IRON MOUNTAIN®
DATA CENTERS

Infrastructure planning report

North America

Contents

The latest trends, opportunities and challenges for digital businesses seeking data center infrastructure in North America

The North American data center market

North America is the largest and most mature data center market in the world, with several thousand data centers, well over 10,000 MW of inventory and a 38% share of the global market. It is also the test bed and driving force behind the majority of global data center growth. What happens in North America soon rolls out in other markets around the world.

The level of data center construction in North America's top markets has risen hugely - by 70% from H1 2023 to H1 2024 (CBRE) with almost 3.9 GW (gigawatts) of capacity under construction. In H1 2024 over 500 MW - roughly equivalent to all of the existing capacity in Silicon Valley - were rolled out in the eight biggest markets. Year-over-year that translates to over 1.1 GW of new capacity. And with the additional impulse of AI, demand will continue to outstrip supply for the foreseeable future.

Primary Markets

With total inventory ranging from 500 MW to over 3 GW, the leading data center markets include Northern Virginia, Dallas and Fort Worth, Texas; Silicon Valley in Central California; Chicago; the New York Tri-State Area; Phoenix, Arizona, and Atlanta, Georgia.

Secondary & Growth Markets

Secondary markets which account for around 20% of total demand include Las Vegas and the Seattle/Quincy cluster; Austin San Antonio; Southern California/LA; Toronto and Montreal in Canada; Columbus, Ohio; and new hubs with exponential growth such as Reno and Salt Lake City. New facility construction is also rising fast in Houston, Denver, Charlotte/Raleigh, Minneapolis, Northern Indiana, Nashville, Arkansas, Vancouver in Canada and Queretaro in Mexico.

Recent Developments

With power and planning constraints in a number of major hubs, AI investment is driving construction in new more remote markets as well as current hubs. Increased demand for high density computing is also creating price disparity between new data centers and legacy facilities. Alternative power supplies are being acquired to take strain off the grid, particularly low-carbon options.

North America

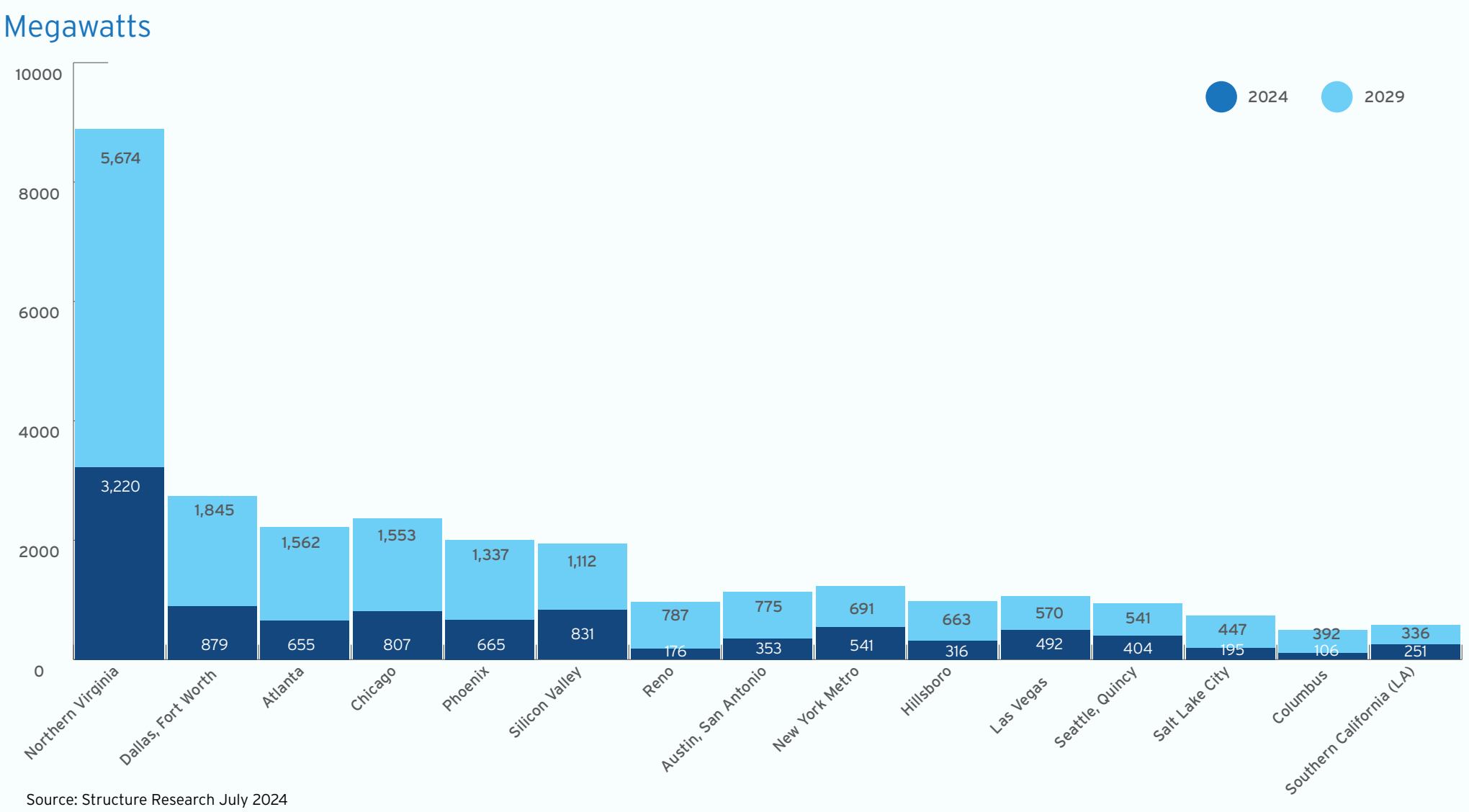
2025
39% of global market
($41 BN)

2029
35% of global market
($66 BN)

CAGR 12.8%

Figures approximate. Source: Structure Research

North America Market projections (MW) 2024 - 2029



Source: Structure Research July 2024

With more flexibility in site selection for Gen AI training facilities and land and power constraints growing in well-established markets such as Silicon Valley, Las Vegas and New York, the primary North American markets are set for a shake up over the next few years.

Opportunities

Given the strength of demand from all sectors, there is a wealth of opportunity in the North American market today. A massive amount of new state-of-the-art infrastructure is under construction across a wider geographic base than ever before.

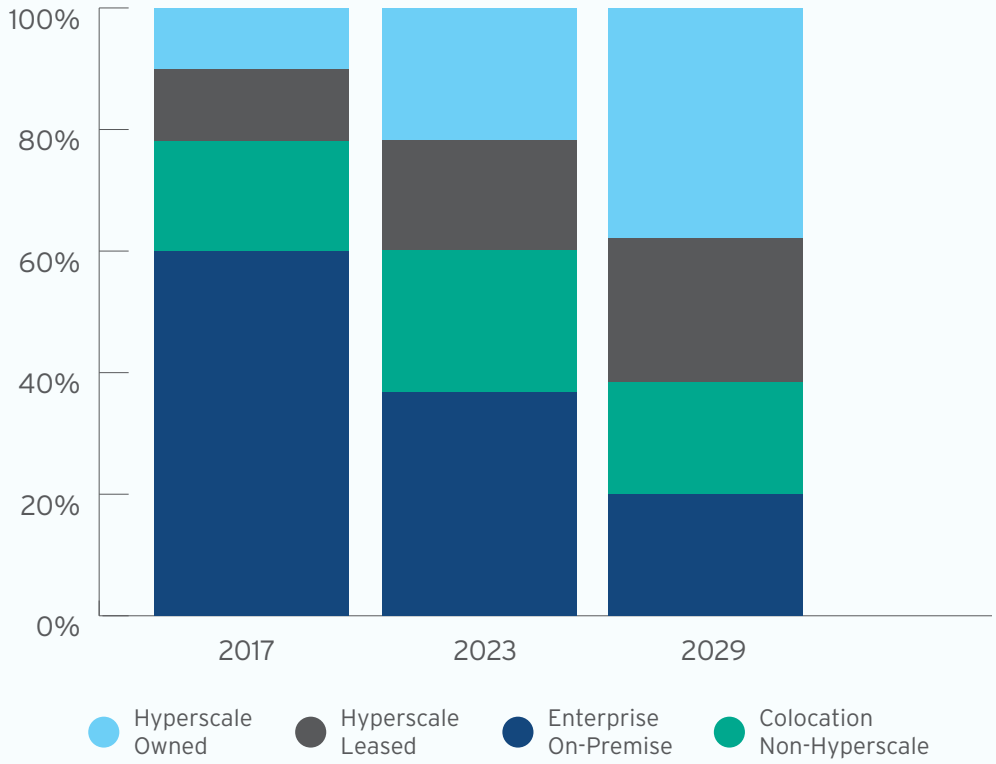
Hyperscale cloud growth

While enterprise and public sector demand remain strong, it is demand from hyperscale cloud providers that is the single largest force in the market currently. According to Structure Research, hyperscale cloud growth is set to run at a new higher CAGR of 25.9% to 2029.

Because it is the prime mover market, today there is much more owned hyperscale data centre capacity in the USA than elsewhere in the world; hyperscalers own twice as much data center space as they lease in this region.

For hyperscalers, the cloud is still the key to growth, overlaid with a bonus layer of AI demand. According to Synergy Research Group, ten years ago businesses spent twelve times as much on their data centre hardware and software as they did on cloud infrastructure, but today they spend three times more on cloud services than they do on their data centre infrastructure. SaaS, social networks, e-commerce and new higher bandwidth apps continue to drive both hyperscale and colocation expansion.

Global data center capacity trends



Source: Synergy Research Group

Global market trends show the steep rise in hyperscale demand and data center ownership, and the rapid drop in on-premise enterprise data centers. Globally, hyperscale clouds now support over 40% of IT load, and Synergy forecast this will grow to 60% by 2029.

AI market development

Overlaid on these strong fundamentals, the growth accelerator is AI. Actual growth of the AI market is still hard to predict currently, with predictions ranging from 19% (e.g. Precedence Research) to 55% (Bain & Company). AI is already creating huge demand for high-power data centers and supporting infrastructure.

This growth is from an extremely strong base in North America. As home to the world's largest tech giants, all of whom are investing heavily in AI, North America had an impressive 37% share of the $540 billion AI market in 2023 (Precedence). New GPU-oriented cloud infrastructure providers are also scaling up to offer services alongside established cloud providers.

Rising secondary markets

Infrastructure for Generative AI training models is less location-sensitive than for traditional consumer clouds. This is changing the data center development landscape. A lot of larger deployments are taking place in smaller, more remote markets with less dependency on low latency but reasonable proximity to core cloud nodes in primary markets. This has created demand for large-scale data center capacity in secondary markets where land and power are less constrained and more affordable.

Edge

Edge data center investment is also growing fast, driven by 5G, IoT, and the demand for low-latency application processing close to data sources. According to Credence Research the edge market is expected to achieve a compound annual growth rate (CAGR) of more than 20% over the next few years.

Subsea connections

North America has a huge range of submarine cable connections, but these continue to proliferate. More diverse landing points like Virginia Beach and Myrtle Beach on the east coast are being added, and new trans-Pacific landings on the west coast are planned for Canada and Mexico. Florida, a key hub for Latin American transit, has new routes from Naples on the west coast, as well as the recently-activated EllaLink cable. New cables will soon boost capacity to the south, including Firmina, Carnival Submarine Network-1, and AMX-3/Tikal.

Clean power growth

Due to the extreme power requirements of GPUs, and the rising profile of data center power usage, the rise of AI is being accompanied by a strong focus on new sources of low-carbon energy. Data centers account for over 2.5% of total U.S. power consumption and this is set to triple to 7.5% by 2030 (Boston Consulting) with total demand from data centers forecast to reach 35 GW by 2030 (Newmark property consultants).

Data center operators already lead in renewables investment and deployment, accounting for two thirds of corporate renewable purchases in the USA. Clean energy PPAs for new facilities across North America will continue to drive huge demand and growth in wind, solar and low-carbon storage solutions, expanding the green grid to support the growing digital economy.

Issues

Most of the issues facing the North American data center market are either related to space and power availability or pricing. The power-intensive requirements of Gen AI are reshaping the market, and in the face of the climate crisis new power supplies need to be carbon-free.

Vacancy rates

Supply is struggling to keep ahead of demand. According to CBRE, data center vacancy rates have plunged to a record low of 2.8%. Chicago experienced the sharpest decrease in availability to 2.4% from 6.7%. In Northern Virginia vacancy rates fell to 0.9% from 1.8% despite an 18% increase in inventory. Northern Virginia's rates are the lowest of any major market worldwide - lower than even space-constrained Singapore. Pre-booking at least two years in advance of completion is now the norm. According to CBRE, nearly 80% of the data center space currently under construction in primary markets is pre-leased.

Power

Lack of power has acted as a drag on expansion in established locations and has started to reshape data center geography. Over the past year, lead times for power delivery have stretched in markets beyond Northern Virginia, the world's primary data center hub. Phoenix, Atlanta, Dallas and Columbus, all of which have been expanding rapidly, started experiencing delays and constraints in supply. Power delivery timelines are also expected to lengthen due to shortages of equipment like transformers, switches and generators.

A more multifaceted approach is emerging to cope with power needs, involving alternative energy sources, regulatory improvements, and infrastructure upgrades. Small modular nuclear reactors are also set to play an increasing part in the power jigsaw, but will take time to power up.

Pricing

Fueled by supply chain issues and strong demand, average pricing increased by 6.5% in the first half of 2024 (CBRE) with the sharpest rise in Georgia (26% year-on-year), mainly due to AI provider demand. This upward trend is expected to continue. A new pricing disparity has also begun to emerge in response to the high densities and new cooling configurations required by AI, with newer or upgraded data centers typically costing more than older ones.

Sustainability reporting

In other global regions, new stringent sustainability reporting obligations are now coming into force. These cover metrics such as cooling efficiency, recycling waste heat and the use of low-to-no-carbon energy. In North America regulations are more localised, with energy efficiency regulations in action in states like California, Washington, and British Columbia in Canada.

However, many of the leading data center operators and customers have already set ambitious emission reduction targets and committed to report on these. In addition, while Scope 3 reporting (reporting of emissions from colocated IT) has not yet become obligatory at a federal level, this may change in the near future, as emission reduction targets loom.

Over 23,000 companies - representing $67 trillion in market capitalization – now disclose their emissions through CDP. According to Accenture, 37% of companies have now set net zero targets. The SEC are currently considering making climate governance, climate-related risks and impacts, and the reporting of Scope 1, 2 and 3 greenhouse GHG emissions obligatory.

Key Hubs

Supply is being snapped up across all of North America's primary hubs. With total inventory ranging from 541 MW to 3.2 GW, the leading data center markets include Northern Virginia, Dallas-Fort Worth; Silicon Valley; Chicago; Phoenix, Arizona; Atlanta and the New York Tri-State Area.

Northern Virginia

The world's leading data center hub shows no signs of a slowdown, despite power and land acquisition issues. Development is spreading to adjacent areas like the I-95 Corridor, 3-4-story facilities are becoming common, and Dominion Energy will deliver more capacity in 2026 and 2028.

Dallas-Fort Worth

Demand is extremely high in the Dallas-Fort Worth area, with pre-leasing of current construction at a record-high of 94.5% (CBRE). Power is hard to come by and prices are rising fast.

Silicon Valley

Silicon Valley is a mature hub at the heart of global tech development with a highly skilled talent pool and excellent connectivity. However due to limited land and power availability alternative west coast hubs are now growing faster.

Chicago

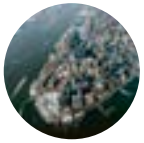
Chicago is a hugely popular data hub both for local commerce (particularly trading) and to serve major markets like Milwaukee, Detroit, Indianapolis, Detroit and Toronto in Canada. The city has exceptional connectivity, with leading Internet Exchanges and new inter-state backbones being built.

Phoenix

Phoenix is a fast-growing hub on the west coast with steep demand and a major multi-Gigawatt development pipeline - the second largest in the US after Northern Virginia. Investment in advanced water management is necessary here to conserve this limited resource.

Atlanta

The industry tends to view Atlanta on the east coast as an alternative to Northern Virginia with low-latency connections to Miami, New York and Chicago. As such it is growing faster than almost any other primary market.



New York Tri-State

New York is space-and-power constrained, leading to significant delays in the completion of new facilities. Retro-fits of older facilities are becoming widespread to adapt them to new higher densities.

Secondary Markets

Secondary markets are growing fast in response to constraints in mature markets, and in line with the greater geographic flexibility of Generative AI training site selection. Currently they account for around 20% of total demand, but this proportion seems likely to grow over the next few years.

- **Las Vegas:** with almost 500 MW of capacity Las Vegas is a popular mature hub serving Silicon Valley in particular and will continue to grow.
- **Reno:** with only 176 MW currently but 787 MW of capacity forecast by 2029 (Structure Research) Reno is set to overtake Las Vegas as the leading Nevada hub.
- **Seattle-Quincy:** aerving Portland and Vancouver, with hydro power availability and tax incentives, this mature market cluster is attracting high levels of enterprise and cloud demand.
- **Austin and San Antonio:** combined under-construction activity more than quadrupled here between 2023 and 2024 to 463.5 MW (CBRE).
- **Toronto:** Canada's primary data center hub and the eighth-largest market in North America, is now successfully transitioning from retail colocation to hyperscale and AI requirements.

Close to the Sierra Nevada on the edge of the Great Basin, Reno's altitude, power and land availability plus its proximity to west coast hubs like Silicon Valley and Los Angeles are driving exceptional growth.

- **Hillsboro:** a key crossroads for Asia Pacific and North American data, Hillsboro offers attractive tax incentives, access to subsea cables, and has an extremely healthy construction pipeline.
- **Southern California/LA:** consistent growth is expected here to feed the growing local needs of the Californian digital economy
- **Columbus:** a strong pipeline with inventory forecast to quadruple by the end of the decade.
- **Salt Lake City:** growing exponentially, with some of the lowest power costs in the US due to its plentiful hydroelectric generation plants and proximity to coastal and central regions.

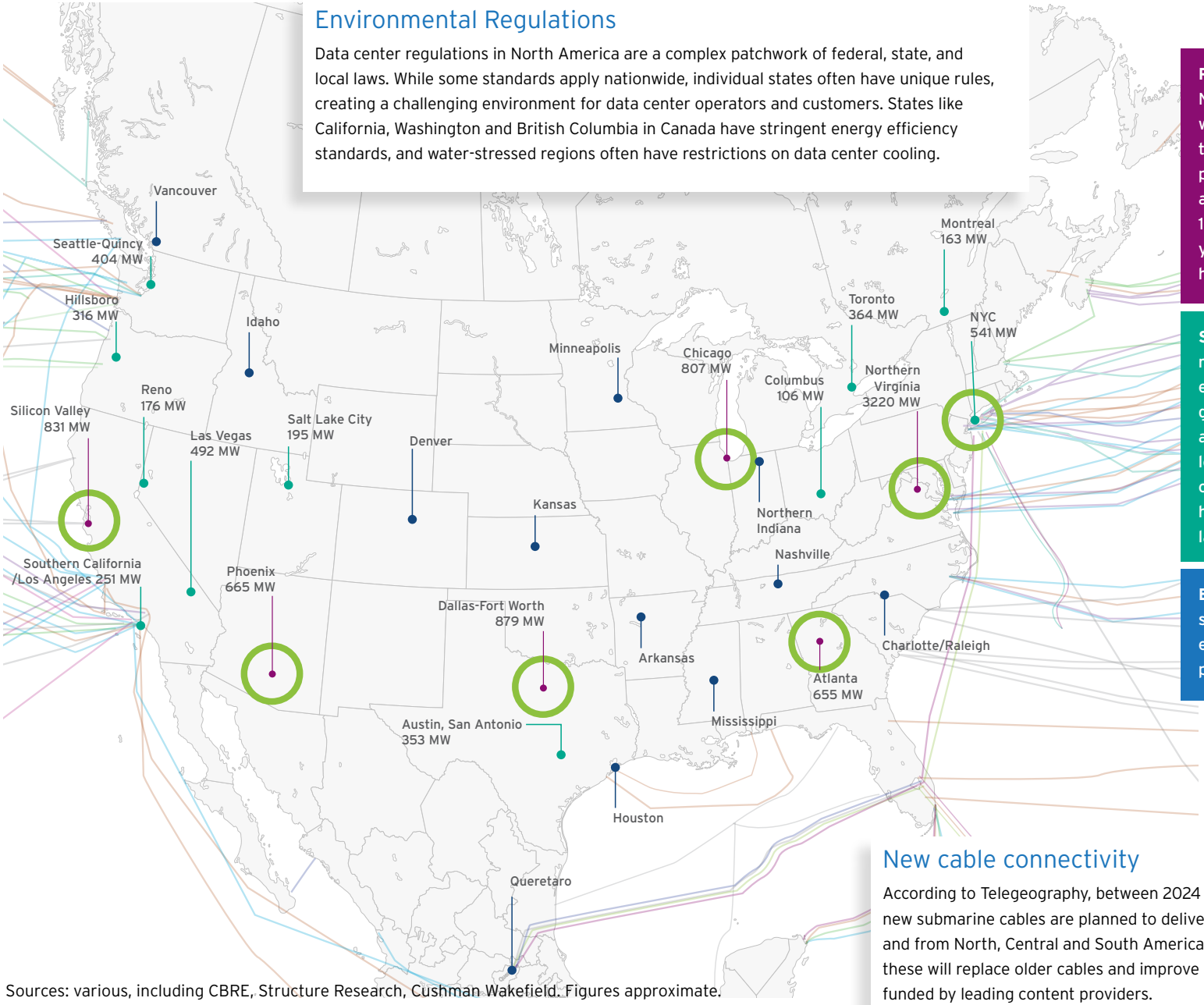
Emerging markets

- Denver
- Charlotte/Raleigh
- Minneapolis
- Montreal
- Houston
- Northern Indiana
- Nashville
- Idaho
- Arkansas
- Vancouver
- Kansas
- Mississippi
- Queretaro

Growth map: key hubs & connectivity

Environmental Regulations

Data center regulations in North America are a complex patchwork of federal, state, and local laws. While some standards apply nationwide, individual states often have unique rules, creating a challenging environment for data center operators and customers. States like California, Washington and British Columbia in Canada have stringent energy efficiency standards, and water-stressed regions often have restrictions on data center cooling.



Primary Markets (over 500 MW built) North America's key data center markets are all world-leading, but Northern Virginia towers above them all. While there are shortages of space and power in key markets this is not dampening demand, and operators are currently achieving growth of over 1 GW a year. All new capacity is being snapped up years in advance and there is a huge pipeline for new high-power-density infrastructure.

Secondary Markets (100 to 500 MW built) Secondary markets account for a growing proportion of overall expansion, compensating for slowdowns. They are generally adjacent to key markets or consolidating around economic hubs. Suitable land, tax incentives, low-carbon power, good connectivity and a cool climate are attractive factors, particularly for high-intensity AI training which is not latency-sensitive.

Emerging (<100 MW built) Accelerated by these same factors, a large number of new metro hubs are emerging. They provide critical diversity for the next phase of market development.

New cloud zones

New cable connectivity

According to Telegeography, between 2024 and 2026 several $billion-worth of new submarine cables are planned to deliver new capacity of up to 1,500 Tbps to and from North, Central and South America. As well as providing new capacity these will replace older cables and improve route diversity, and many are being funded by leading content providers.

Sources: various, including CBRE, Structure Research, Cushman Wakefield. Figures approximate.

Outlook

The North American data center market is growing faster than ever, with a huge amount of new capacity either under construction or in the pipeline. Hyperscale cloud growth, and accelerating service provider and enterprise requirements will continue to support this trend.

With the addition of tailored infrastructure for AI and ML, the market has been boosted to new levels, accelerating construction to the same levels as the fast-growing Asia Pacific market. Structure Research forecasts that overall capacity will grow from a 2024 base of 10.373 GW to 22.58 GW by 2029.

Constraints and diversification

More than doubling inventory in five years will create significant tensions between demand and supply, and the issues the market is experiencing are the flipside of this growth. Space and, particularly, power, are in short supply in mature markets.

While primary markets are still growing fast, these constraints are creating opportunities for new data center hubs, accelerating secondary markets in nearby locations where space and power can be obtained. This trend is likely to increase, as GenAI training is less latency-sensitive than traditional cloud services and IoT and 5G drive edge developments. New primary markets will emerge as part of this process, and secondary markets will become more numerous and more widespread in order to deliver a more diverse set of data processing services.

Green growth

The sharp rise in power consumption driven by power-hungry GPUs will drive design innovation. To keep costs under control and meet emission reduction targets while managing a clean energy transition, efficient components and operations and low-carbon power procurement and storage are becoming key factors for success. New power sources ranging from renewables to nuclear will become non-negotiable requirements for both operators and their end-users, and federal regulations are likely to reinforce these requirements.

Iron Mountain Data Centers in North America

Iron Mountain Data Centers owns and operates a range of highly scalable data center campuses and stand-alone facilities across North America, and is growing fast.

Northern Virginia

We have three hugely scalable data center campuses in Virginia. The most developed is in Manassas in Northern Virginia, offering more than 2,000,000 ft2 of space and 280 MW of power spread across five operational and four planned data centers. A few miles away, we have an interconnected 40 acre site with 150 MW of capacity. 95 miles south in Richmond, we have another recently-acquired 66 acre site with the potential to scale to 200 MW and beyond.

Phoenix (AZP-1, AZP-2, AZP-3)

Just 10 minutes east of downtown Phoenix, Iron Mountain's AZP-1 and AZP-2 offer 89 MW of power and an unbeatable range of interconnection opportunities. A third Phoenix data center, AZP-3, is currently being built.

Scottsdale (AZS-1)

Also in the Phoenix area, AZS-1 offers 121,000 ft^2 of space and 5.8 MW of power with in-house access to a wide variety of carriers and clouds.

Western Pennsylvania (WPA-1)

WPA-1 is located 220 feet underground, using a natural cooling lake to optimize energy efficiency and offering unbeatable levels of security and protection.

Ohio (OHS-1)

OHS-1 is a 44,000 ft^2 facility, strategically placed to serve multiple Midwest markets, including Columbus, Cincinnati and Indianapolis.

New Jersey (NJE-1)

NJE-1 is a TIA-942 certified purpose-built 830,000 ft^2 / 30 MW colocation facility on a 43-acre campus located a short drive southwest of NYC, with a huge rooftop solar array.

Miami (MIA-1)

Located on a 3.4-acre campus in Central North West Miami our 150,000 ft^2, 16 MW MIA-1 data center will soon be ready for service.

Kansas City (KCM-1)

Constructed underground, Iron Mountain KCM-1 is a 50,000 ft^2 facility with 3.9 MW of power and exceptional reliability and efficiency.

Denver (DEN-1)

DEN-1 is a purpose-built 180,000 ft^2 / 14.4 MW colocation facility on a 6-acre campus located just 10 minutes from downtown Denver in one of the lowest-risk US metros for natural disasters.

Chicago (CHI-1)

Located in the Chicago suburb of Des Plaines, Illinois, our Chicago data center will offer a mix of retail colocation and hyperscale data center solutions. At full build, CHI-1 will offer 36 MW of IT capacity, with the first 8 MW coming online in 2025

Boston (BOS-1)

Iron Mountain BOS-1 is a LEED-certified 21,000 ft^2 facility with 3.6 MW of power and significant potential to scale up on our 58-acre campus.

Sources

Market Forecasts & Trends

https://www.structureresearch.net (subscription)

https://www.cbre.com (North America)

https://www.datacenterknowledge.com

https://www.precedenceresearch.com

https://www.cbre.com (global)

https://www.cushmanwakefield.com

Hyperscale & AI Trends

https://thetechcapital.com

https://www.precedenceresearch.com

Edge

https://www.credenceresearch.com

Renewables/Low Carbon Energy

https://www.reuters.com

https://www.datacenterfrontier.com

https://www.spglobal.com

Power Pressure & Pricing

https://gridstrategiesllc.com

Connectivity

https://blog.telegeography.com

Infrastructure Planning Report

North America

Iron Mountain Data Centers operates a global colocation platform that enables customers to build tailored, sustainable, carrier and cloud-neutral data solutions. As a proud part of Iron Mountain Inc., a world leader in the secure management of data and assets trusted by 90% of the Fortune 1000, we are uniquely positioned to protect, connect and activate high-value customer data. We lead the data center industry in highly regulated compliance, environmental sustainability, physical security and business continuity. We collaborate with our 1,300+ customers in order to build and support their long-term digital transformations across our global footprint, which spans three continents.
IRONMOUNTAIN.COM/DATA-CENTERS

To discuss your plans and requirements please contact us on:

US: 833.IRM.colo
NL: +31 800 272 4433
UK: +44 844 417 8379
DE: +49 800 408 0000

Email: datacenters@ironmountain.com

IRON MOUNTAIN®
DATA CENTERS