Qatar Investment Authority In collaboration with **datacenterHawk**

Data center foundations

A review of the data center industry & current trends

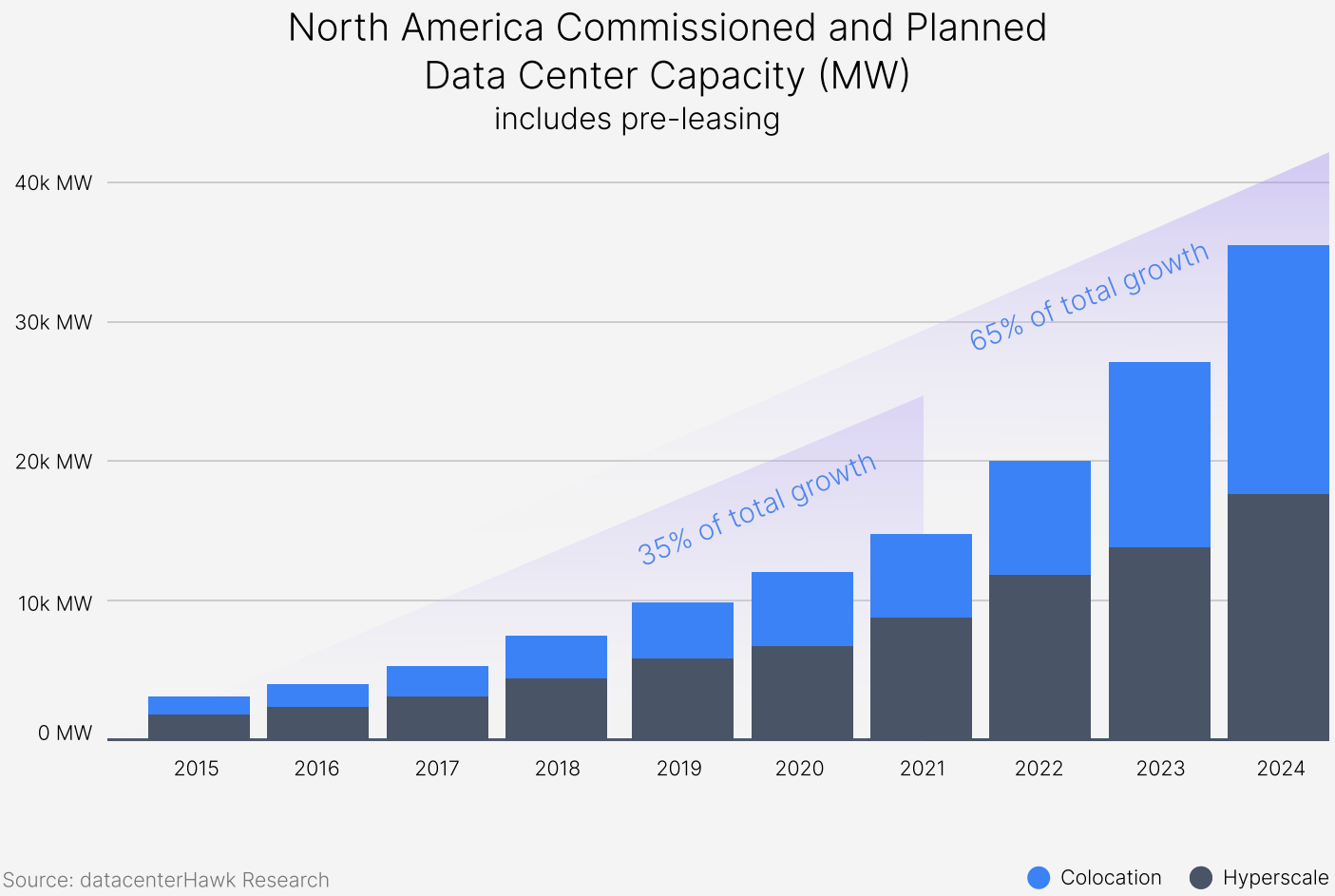
White Paper, June 2025

QIA جهاز قطر للاستثمار QATAR INVESTMENT AUTHORITY

DATACENTERHAWK

Executive Summary

This report provides a comprehensive analysis of the data center industry, covering market trends, investment opportunities, and the technological advancements driving growth. The industry has grown substantially over the last 10 years, and has nearly doubled its global size within the last three years.



Source: datacenterHawk Research

Data centers vary in size and function, ranging from hyperscale facilities operated by major cloud providers like AWS, Google, and Microsoft, to enterprise data centers serving mid-scale IT needs, and interconnection hubs facilitating high-speed data exchanges. Hyperscale data centers, in particular, have grown substantially, with some facilities now exceeding 300 MW to accommodate the expanding requirements of AI and cloud services.

Contents

From an investment perspective, data centers have emerged as a high-performing asset class, often outpacing traditional commercial real estate investments. Investors are engaging through various strategies, including sale-leaseback transactions, direct acquisitions, and joint ventures between financial and technology firms. Recent industry activity underscores this trend, with major partnerships such as those between Microsoft, GIP, and BlackRock focusing on AI-driven data center infrastructure.

AI continues to reshape the data center landscape, with AI workloads expected to represent 70% of total data center demand by 2030. This shift has introduced new challenges, including the need for high-density computing, advanced cooling systems, and high-speed networking capabilities. The increasing demand for AI infrastructure has also led to the rise of GPU leasing services, where companies like CoreWeave are capitalizing on AI chip shortages by offering GPU-as-a-Service (GPUaaS).

Energy consumption remains one of the most pressing concerns for the industry, as data centers require vast amounts of power, placing significant strain on energy grids. Sustainability efforts are now a critical focus, with companies investing in renewable energy sources such as solar, wind, and nuclear power to reduce their environmental impact. Some hyperscale operators are exploring on-site power generation solutions, including small modular nuclear reactors (SMRs) and natural gas plants, to ensure energy security and operational continuity.

Overall, the data center industry is undergoing rapid transformation, driven by technological advancements and increasing digitalization. While investment opportunities remain strong, addressing power constraints and sustainability challenges will be crucial to maintaining long-term growth and operational efficiency.

The Evolution of Data Centers

A data center is a facility designed to house servers and other IT infrastructure that support a company's technology by storing, processing, and distributing large volumes of digital data. Data centers come in a variety of sizes, ranging from IT floors in larger office buildings to massive hyperscale data centers that span hundreds of thousands of square feet.

Data centers are essential to nearly everything we do online. Anything from online shopping, emails, social media, online gaming, credit card transactions, media streaming, and much more are all processed through data centers. As technology evolves and cloud computing continues to grow, data centers are becoming increasingly important.

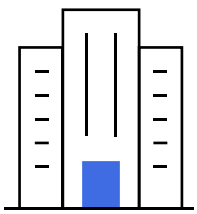
Companies typically have three options when it comes to their IT infrastructure.

- Self-built – The company builds, owns and operates its infrastructure
- Colocation – The company leases infrastructure from a data center provider, either as a single tenant or shared with multiple tenants
- Cloud – The company leases virtual data center space from a cloud provider

Types of data centers

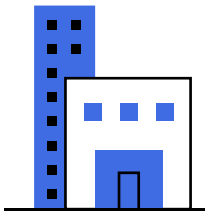
Nearly every company has some sort of data center need, and those needs often serve a unique purpose. As a result, there are a variety of different types of data centers, designed to offer robust security, low latency, high density, or to meet other unique client needs. This section will provide an overview of common types of data centers built globally.

Carrier Hotel



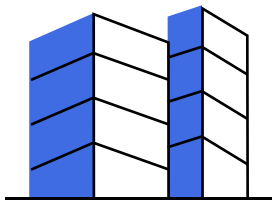
- Small leases 250 kW <
- Focused on connectivity
- Large number of tenants

Wholesale



- Medium sized leases (250 kW - 8 MW)
- Fit for data processing and storage
- Shared between several large tenants

Hyperscale



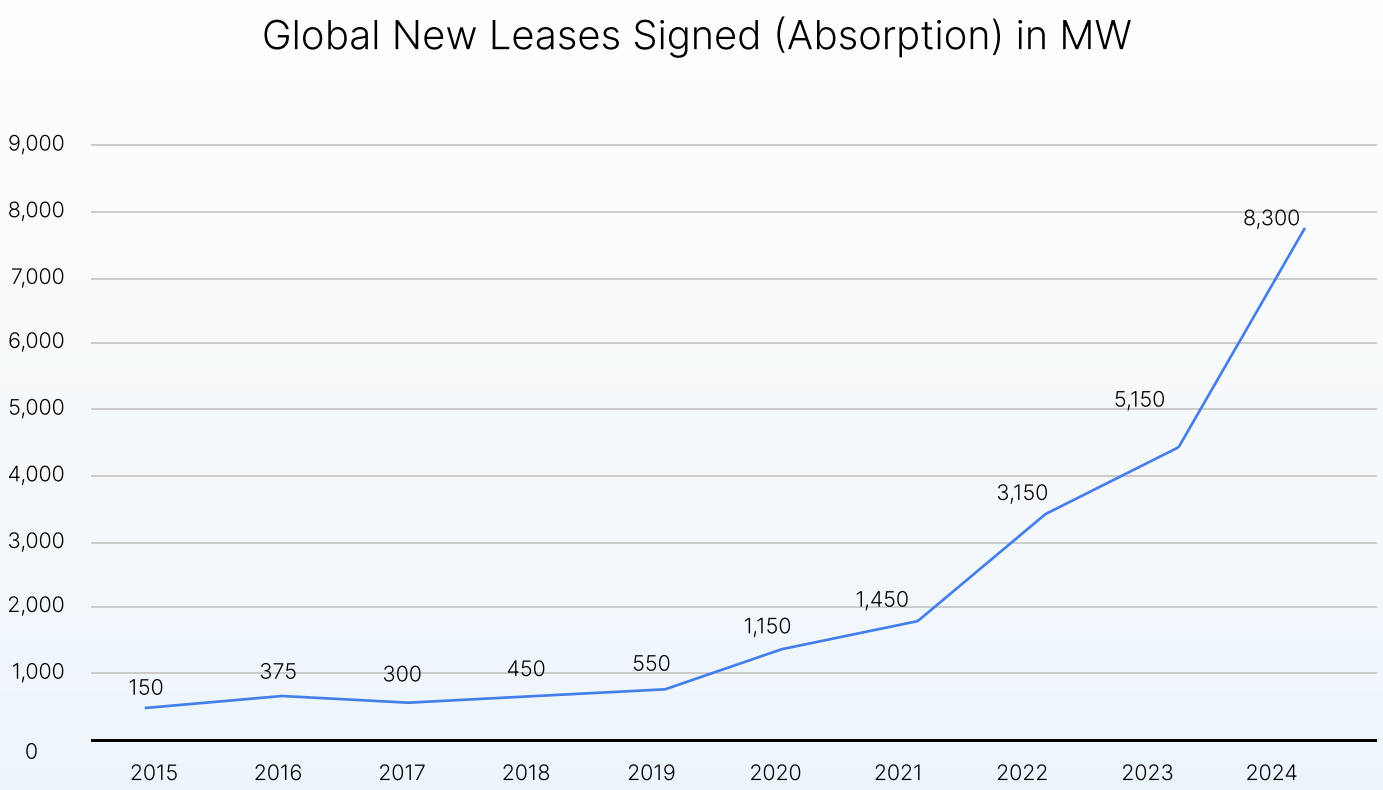
- Large facilities (10 - 300 MW)
- Cloud or AI operations
- Single tenant

Hyperscale

In broad terms, hyperscale data centers are facilities built on a scale much larger than previous development, often 50 MW or more within a single facility. These facilities are built to accommodate the massive storage and computing needs of major cloud and tech companies, like AWS, Google, Meta, Oracle, and Microsoft. Hyperscale data centers generally feature modular designs, enabling rapid expansion and design flexibility. They also have robust cooling infrastructure to provide higher densities of power per rack.

To manage such large operations efficiently, hyperscale data centers rely heavily on automation and software-driven monitoring systems, which optimize resources and detect issues in real-time. High levels of redundancy, including uninterruptible power supplies (UPS), backup generators, and multiple network connections, ensure continuous uptime

and reliability, even in the face of failures or during maintenance. This combination of scalability, efficiency, and resilience makes hyperscale data centers critical for supporting the infrastructure behind cloud services, artificial intelligence training, and data-intensive applications.



Source: datacenterHawk Research

Hyperscale catalyzed by cloud adoption

Prior to 2018, large-scale development often involved campuses with multiple data center buildings, but each individual data center would typically only offer up to 20 MW. Increased digital transformation, particularly during the COVID pandemic, resulted in substantial demand for additional cloud infrastructure. Simultaneously, opportunities for large plots of land became increasingly scarce. Developers were required to maximize the amount of power they can deliver per square foot, resulting in much larger data centers.

Hyperscale company growth strategies

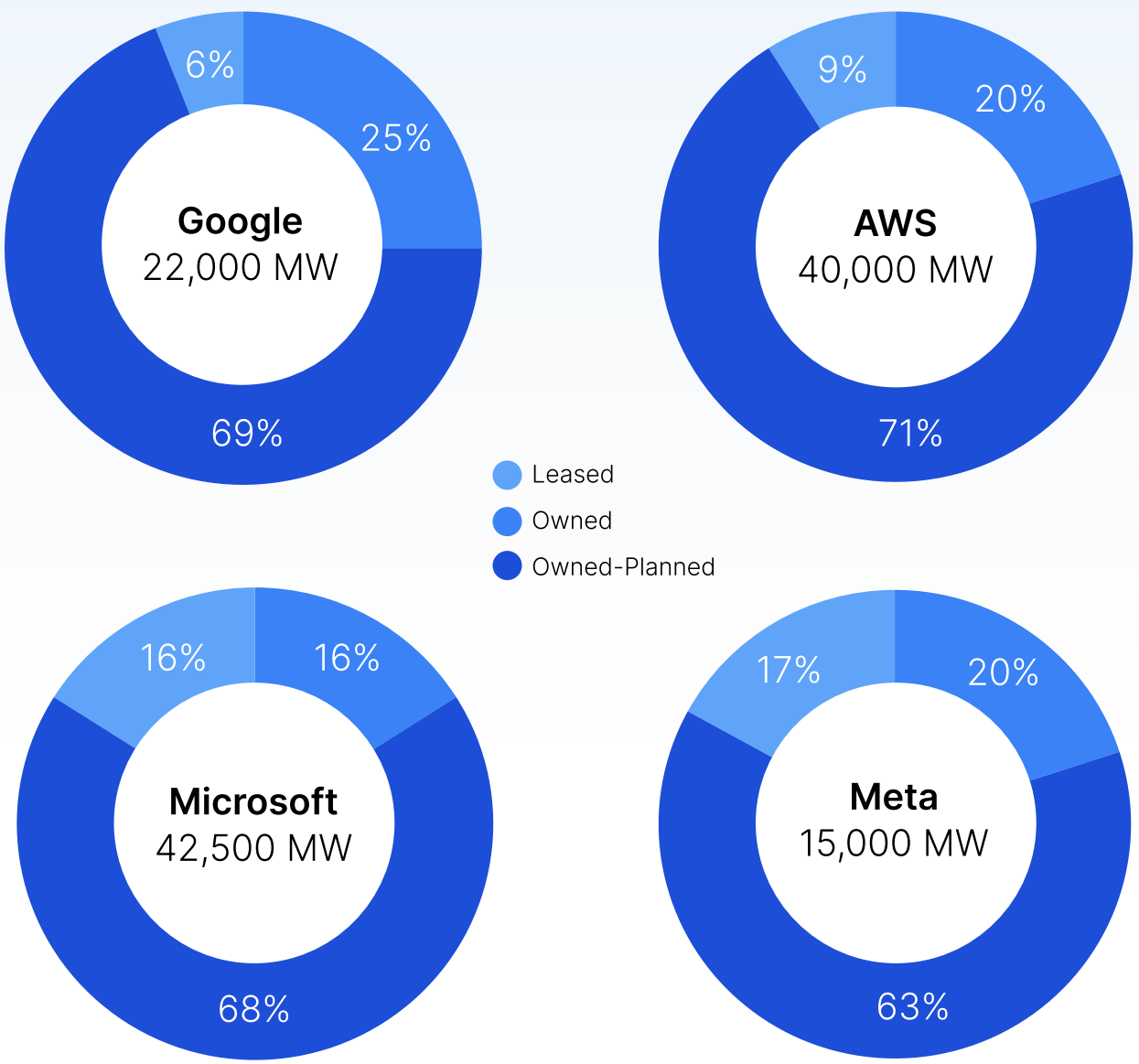
When hyperscale users (e.g., cloud service providers like AWS, Google, and Microsoft) need data center capacity, they often face the decision of self-building versus leasing. Each approach has its advantages and disadvantages, depending on factors like cost, scalability, time-to-market, and control.

FACTOR	SELF-BUILD	LEASING
Cost	High CapEx, lower OpEx long-term	Lower CapEx, higher OpEx long-term
Time-to-Market	Slower (years to build)	Faster (months to deploy)
Control	Full control over infrastructure	Limited to provider's offerings, although there's more control in a build-to-suit scenario
Scalability	Tailored to projected long-term needs	Easier to scale incrementally
Risk	Greater risk during construction	Shared risk with the lessor
Customizability	High	Moderate to low, although more customizability in a build-to-suit scenario
Geographic Expansion	Slower, requires local expertise	Faster, utilizes provider networks
Operational Complexity	High, requires in-house expertise	Low, outsourced to provider

In previous years, companies generally had more a more consistent approach to their lease vs build strategy. Increased global cloud adoption from 2020 onward, however, companies began diversifying their approach with a mixture of shell, turn-key, and BTS leases along with self-builds.

In major markets in North America and Europe, hyperscale companies often have a mix of owned and leased data centers within the same market. In Asia-Pacific, Latin America, and Middle East/ Africa, hyperscale groups prefer to enter a market through leasing and expand their presence with that data center provider.

Global Hyperscale Portfolio (MW) - Owned vs Leased



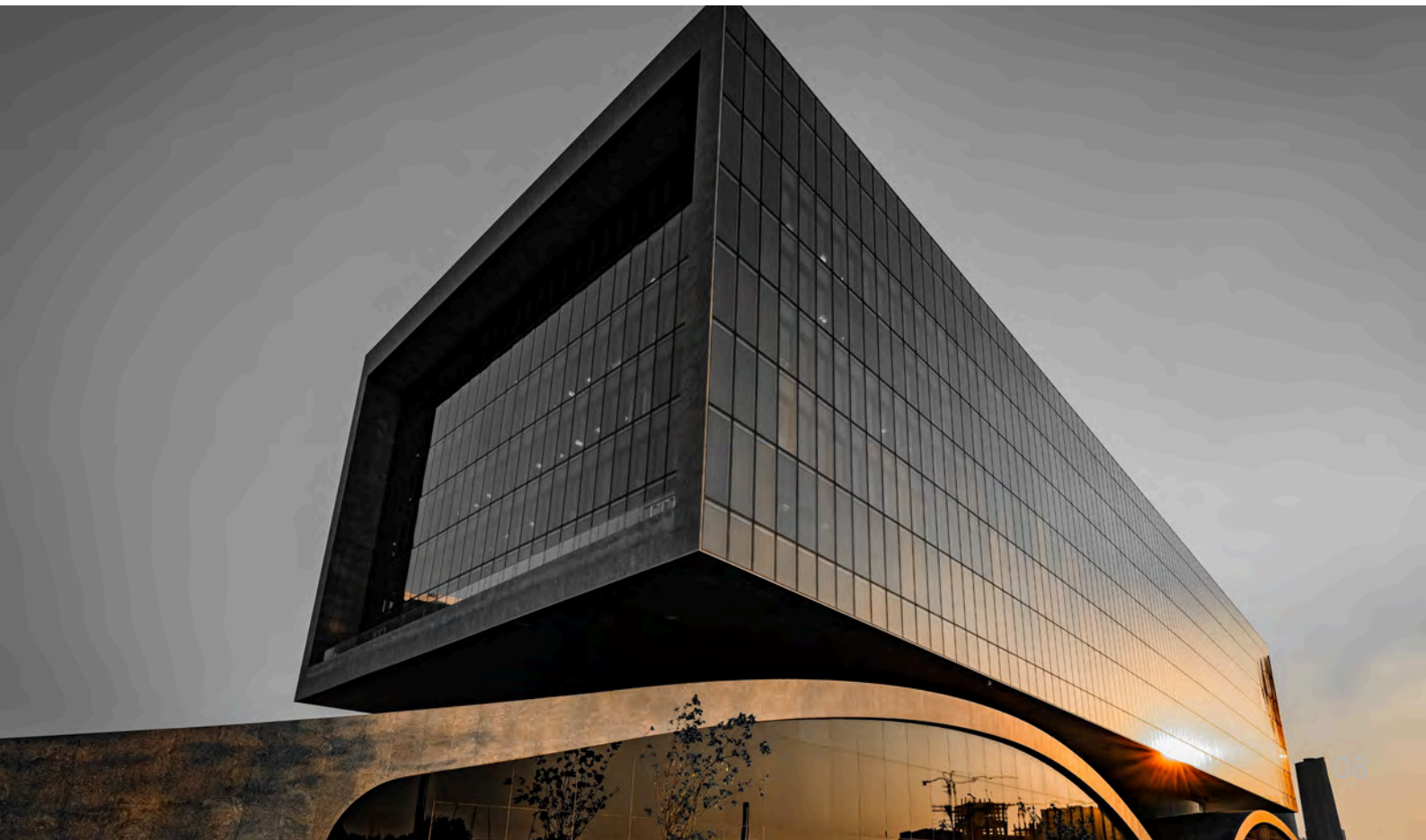
Source: datacenterHawk Research

Recent hyperscale activity

Hyperscale development is occurring around the world and is less focused on major markets, opting instead to focus on anywhere within a country that power is available. For example, Madrid and Barcelona are major markets in Spain, though Stoneshield is planning a €3.6 billion 500 MW data center campus near Bilbao, and Microsoft is investing €10 billion in a new campus in Zaragoza.

Hyperscale projects often include collaboration between the data center operators and the local governments. Sometimes this takes the form of governments proactively investing in data center and IT parks particularly zoned for data center development, like the Sedenak Tech Park in Johor or France and UAE's investment to construct a 1 GW AI park in France.

Emerging data center regions, particularly in APAC (Asia-Pacific) and MENA (Middle East-North Africa), are receiving substantial capital investment. The newly formed venture firm Fir Hills Inc intends to invest up to $35 billion in a 3 GW AI campus in Seoul. Oracle also recently announced its intent to increase its investment in the UAE "five-fold", while Equinix plans to invest over $1 billion in a single data center in Saudi Arabia.



Wholesale

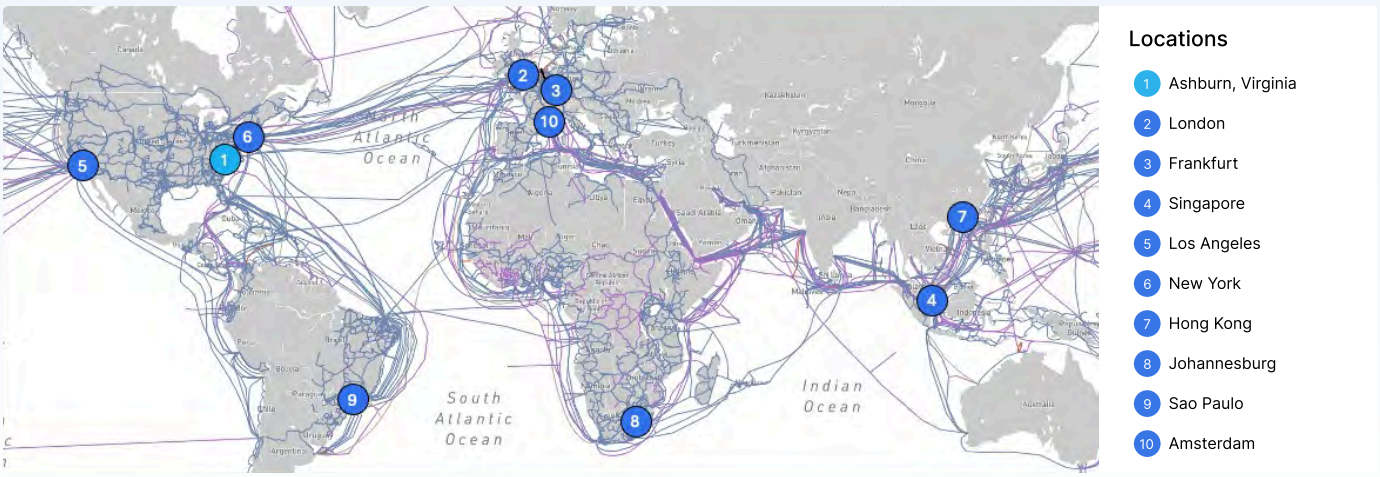
Enterprise, or wholesale colocation data centers, are large facilities designed to meet the mid-scale IT needs of multiple businesses. While hyperscale data centers are often single-tenant, enterprise data centers can house the IT infrastructure of multiple clients. These data centers are divided between multiple floors, suites, or data halls. Each floor or hall typically leased to a single tenant, though some floors and halls can be further subdivided for several clients to colocate in that space. The tenant can customize their suite to their specifications and can even request specific power, cooling, and generator infrastructure for their suite. Enterprise data centers are often built to offer 20-30 MW of leasable capacity that can be used by multiple tenants. Enterprise requirements are generally sized in the 250 kW – 5 MW range, though the rapidly escalating size of hyperscale requirements is expanding the enterprise range, with some enterprise requirements approaching 10 MW.

Carrier Hotel

Interconnection hubs, also referred to as carrier hotels, are data centers that have connections to a substantial network of telecom and fiber carriers. Inside carrier hotels, data and traffic is exchanged from one network to another, ultimately linking the end user to wherever their data is being stored or application is running. Carrier hotels are critical to a market's local network and to the larger national fiber backbone. Most carrier hotel leases are small, under 250 kW. The size of the requirement and the rich network ecosystem offered at carrier hotels results in leases that are much higher cost than enterprise or hyperscale leases.

Global Data Center Interconnection Hubs

Since the late 1990s, a small number of locations have served as the most critical hubs of the global internet



* Ranking based on QIA research regarding fiber density, cloud on-ramps, subsea cable landings, and network density.
Base map is Infrapedia global subsea cable map.

Data Centers Proving to be a Resilient High-Value Real Estate Investment Class

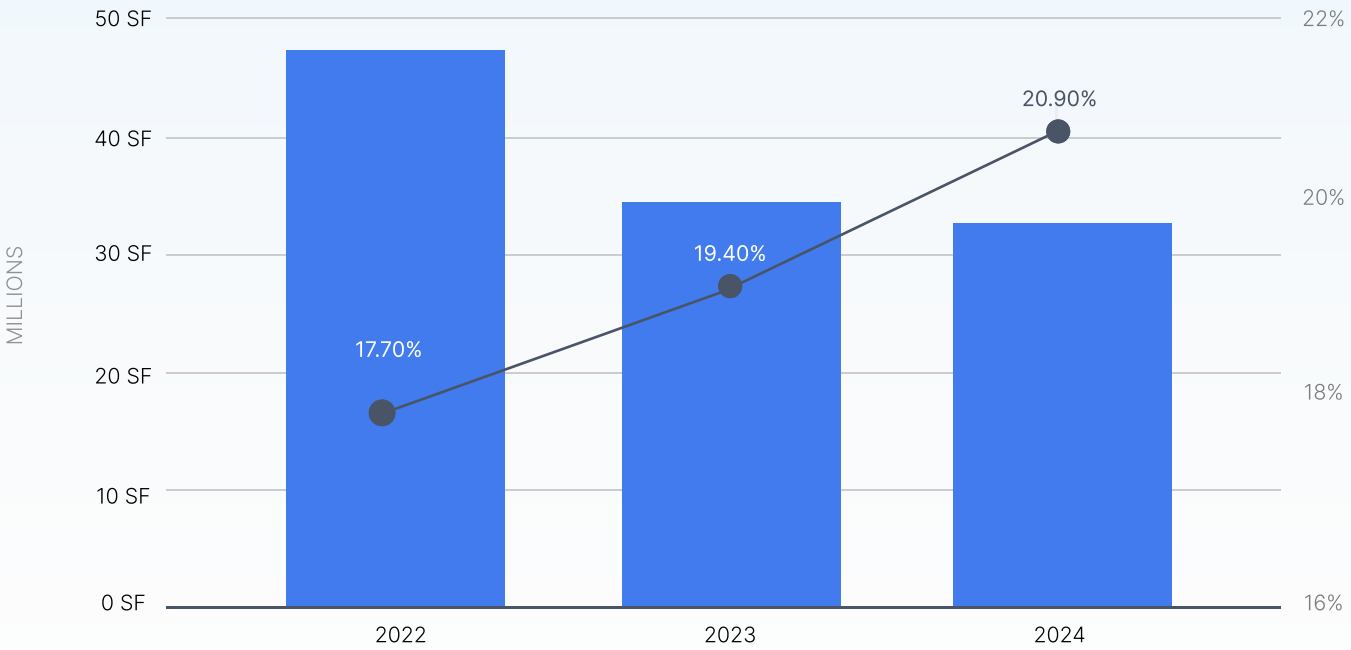
Data centers vs traditional real estate

Although often considered in the same category, data centers are substantially different from traditional commercial real estate. While office and industrial assets are often measured by total space, either square feet or square meters, data centers are measured by power in kilowatts (kW) or megawatts (MW). While racks and servers take up physical space, the important metric is how much processing power can be handled on the racks and in the entire data center. Demand and technological improvements have substantially increased the size of data centers in recent years, with some individual facilities exceeding 300 MW. Most data centers, however, are 1-100 MW.

Data centers are also more complex than traditional real estate assets. According to the Uptime Institute's 2023 report, a data center outage costs operators approximately US$12,900 per minute. With an average outage lasting about 90 minutes, the total cost typically ranges from US$1.0 million to US$1.5 million. To help mitigate this, data centers house complex mechanical systems, strict security, diverse connectivity equipment, and robust IT infrastructure. As a result, data centers require substantially more capital to develop and are similarly more valuable than traditional real estate assets.

Office leases are generally structured on a cost/SF basis, with the tenant paying to use a set amount of physical space. Colocation leases, however, are typically on a cost/kW basis.

New Office Space (SF) vs Vacancy Rates

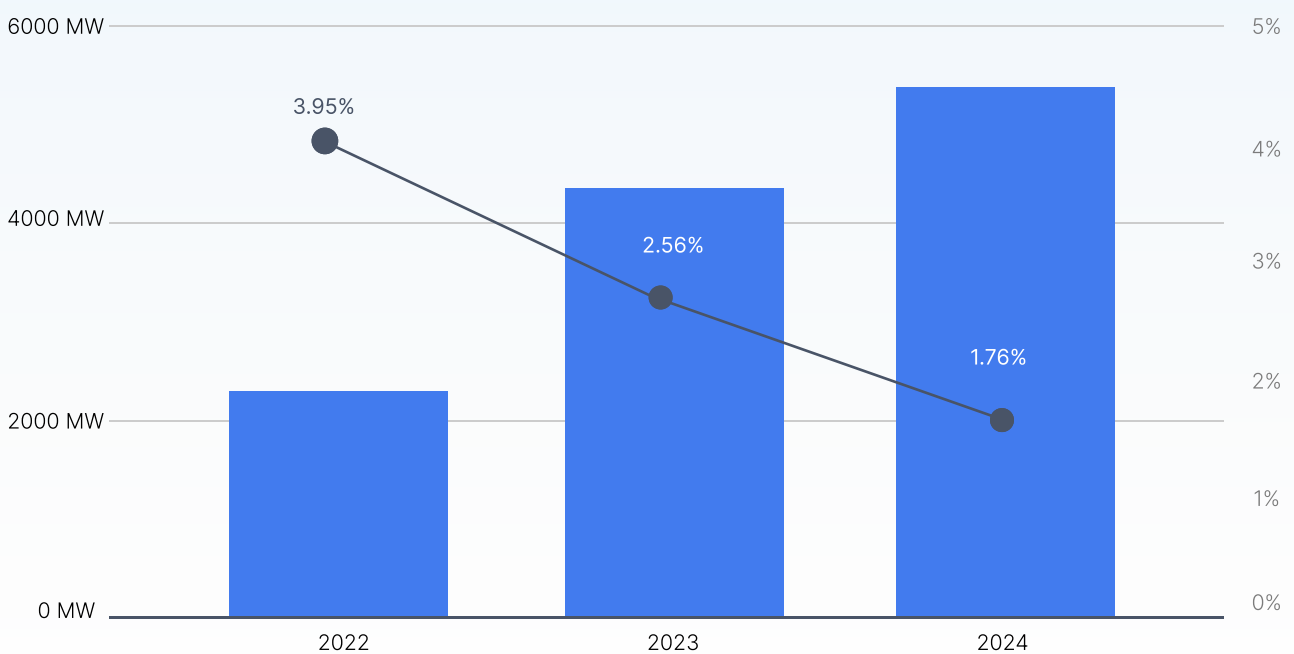


Source: Cushman Wakefield

Compared most other real estate asset classes, data centers are one of the few that are increasing in new deliveries and overall demand.

New Commission Data Center Power (MW) vs Vacancy Rates

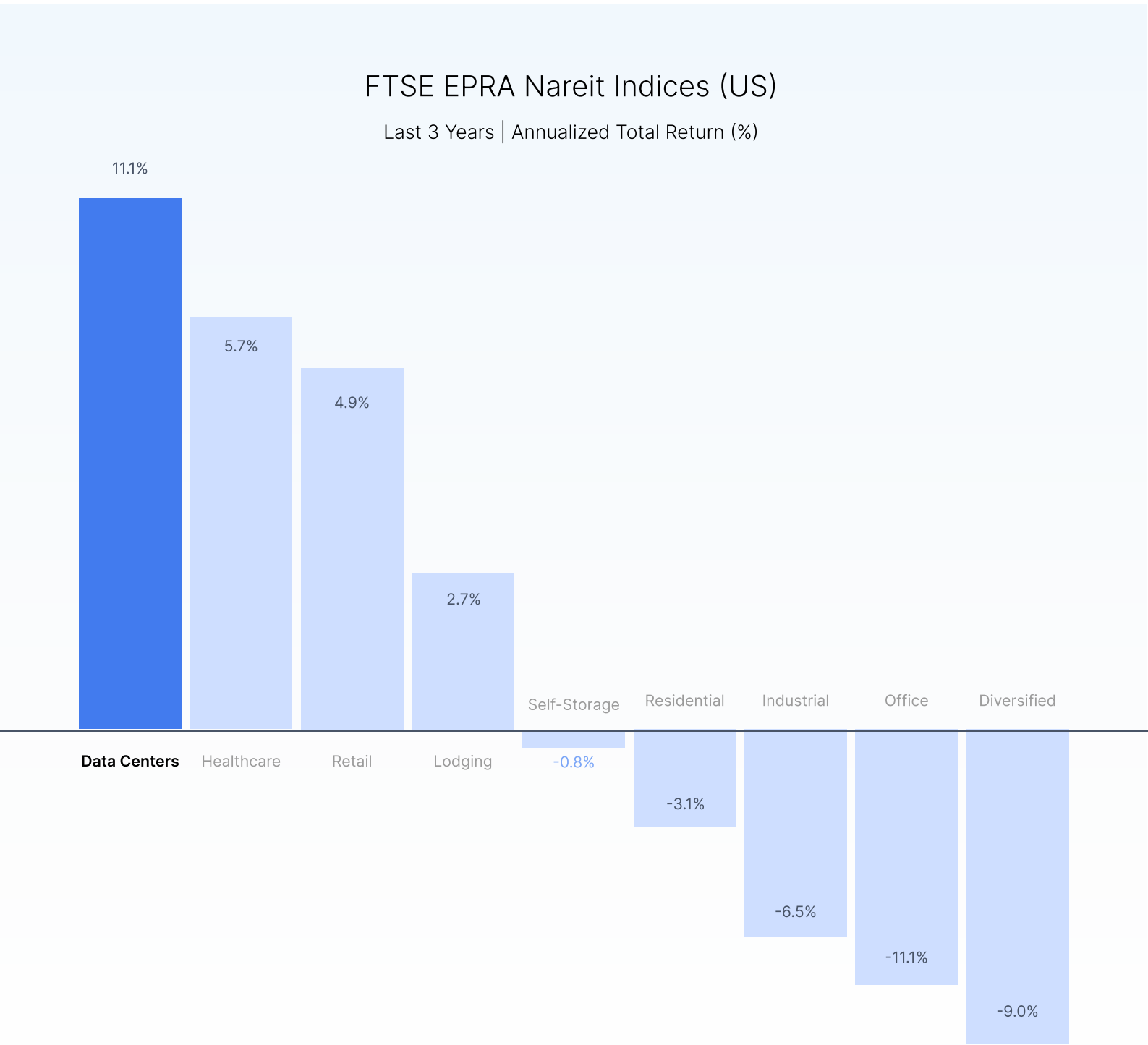
includes pre-leasing



Source: datacenterHawk Research

Data centers as an investment class

As stated, data centers are valuable real estate assets and, as such, are an emerging area of investment. As an emerging asset class, data centers have recently outperformed other forms of commercial real estate. Investment in the industry comes in a variety of forms.



Source: FactSet (data retrieved: January 2025)

Leading Global Providers by Power

PROVIDER	DELIVERED AND LEASED (MW)	PRE-LEASED (MW)	PLANNED (MW)	TOTAL (MW)
QTS	1,050 MW	2,075 MW	7,300 MW	10,425 MW
DIGITAL REALTY	2,625 MW	425 MW	2,500 MW	5,550 MW
STACK INFRASTRUCTURE	650 MW	1,250 MW	3,600 MW	5,500 MW
CloudHQ	575 MW	1,200 MW	3,050 MW	4,825 MW
VANTAGE DATA CENTERS	850 MW	775 MW	2,750 MW	4,375 MW
POWERHOUSE DATA CENTERS	0 MW	725 MW	3,625 MW	4,350 MW
Aligned Energy	675 MW	650 MW	3,000 MW	4,325 MW
CyrusOne	1,050 MW	200 MW	2,750 MW	4,000 MW
edged	50 MW	150 MW	2,950 MW	3,150 MW
NTT	1,100 MW	250 MW	1,500 MW	2,850 MW
EQUINIX	1,525 MW	100 MW	950 MW	2,575 MW
edgeconnex	325 MW	575 MW	1,675 MW	2,575 MW

Types of data center investment

Sale-Leaseback

A simple option for investment is the acquisition of individual data center assets through sale-lease transactions. Sale-leasebacks typically involve data centers owned and operated by non-hyperscale companies. These data centers are moderately sized, 10-20 MW, and built with robust infrastructure. Investment groups can acquire these assets and lease them back to the original tenant on a long-term basis.

Data center acquisition and improvement

Instead of acquiring existing assets and leasing back to the original owner, investors can also acquire data centers that aren't fully occupied and may need some improvements at a high cap rate. By doing so, they can make improvements to the facility and secure a long-term tenant. This increases the value of the asset, both in terms of the infrastructure present and the revenue stream available, and enables the investor to sell the data center at a substantial return.

Company acquisition and investment

Large-scale capital investment often involves providing funding for data center companies or the outright acquisition of an entire data center company. This can be done with the goal of consolidating several data center companies under one logo, or to act as a holding company with each individual company retaining their identity. Short of full acquisition, capital groups can also provide funding for data center providers' future projects. Examples of this include BlackStone's acquisition of QTS for $10 billion in 2021 or Digital Bridge's acquisitions of Yondr, Vantage, Switch, DataBank, and others.

Partnerships and joint ventures

Another common strategy is for capital groups to partner with data center developers for a specific purpose. These can be multi-group partnerships between capital groups, tech companies, data center developers, federal groups, and more, or joint ventures between data center developers and a source of funding. This can be done to develop a particular type of data center or to enable data center development in a new region. One example of this is AdaniConneX, a 50/50 joint venture between Indian holding company Adani Enterprises and the data center provider EdgeConneX. Through the partnership, EdgeConneX can bring its global data center development and operations expertise to India through Adani's capital and brand recognition in the region.

Recent investment activity

Billions of dollars are being invested into the data center industry, primarily to support massive AI initiatives.

Stargate

OpenAI has plans to develop five to ten data center campuses, primarily in the US, each capable of offering 1 GW or more. The scope of the project is immense, and will likely require the development of new power generation infrastructure. The company is partnering with SoftBank, MGX, and Oracle, with aims to generate $500 billion in funding.

GAIIP

BlackRock, Global Infrastructure Partners, MGX, and Microsoft formed a partnership, GAIIP, to aid in the building out of AI data center and power infrastructure. The partnership raised initial funding of $30 billion, with aims for $100 billion.

xAI

Elon Musk's xAI platform is growing out its presence in North America. Since its launch, xAI has raised more than $11 billion in funding to support the acquisition of chips at its AI data centers in Atlanta and Memphis. The company is actively pursuing additional options for further funding.

Data center investment risks

The data center industry is not immune to risks, and the volume of capital involved amplifies the risks. Despite demand being at an all-time high, there is no guarantee a client will be secured. The scale of development today often requires ownership of land before power can be procured, requiring larger capital expenditure with less security. For example, Equinix's proposed DB8 data center in Dublin originally had power from the grid, which the grid operator was then unable to fulfill. Equinix requested the development of a gas-powered generation plant, which was denied.

Market tendencies can also change, impacting the value of the data center asset. For example, AI data center requirements are particularly robust and can't be placed in most data centers, unless specifically built for AI. As a result, a large amount of data center assets aren't suitable for one of the leading sources of data center demand. Market competition can also change, like the introduction of DeepSeek into the AI sector. DeepSeek claims it was able to build out an AI model that performed at the level of ChatGPT and other leading AI models for a fraction of the cost, creating hesitation around the amount of capital OpenAI, xAI, and others claim they require. However, most AI projects' ability to raise capital has been relatively unaffected since the release of DeepSeek.

Most real estate investors know of the secular tailwinds and strong fundamentals of hyperscale data centers. However, QIA and datacenterHawk have found that much less is known about the subsector's unique risks, many of which would be familiar to investors in other types of long lease duration single-tenant real estate.

Best in class hyperscale data centers, which minimize these risks, have several key characteristics:

Cloud and Artificial Intelligence Driving Substantial Demand for Global Data Center Capacity

Cloud

What are cloud data centers

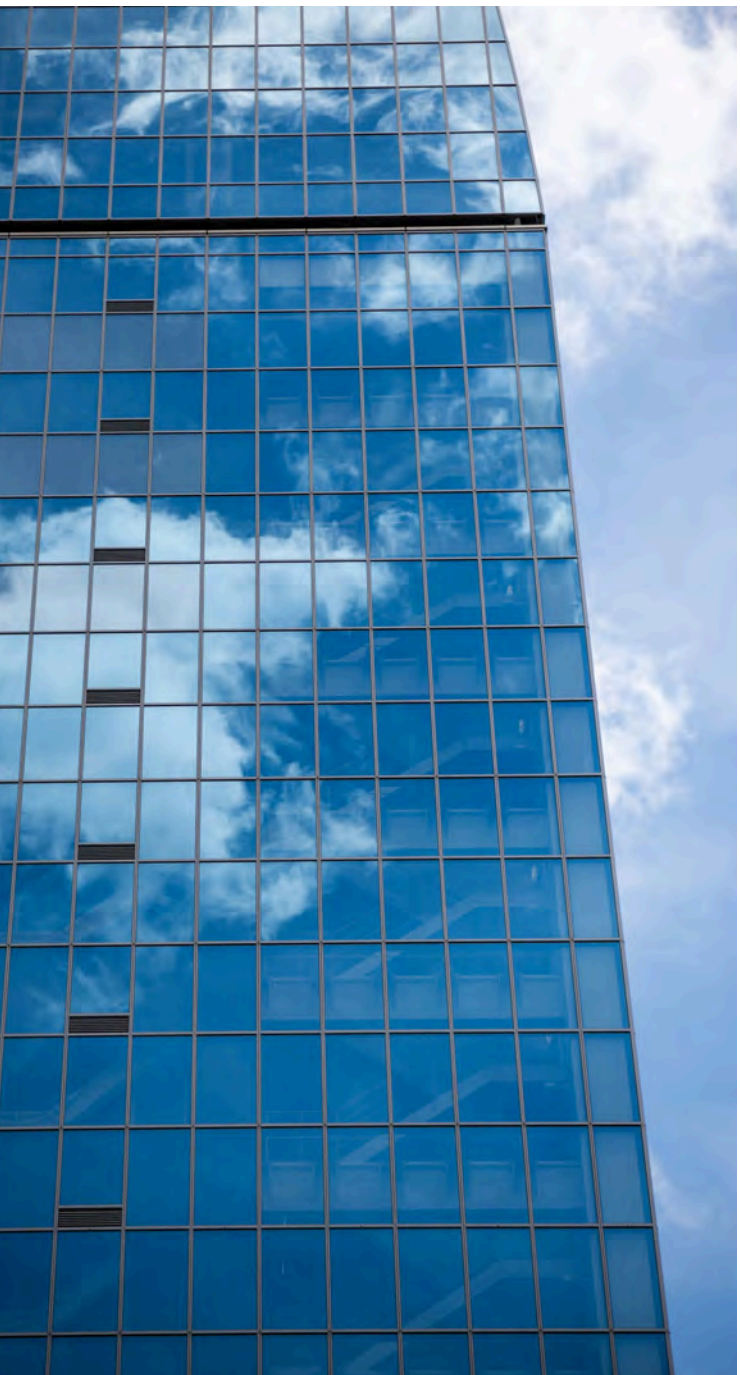
The cloud is a term used to define virtual IT infrastructure. While traditional data centers are tied to a physical location, the cloud is a virtual environment supported by multiple data centers but not tied to a particular one. When a client wants to lease space in the cloud, they are not leasing a physical server in a data center, but a virtual one in that provider's cloud ecosystem. To function effectively data centers that support the cloud are large-scale, feature robust infrastructure, and are highly connected, ensuring the cloud provider can do business with clients around the world.

Why companies choose the cloud

The two primary benefits of housing IT infrastructure in the cloud are risk mitigation and infrastructure elasticity. Since the cloud is a virtual environment shared among multiple locations, an issue at one location does not necessarily result in total loss of service for the user's systems. The virtual nature of the cloud also allows substantial flexibility. In a physical environment, adding more IT infrastructure involves the procurement, installation, and setup of new racks and servers. In the cloud, this can be done almost instantly, enabling rapid scaling up and down based on the user's needs. Although the costs vary substantially based on the application and system being run in the cloud, shifting to the cloud can also be a less expensive option than physical infrastructure.

Impact of the cloud

The introduction and general adoption of the cloud led many to believe it was the end of traditional colocation and physical infrastructure. For many small to mid-sized requirements, operating in the cloud is a much simpler solution. Previously companies of that size would operate small on-premise data centers with dedicated IT teams or lease from a data center provider. The cloud proved to be a better solution in many cases.



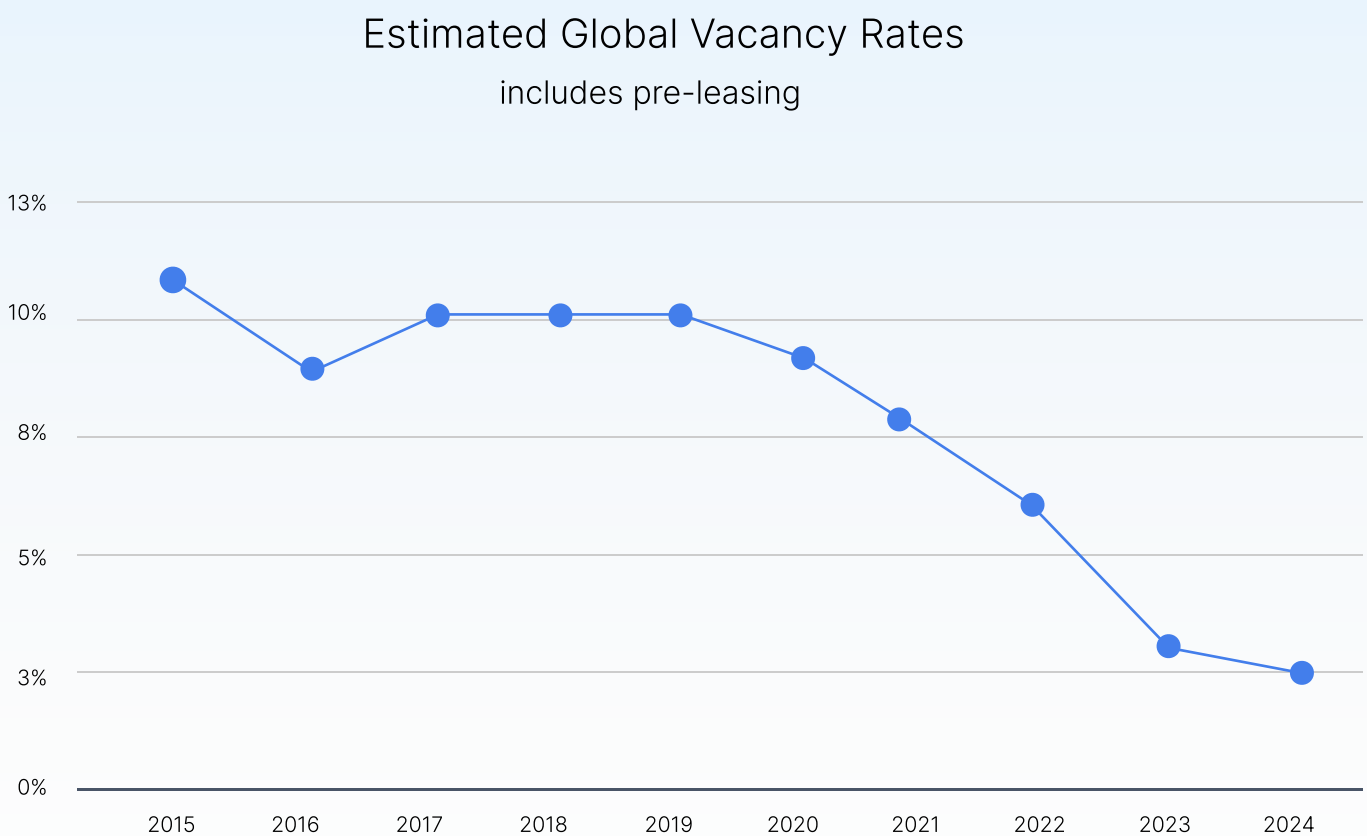
While cloud adoption did prove to be an attractive alternative, it did not materially eliminate data center demand. Firstly, many companies found an all-cloud approach to be more complicated and costly than their previous physical data center solution. As a result, it's common for companies today to employ a hybrid approach, with a mix of cloud and physical data center infrastructure to support their IT needs. Secondly, demand for cloud infrastructure generated additional demand for data center space.

While the cloud is a virtual environment, it is supported by physical data center infrastructure. Cloud companies do build their own data centers, but also lease data center space from data center providers. Increased demand for cloud infrastructure resulted in a substantial uptick in data center leasing.

Emergence of hyperscale

The uptick in data center leasing in response to cloud demand was on a scale that hadn't yet been witnessed in the data center industry. Beginning in 2017-2018, cloud providers escalated their leasing from individual halls to entire data centers. The COVID pandemic spurred further digital transformation and cloud migration, only increasing demand for large volumes of data center space. Hyperscale data centers grew from 30-40 MW to 70-80 MW and larger. Data center projects also grew in scale, from a single or a few data centers, to expansive multi-building campuses, some offering over 1 GW of total power. Campuses are now often fully leased to a single tenant.

Considering the high costs associated with delivering data center space, it's incredibly rare for there to be enough vacant capacity available in a single data center to support a hyperscale requirement. This means nearly every hyperscale transaction is a pre-lease, meaning the client is leasing the space before it is actually constructed, resulting in an ever-declining vacancy rate.



Source: datacenterHawkResearch

Artifical Intelligence (AI) and Machine Learning (ML)

What is AI

AI and machine learning applications demand substantial processing power, which is supplied by data centers equipped with high-performance computing (HPC) systems. These facilities support model training, data processing, and real-time inference for AI-driven applications.

Generative AI (GenAI) is driving significant demand in the data center sector. GenAI refers to a type of artificial intelligence capable of generating new content—such as text, images, code, music, and videos—based on patterns learned from large datasets. Unlike traditional AI, which focuses on analysis and prediction, GenAI models like GPT (Generative Pre-trained Transformer), DALL·E, and Stable Diffusion are designed to produce human-like responses, creative works, and realistic simulations.

Impact of AI growth

The rise of AI-specific chips, including GPUs (Graphics Processing Units) and TPUs (Tensor Processing Units), is reshaping the data center industry by increasing power consumption, cooling requirements, and infrastructure complexity. AI workloads demand high-density compute clusters, prompting data centers to adopt liquid cooling and more efficient heat dissipation methods. The surge in AI-optimized cloud services is also driving hyperscalers and colocation providers to redesign data halls, with power-per-rack requirements for AI-heavy deployments exceeding 40 kW. Additionally, networking and storage bottlenecks are becoming critical challenges, requiring high-speed interconnects and high-bandwidth memory to support the massive data transfer and computational loads generated by AI models.

Power-per-Rack Requirements for AI vs Non AI Workloads

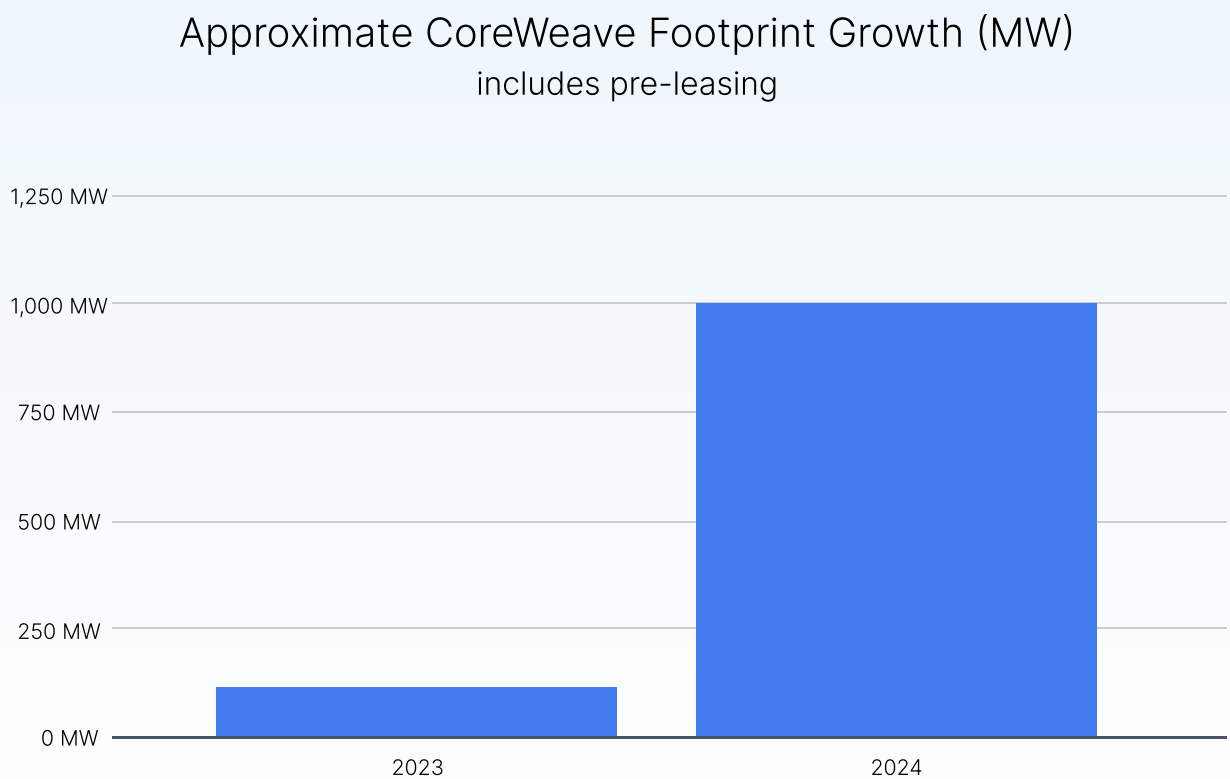
Source: datacenterHawk Research

AI growth driving investment in renewables

Beyond infrastructure, the energy intensity of AI is accelerating the transition to renewable energy sources, pushing data centers toward sustainability efforts aimed at reducing carbon footprints. Hyperscale companies are investing in solar, wind, and nuclear power, while discussions on energy-efficient AI models and green computing are gaining momentum. The shift toward AI-first architectures is transforming data center design, scalability, and operational models, spurring growth in hyperscale, colocation, and edge AI deployments. As AI adoption grows, the industry must innovate in cooling, power efficiency, and high-performance networking to support the next wave of AI-driven workloads.

Companies leasing chips as a new business opportunity

High demand for AI infrastructure is resulting in a shortage of AI-specific GPUs. The manufacturing of these chips and components is limited, and most supply is accounted for before manufacturing. Companies like CoreWeave and other AI focused companies are using this as an opportunity for additional revenue by leasing the operation of their GPUs to clients that need them. This GPU-as-a-Service (GPUaaS) is often charged on a per GPU per hour. Although AI GPUs are costly, these companies are able to quickly recoup their costs through this service.



Source: datacenterHawk Research

Forces behind AI growth

AI data center development is primarily driven by two types of companies: traditional hyperscale companies like Meta, Oracle, AWS, Google, NVIDIA, and Microsoft, as well as GPUaaS companies like CoreWeave, Lambda Labs, and xAI. For both types of companies, the motivation for development is the same. Companies need space and GPUs to train their AIs, and will either build it themselves or lease from a company that has chips. The primary difference between a company leasing or operating their own AI infrastructure is the amount of capital the company can spend on GPUs and infrastructure and whether they have the expertise to operate it. Most companies, other than hyperscale, lack the capital and expertise, but can still develop their business through the use of GPUaaS.

Where AI is growing

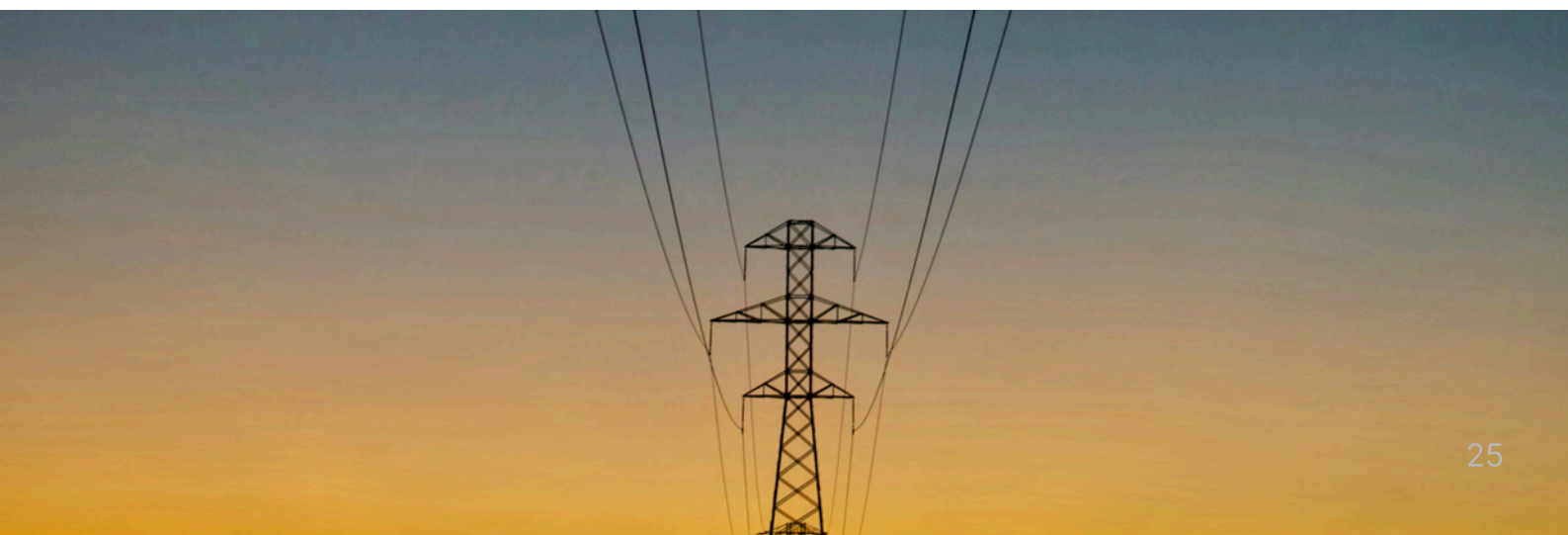
AI infrastructure and learning is less dependent on physical location and latency, meaning AI data center can be deployed in more rural and undeveloped areas without materially impacting operations. Given the scale of power and land required for AI, smaller markets can be an attractive option due to availability of inexpensive power. Markets that are too remote, however, may not be suitable candidates due to a lack of technical workforce and difficulties relocating a workforce to the area. As a result, development is still centered around major areas of population, though often substantially further away from the city center than before.

Data Center Demand Contributing to Energy Generation and Transmission Obstacles

As stated, the primary resources consumed in data centers is power. The power consumed is also on a scale much higher than other industries. To support this, data centers need robust power infrastructure, both in terms of generation and transmission.

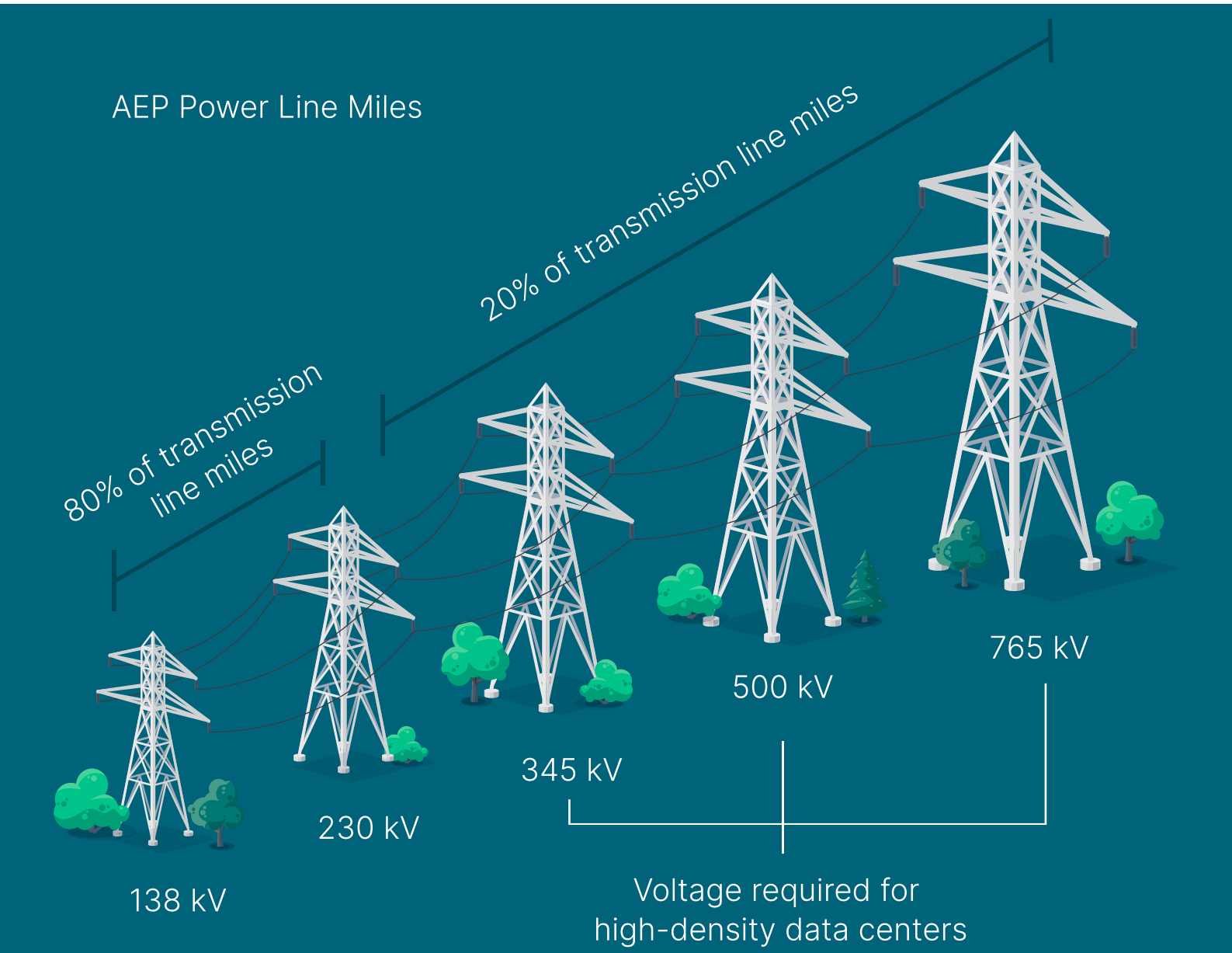
Generation

The amount of power generated and the source of that power is important for data center development. When considering a market, data center developers must consider the total amount of power available on the grid and other users that may be competing for that power, particularly with the scale of developments today. A grid may have excess power that can easily be used by data centers, but also may have trouble supporting several new 1 GW+ data center projects. The way the grid's power is generated is also important. Given the amount of power data centers consume, environmental impacts are under higher scrutiny. There is a push for data center projects to utilize renewable energy sources and many data center companies are tying loans and interest rates to their sustainability metrics.

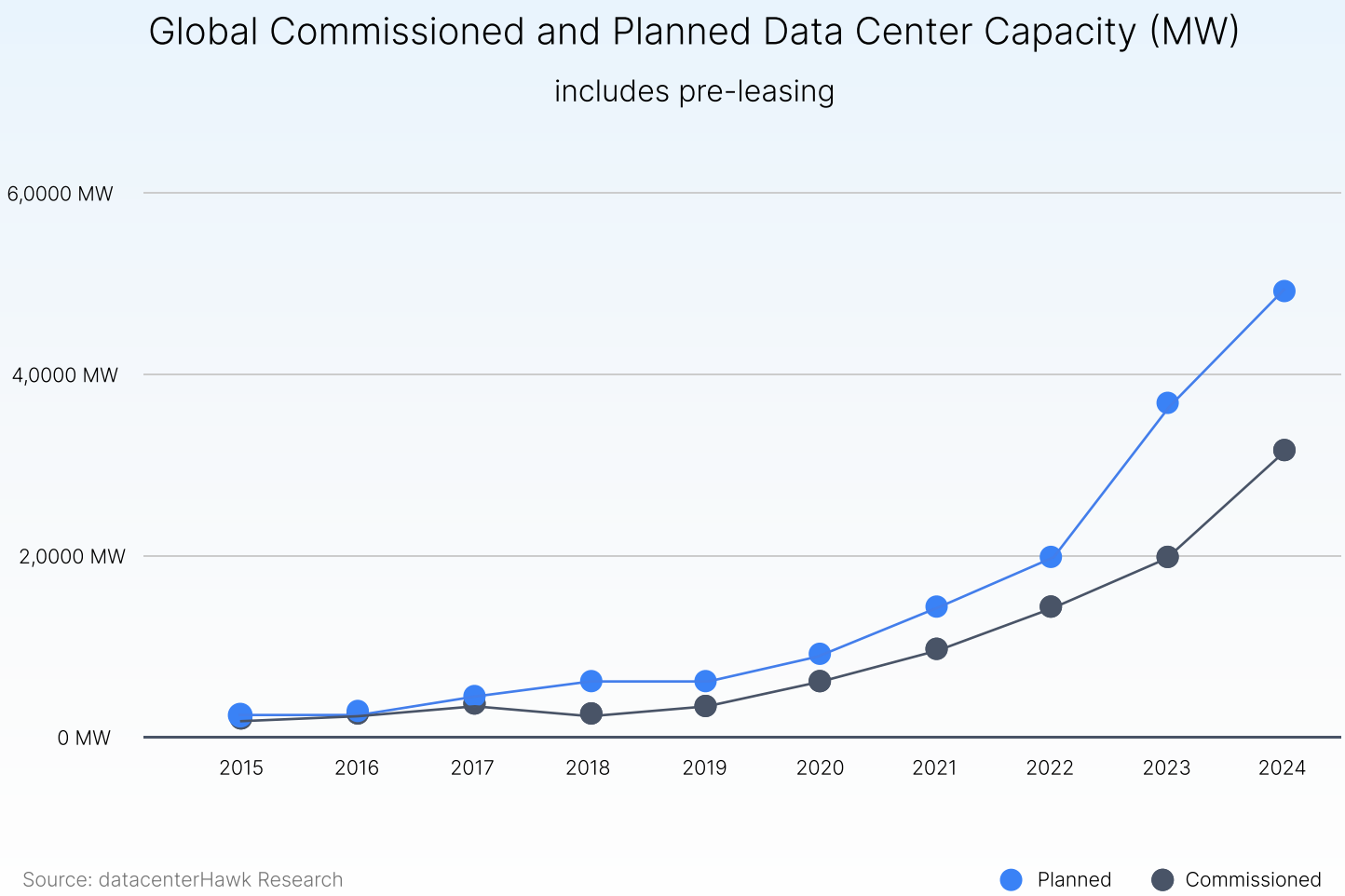


Transmission

Total power supply is critical for data center development, but also the transmission infrastructure in a market that delivers the power to the data center itself. Power is generated at the respective power plants and then transported from the plant to the larger grid through transmission lines. The transmission lines run to substations that convert the power to a voltage that is more usable for the end customer. Substations can be measured by the amount of power available to be distributed to end users, often in megawatts. Substations are typically shared between multiple consumers, but can also be built as dedicated private substations for a single user. To mitigate the disruption of their power supply, data centers can be connected to a substation though multiple lines in the event one line is damaged. They can also connect to multiple substations for further risk mitigation.



Source: https://interchange.puc.texas.gov/Documents/55249_7_1414447.PDF



The rapid scaling up of demand across the industry, as well as the size of each individual project presents a challenge for utility companies, both in terms of generation and transmission. These power issues are the foremost bottleneck for new data center development globally.

Grid traffic driving “behind the meter” development

Demand can quickly exceed surplus power available in the grid, even in emerging markets with little data center activity. New generation plants can help alleviate the strain on the grid, but the delivery of such projects is often lengthy and costly and can be a setback for company’s sustainability initiatives. To expedite the process and reduce the reliance on the grid, some hyperscale projects are implementing “behind the meter” solutions, developing their own power generation infrastructure and connecting it directly to their campus through direct and private transmission lines.

New substations and HVTs in development to meet scale

Even if power is in supply, the transmission infrastructure to delivery that power is often lacking. The high-voltage transmission lines (HVTs) necessary are uncommon and utilized quickly by existing projects. Shared substations are also too small, meaning nearly every new campus requires an on-site substation to be built.

In response to the amount of new infrastructure that utility companies need to install, utilities are passing some of those costs on to the data center operator and requiring payment for the power delivered, even if a tenant has yet to be procured. Developing new infrastructure comes with high costs and the utility companies are seeking opportunities like this to de-risk their capital expenditures.

Developers considering natural gas and nuclear generation alternatives

An alternative solution for generation and power bottlenecks is on-site power generation, either as a supply for the total power a site needs or as “bridge power” the help the site operate until power can be delivered by the utility company. Two primary areas of potential for on-site self generation come from natural gas and nuclear power. In many regions, natural gas is a readily available resource, and site selection teams are actively considering sites with direct connections to natural gas lines where they can build generation plants. While still in the theoretical stage of deployment, hyperscale companies are also considering nuclear power generation, particularly on-site SMRs (small modular reactors). While debate about the safety and utility of nuclear power is ongoing, proponents of nuclear power claim that SMRs would be a clean, quick, and effective solution to alleviate power supply and transmission concerns.

Conclusion

Looking ahead, the data center industry is poised for continued expansion, driven by the accelerating demand for AI, cloud computing, and high-performance digital infrastructure. As hyperscale facilities grow larger and AI workloads dominate processing demands, the sector will need to adapt through innovative solutions in power management, cooling technologies, and network optimization. The rise of GPU leasing services and AI-driven infrastructure investments signals a shift toward more dynamic and flexible computing models, ensuring that the industry remains at the forefront of technological progress.

However, as the industry scales, sustainability and energy efficiency will become defining challenges. The increasing strain on power grids highlights the urgent need for renewable energy integration, on-site generation, and next-generation efficiency solutions. Companies that successfully balance rapid growth with sustainable innovation will shape the future of the industry, securing long-term resilience and profitability. By embracing cutting-edge advancements and strategic investments, the data center sector is set to remain a cornerstone of the digital economy, fueling the next era of technological evolution.

About datacenterHawk

datacenterHawk empowers people to make better decisions around data center real estate by providing the most accurate and useful information to the market through a subscription-based platform and consulting services.

What makes datacenterHawk uniquely positioned as a market leader:

Deal-backed data
Tracking deals means our data moves 6x more per quarter, 6-12 months earlier than just tracking construction.

Experienced team
Our regional experts have 2.5x more in-region industry experience than the market average and are equipped to deliver the most relevant, detailed data

Boots on the ground in 5 continents
North America, South America, Europe, Asia, & Australia

Faster updates
We update our data 4x more often and 3x faster than the market average.

60+ markets tracked quarterly
With market coverage expanding into emerging regions.

Trusted by industry leaders for 10+ years
Teams large and small have trusted datacenterHawk for over a decade to help guide their largest strategic decisions.

datacenterHawk was founded on the vision that companies making million- and billion-dollar decisions around digital infrastructure needed a trusted resource like those that existed for other commercial real estate sectors.

hello@datacenterhawk.com | datacenterhawk.com

7616 Lyndon B Johnson Fwy, Suite 720 | Dallas, TX 75251