and Cipher Mining, have pivoted from cryptocurrency mining, leveraging their existing power infrastructure and sites. The credit profiles of these borrowers are very different from hyperscalers. As a result, the structures used by these borrowers to access debt markets are different, and will be covered in one of the next articles of this series.

The market is bifurcating: hyperscale sits in one category, while edge, enterprise, and wholesale segments offer differentiated risk-return profiles for value-add and core-plus investors, often with faster paths to revenue. The edge thesis is distinct from hyperscale — edge data centers typically have a shorter path to revenue, with inference workloads requiring computing closer to end users predicted to drive the next deployment phase. Some market participants anticipate a shift toward edge over a five-to-seven-year horizon.

* * *

To understand the risks that confront financial investors across the data center industry, it is helpful to begin with the contractual foundation that underpins most data center financing — the lease agreement — and then examine, in turn, the execution, operational, market, and regulatory risks that can threaten that foundation.

* * *

4. The Lease Agreement: Foundation of the Revenue Model

At the heart of every data center financing is a deceptively simple question: who is paying the rent, and how reliable is that cash flow? Long-term lease or colocation agreements with investment-grade hyperscalers serve as foundational credit support in project finance and other financing structures for data centers. Hyperscalers frequently commit to triple-net leases under which tenants cover property taxes, building insurance, and maintenance costs in addition to base rents.

However, the fact that many data center projects are financed on the strength of a single anchor tenant introduces concentration risk. Financing parties should consider the implications of tenant default, early termination rights, termination payments, and the re-leasing market for built-to-suit facilities in the event of a vacancy.

Some argue that the era in which hyperscalers could unilaterally dictate lease terms is largely over. In fact, others argue that power-ready developers now hold more negotiating leverage than hyperscaler tenants. Offtake contracts must now be bankable, meaning that more aggressive termination rights and what many considered onerous confidentiality provisions have given way to more balanced terms.

Sub-investment-grade neocloud tenants introduce a new category of risk, such that investors need to hedge through parent guaranties and hyperscaler credit "wrappers," or underwrite the facility's intrinsic re-leasing value.

5. Power: The Primary Bottleneck

Beyond construction timelines, the single largest constraint on data center development is power. A new build can take multiple years to reach power. Projects built in 2022 took five years from interconnection request to commercial operations, compared to fewer than three years in 2008.

Data centers are energy-intensive operations. A recent report from the U.S. Department of Energy predicts that data centers will account for between 6.7% and

12.0% of total U.S. electricity consumption by 2028. At the same time, there is an estimated shortfall of 45 GW of power in the U.S. through 2028. Electric grid interconnections often take years, and that delay is growing in many markets. As a result, power availability, not capital, is the primary constraint on data center development. This makes Bring-Your-Own-Power (BYOP) solutions increasingly attractive despite their complexities and substantial cost to build, operate and maintain. De-bottlenecking solutions include converting crypto mining sites into data centers, siting at large U.S. nuclear power plants, co-locating new natural gas-fired power plants, and deploying on-site fuel cells. Natural gas is expected to dominate data center power over the next five years, but natural gas air permits are a major gating issue in many jurisdictions. Nuclear energy is attracting hyperscaler offtake agreements from Microsoft, and others, but new-build timelines and cost overrun risks remain material. Fast evolving Federal Energy Regulatory Commission, grid operator, and state rules further intensify the complexity of powering AI today.

Some participants are separating the financing of BYOP solutions from the data center components, tapping both the project finance capital market for power and the commercial real estate capital market for data centers.

6. Construction Risk

Construction risk follows naturally from the lease structure: if a facility cannot be completed on time and on budget, the lease — and any associated backstop — may never take effect.

Even when each contractor involved in the project is experienced, there may be externally-driven delays that can trigger lease termination rights and jeopardize any hyperscaler backstop.

Construction delays can unravel the credit architecture of an entire deal. Rating agencies are focused on construction-related risks, including delays, contingency budgets and other cost overrun risks and contractor experience.

7. Technology Obsolescence and the "GPU Debt Treadmill"

A key debate regarding potential cracks in data center financing revolves around GPUs and the risk that the technology lifecycle may not align with the longer financing maturities for the facilities that house them. While data center facilities have a lifespan of decades, the average lifecycle of a GPU/TPU is around seven years, though some believe it could be as little as three years.

This has been referred to as the "GPU debt treadmill." Even if the financing structure is ring-fenced and backed by an investment-grade counterparty, the real risk may lie in whether an equity issue today evolves into a credit problem over time.

The "powered shell" colocation model, in which providers supply a powered structure to house computers for HPC and AI tasks but do not provide the IT equipment, mitigates technology risk for the landlord. However, rapid advances in chip technology may require older data centers to upgrade power supply and cooling systems, and there is a risk of existing tenants opting to move to newer, cheaper facilities with more up-to-date designs, rather than renewing leases on existing sites.

8. The AI Demand Question and Bubble Risk

No honest assessment of data center financing can avoid the central question: will the demand materialize?

Morgan Stanley considers concerns that hardware and software efficiency gains will reduce overall demand for compute to be "overblown" and expects any excess capacity to be picked up by other players. The bank expected that GenAI would yield a positive ROI starting in 2025, with a $1.1 trillion revenue opportunity in 2028 and margins approaching 70%. The available evidence suggests that 2025 largely validated the expectation, although the returns are measured primarily through productivity improvements rather than directly traceable revenue growth.

But not all observers are sanguine. As Bloomberg has reported, never before has so much money been spent so rapidly on a technology that, for all its potential, remains largely unproven as an avenue for profit-making. The companies that ignited the AI investment frenzy, like OpenAI and Nvidia, have been instrumental in keeping it going by inking large and sometimes overlapping partnerships with cloud providers, AI developers, and other startups, and are seen as ratcheting up the risks of a possible AI bubble. S&P noted that colocation providers might not be inclined to give their neocloud tenants leeway if their "business strategy shifts significantly over the next year and demand for the planned AI chips to be placed in the facility diminishes."

Refinancing risk may become a real threat if credit quality deteriorates, as these companies will need substantial new capital to pay down outstanding debt.

9. Environmental Permitting and Community Opposition

Permitting challenges and local resistance have become serious obstacles to data center development. Large data centers have an enormous environmental footprint. They can consume up to 5 million gallons of water per day, place upward pressure on residential power costs, and generate friction with local communities due to air and noise pollution and negative visual impact.

Community opposition is a real headwind. State-level moratoria remain a genuine risk absent federal preemption and notwithstanding recent increased industry engagement with local and state stakeholders. The New York state moratorium, which pauses environmental permits for hyperscaler data centers for up to one year, went into effect July 14, 2026. This is a global trend: during London Climate Action Week, 40 mayors worldwide endorsed a pact led by C40 Cities to restrict resource-intensive AI-driven data centers and require developers to meet strict environmental and economic benchmarks.

Market participants who fail to account for permitting timelines and community engagement may find themselves with stranded sites.

10. The Road Ahead: Private Capital Steps In

Because of their ambitious AI-related capital expenditures, the hyperscalers already account for a large share of new corporate borrowing. Goldman Sachs expects liquid credit market saturation and issuer concentration constraints to become "somewhat more binding in coming years." This dynamic will likely drive further innovation in financing structures and greater reliance on private markets. Private credit sits at the intersection of significant AUM growth in higher-rate environments and the complex, large-scale, customized financing needs of the AI buildout.

There are several trends that will continue to evolve. Lender groups are likely to adopt a more cautious approach to deploying capital into the sector generally, with more scrutiny around whether the substantial upfront investment in data centers will translate into profitability. Existing data center debt will need refinancing in the coming years, creating both opportunity and risk. Asset-based finance will play a key role, providing a beneficial arrangement for all parties involved. Hybrid structures blending project finance principles with real estate and capital markets terms will become more common and attract a broader range of lenders, including insurance companies. The financing toolkit for data centers is expanding rapidly, and the structures emerging today will likely define capital markets innovation for the next decade.

* * *

In the articles that follow in this series, we will examine these capital solutions in detail. Our next installments will focus on the pioneering senior secured bond issuances from companies like TeraWulf, Cipher Mining, and Applied Digital — hybrid instruments that blend project finance discipline with high-yield bond market access — and the transformative role of structured joint ventures and investment-grade private credit in financing the world's largest data center projects. In future installments, we will discuss where traditional commercial real estate financing fits within these evolving hybrid frameworks as many data center financings continue to be structured around

commercial real estate loan instruments, including mortgages, deeds of trust, construction loans, and collateral assignments of leases, even as deal sizes climb into the tens of billions and sponsors increasingly layer on project finance and securitization structures. For financial sponsors, understanding these structures is no longer optional. It is the price of admission to one of the most consequential investment cycles of our time.

…

CLEARY GOTTLIEB