COLOCATION & CLOUD

6-12 REPORT

Knowledge is power, as the saying goes, and data is essential to knowing how to run a business efficiently and profitably. But how much electrical power is needed to run a business in the age of artificial intelligence?

That is one of the key questions enterprises will confront in 2024 and beyond. For growing companies navigating the constantly evolving technology landscape, scalable IT infrastructure is key. A flexible tech stack allows organizations to stay agile and can also help create an improved customer experience, optimize resources and processes, and accommodate new or evolving solutions. Due to this focus on scalability, the adoption of cloud services has notably increased as solutions became more easily integrated into existing IT environments. However, within the last two years, questions about performance, security, and cost of these cloud services have led some organizations to reconsider the choice in adopting.

AVANT and Trusted Advisors are seeing increased use of colocation services in third-party data centers to address these issues. It's not just Fortune 500 companies that are using colocation, either — small- and medium-sized companies in traditionally conservative industries such as regional banks and credit unions are also leveraging these services. In the healthcare vertical, Trusted Advisors are seeing everything from multi-state organizations to clinics and specialty practices now buying colocation services, whereas previously, these entities relied entirely on equipment located on-premises. Across all industries, power-hungry AI applications are only going to increase the need for colocation services over on-premises data centers.

About AVANT

AVANT is a platform for IT decision-making and the nation's premier distributor for next-generation technologies. AVANT provides unique value to its extensive network of Trusted Advisors with original research, channel sales assistance, training, and tools to guide decision-making around IT services that promote business growth. From complex cloud designs to global wide-area network deployments to the latest in security services, AVANT sets the industry standard in enabling its partners and clients to make intelligent, data-driven decisions about services, technology, and cost-effective communications. For more information, visit **www.goavant.net**, or connect on Twitter and LinkedIn.

Cloud and Colocation: By the Numbers

AVANT is among the largest distributors of colocation and managed cloud service providers in the United States through an ecosystem of thousands of Trusted Advisors supporting customers with a platform for IT decision making. As such, AVANT is in a unique position to understand the patterns, behavior, and tendencies of the participants in the internet infrastructure landscape. Through the normal course of business, they regularly collect end users' purchase needs and decisions for a wide range of communications and infrastructure services through their Interactive Quick Assessments (IQAs) process.

Taking a deeper look into the colocation and cloud market in 2024 and beyond, AVANT partnered with **Structure Research**, an independent research and consulting firm, to analyze IQA data for the five-year period from 2018 to 2023. Structure Research, which focuses on the internet infrastructure market's cloud and data center segments, also provided insights based on its own global dataset on the data center and colocation market.

Four Macro Trends Driving Colocation Growth

In the past decade plus, the question on the minds of many IT professionals and their C-Suite executives was "What is our cloud strategy?" Digital transformation plans included roadmaps for migrating on-premises applications to rapidly growing cloud providers. In 2020, the COVID-19 pandemic further supported this trend as the work force practiced social distancing and access to physical servers in data centers became more restricted. In the past year, AVANT saw several new trends emerge:

1. Repatriation
2. Post-COVID corporate real estate changes
3. AI's impact on colocation space and power inventory
4. A demand for increased connectivity requirements

Trend #1: Repatriation — To the Public Cloud… and back?

In the last few years, some organizations that shifted to the cloud have begun to reverse their decision. "Customers came to us and said that moving everything to the cloud did not meet expectations," said Chip Hoisington, AVANT Senior Director of connectivity, colocation, and wireless.

According to Hoisington, part of the reason for this sentiment was the high cost that organizations incurred as a part of the transition to the cloud, but another was application performance and availability issues. "Enterprise customers saw that some of their database, storage applications, and services weren't optimized for the cloud, causing performance issues," he said. Some AI workloads and other applications in the public cloud can also cost a lot more for customers. According to 451 Research, 48% of businesses have migrated workloads out of AWS, MS Azure and Google (GCP); and 96% of those surveyed cited cost efficiency as the main reason for repatriation.

Location of IT Infrastructure

The data in Figure 1 was collected by AVANT from IT professionals seeking to purchase colocation space. In 2023, customers seeking to purchase colocation had the highest percentage of their infrastructure currently in the cloud at 16%, compared to the previous high of 14.5% in 2021. These customers also reported that 58.7% of infrastructure was located on-premises. It is interesting to note that the companies seeking colocation space in 2023 also had the lowest current percentage (22.1%) of their infrastructure currently in colocation facilities, showing a decline from 36.9% in 2020. Customers who currently have the lowest historical percentage of current workloads in colocation and the highest historical percentage of workloads in on-premises and cloud are seeking to purchase colocation.

FIGURE 1

Current Location of IT Infrastructure for Orgs Seeking Colocation

Source: AVANT IQA data
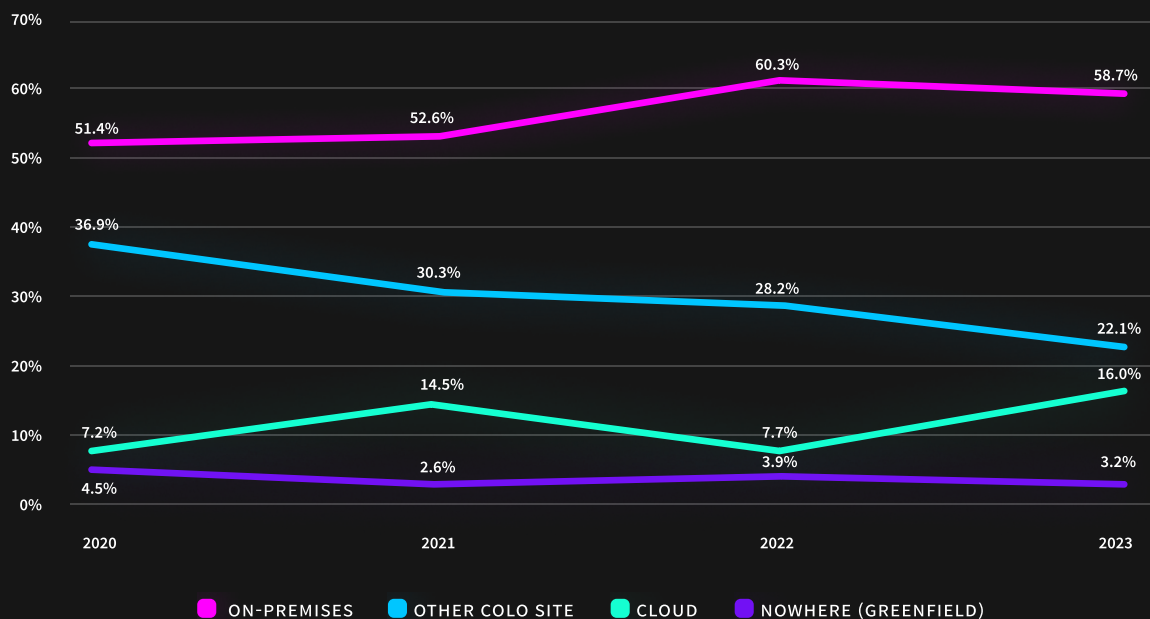
Trend #2: Post-COVID Corporate Real Estate

The second major post-COVID trend was driven in part by the macro changes in utilization of corporate office space and the trend toward a remote workforce. As corporations look to reevaluate their corporate office space needs and cancel leases, existing on-premises IT infrastructure needs to be relocated. With that said, we are seeing many organizations move their IT infrastructure equipment to colocation facilities.

Over the last five years, survey data (Fig. 2) reveals that the **average** amount of infrastructure located on-premises is 55%. This indicates a significant amount of digital infrastructure lingering on site despite the availability of public cloud for well over a decade.

FIGURE 2

Location of IT Infrastructure (5-Year Avg.)

Source: AVANT IQA data

13%

3%

29%

5-Year
A V E R A G E

- ON-PREMISE
- OTHER COLO SITE
- CLOUD
- NOWHERE (GREENFIELD)

55%
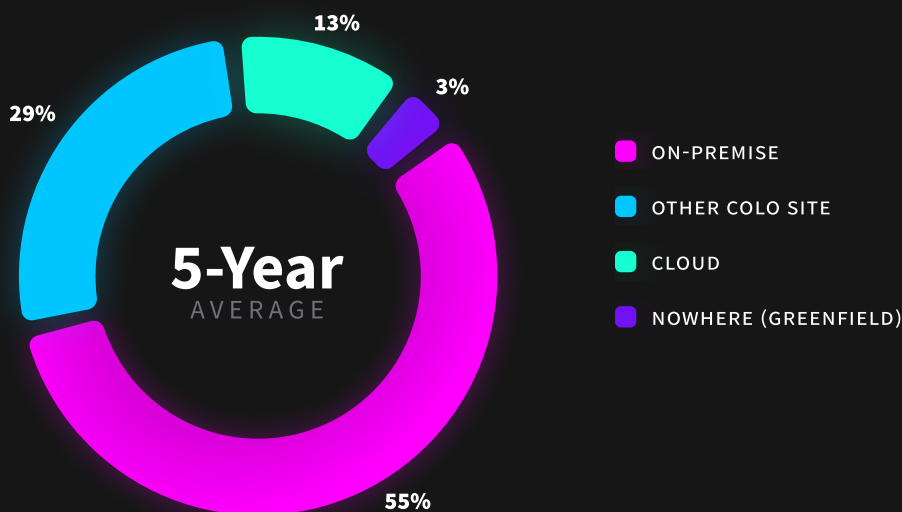
The shift towards the cloud is incremental and still in motion, according to Philbert Shih, Managing Director of Structure Research. Meanwhile, colocation remains a viable and popular choice for off-premises deployments, highlighting its continued relevance in the market. Almost 30% of organizations opted for colocation as a preferred off-premises deployment strategy, according to AVANT data.

In the end, however, on-premises infrastructure has still been the prevalent choice for many organizations. The AVANT data aligns with the Structure Research data. The difference between these 2 data sets is that the AVANT data is looking at customers who are currently looking to move their infrastructure to colocation, while Structured Research data is based on general IT deployments with the enterprise market at about 47% on-premises and 53% outsourced off-premises.
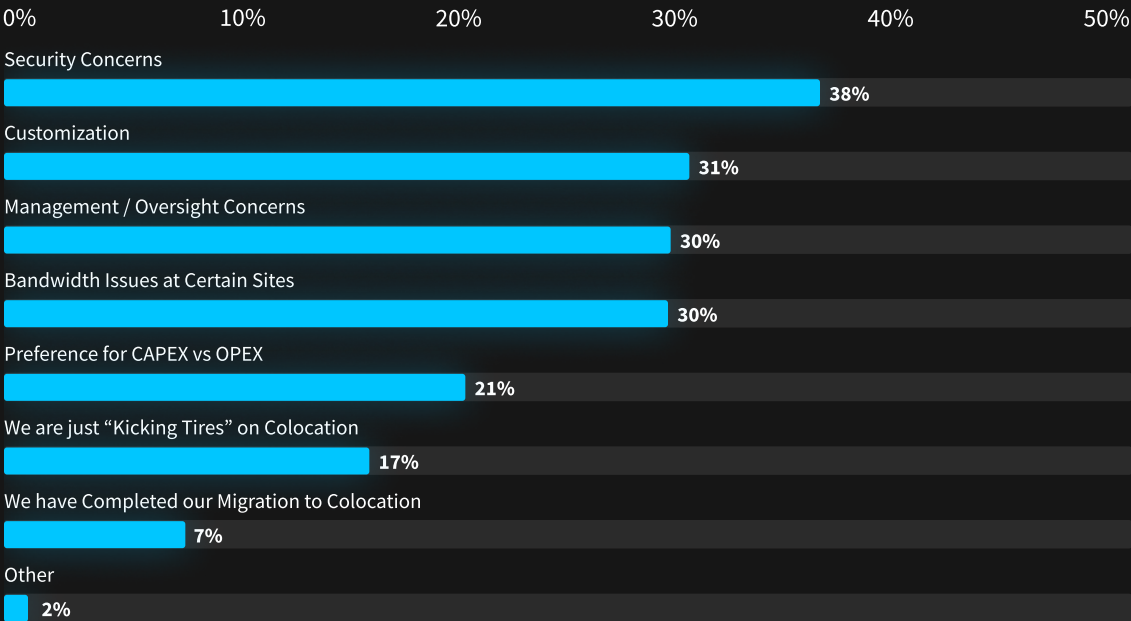
It is not entirely surprising that some organizations continue to persist in their preference for on-premises deployment, according to Shih. Some of the preference for on-premises infrastructure is about organizational inertia; simply put, it takes planning and engineering effort to refactor applications for the cloud. On the other hand, the desire for flexibility and protection from threats such as ransomware are driving enterprises to a hybrid cloud strategy. A survey of 1,324 enterprise IT decision-makers from research firm IDC suggests 50% of applications were on-premises in 2022 (including both private cloud and non-cloud infrastructure), but that figure will decline to 35% by 2024 as organizations move more applications to hosted private cloud and IaaS offerings.

AVANT's survey data from the 2024 State of Disruption Report also explored why companies have not yet moved their workloads from on-premises to colocation facilities. In Figure 3, 38% of respondents cite security concerns as the top reason. Only 7% said that they have completed all their migration plans.

FIGURE 3

Reason Some Companies Have Not Migrated to Colocation Services

Source: AVANT 2024 State of Disruption

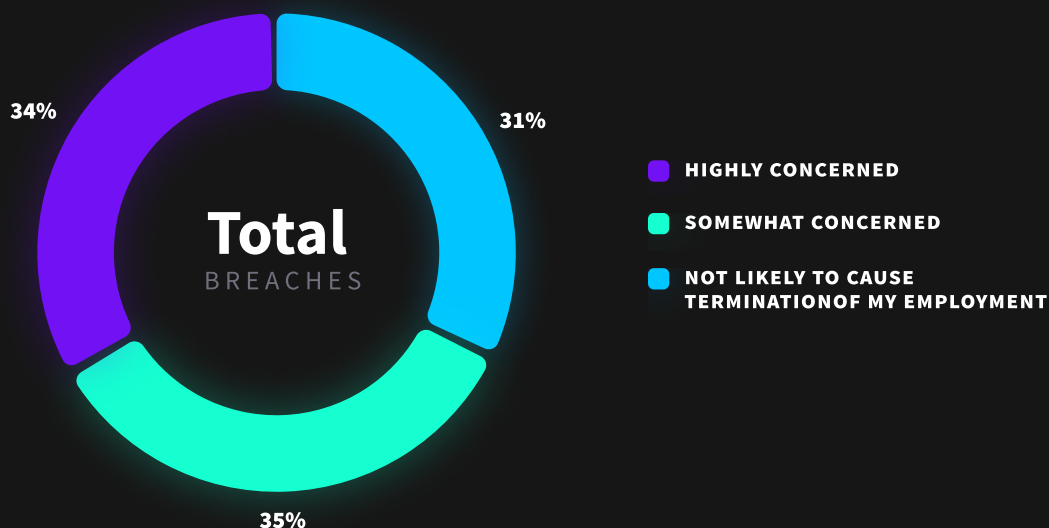|		
Security Concerns		38%
Customization		31%
Management / Oversight Concerns		30%
Bandwidth Issues at Certain Sites		30%
Preference for CAPEX vs OPEX		21%
We are just "Kicking Tires" on Colocation		17%
We have Completed our Migration to Colocation		7%
Other		2%

The security concerns are tied to an underlying motivation: fear. Over two-thirds of IT decision-makers are concerned that their employment might be at risk in the event of a cyber breach (Fig. 4).

FIGURE 4

Breaches and Perceived Risk to Employment

Source: AVANT Analytics State of Disruption 2024

34%

31%

35%

Total
BREACHES

- HIGHLY CONCERNED
- SOMEWHAT CONCERNED
- NOT LIKELY TO CAUSE TERMINATIONOF MY EMPLOYMENT

Trend #3: AI's Increasing Consumption of Colocation and Power Inventory

Enterprises can "repatriate" some servers back to a data center, but the growth of other workloads will make this harder, if not impossible, to handle. Specifically, AI workloads are driving up power density requirements. Historically speaking, enterprise customers have a requirement of 5kW per cabinet, Hoisington noted. Yet in 2023 and 2024, some customers were requesting three to four times as much power per cabinet because of the increasing use of GPUs (Graphics Processing Units), and that trend is likely to spread and grow as more enterprises embrace AI, he predicts.

FIGURE 5

2023 Total Power Required for Colocation Deployments (in kW)

Source: AVANT Analytics
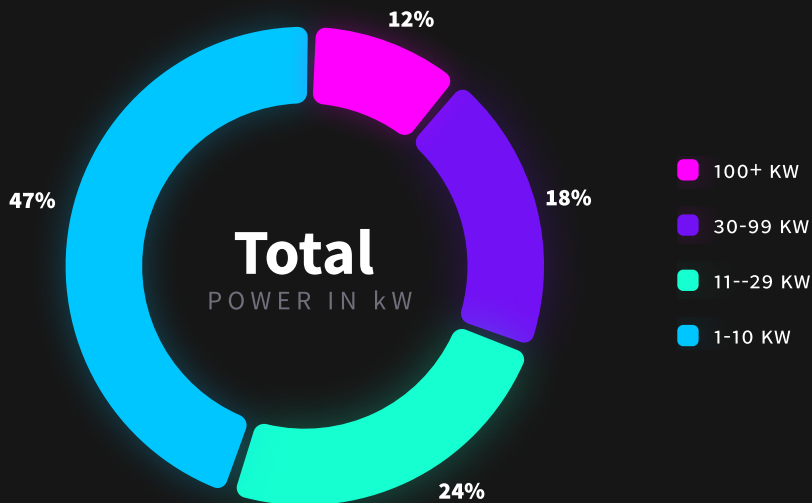
Retail colocation power requirements remain relatively small, as revealed by AVANT's analytics. Data from 2023 shows that 47% of respondents are deploying under a total of 10kW of power. Another 24% use between 11 and 29kW, while 17.4% need between 30 to 99kW. Total power requirements of over 100kW have grown in 2023 to 12% of customer requests.

Most deployments still consist of single-digit cabinet counts, with low power consumption and densities. Until 2023, growth in power requirements for a typical retail colocation deployment was not climbing quickly. The data suggests that the types of workloads sitting in retail colocation are production-grade and steady-state, but with an increasingly legacy profile, according to Structure Research's Shih.

AVANT data shows a significant majority of retail colocation workloads come in at a density of less than 5kW per cabinet, generally falling between 59% to 67% from 2020 to 2023. There is some growth observed in rack densities in the 6-10kW and 11-15kW per rack range. Shih notes that rack densities above these power ranges were rare in the retail colocation space in 2022, but they are starting to come on the radar with the rise of AI in 2023 and beyond.

One of the key reasons for the growth in rack densities above 15kW is the power consumed by GPUs. A single Nvidia H100 GPU can consume up to 300 watts of power, and a complete Nvidia DGX H100 system has an 8U rackmount form factor and a maximum power usage of 10.2kW.

Historically speaking, enterprise customers require 5kW per cabinet, AVANT's Hoisington notes. In 2022, some larger customers were starting to deploy on the order of 8.6kW, and by 2023, customers were requesting 15 to 20kW per cabinet because of the increasing use of GPUs. While technology and digital media companies are driving some of this demand, sourcing this kind of power requirement is leading a growing number of mid-market companies to start moving to colocation facilities, Hoisington said.

"What we're seeing by mid-2023 are load requests in the 20 to 30kW range," notes Hoisington, with these types of requests increasingly becoming the norm rather than the exception. The general impact this has on the market is an increasing consumption of available power capacity within data centers. Customers need to be aware of this trend even if they are not utilizing power hungry Nvidia platforms, as it will impact resource availability and increase pricing in some local colocation markets.

Trend #4: Increased Connectivity Needs

As organizations started using the cloud to scale during the pandemic, new connectivity to cloud requirements, such as direct connects to AWS, Azure, or Google, also surged to a peak of 42% in 2021, according to AVANT IQA data. This is due to the need to ensure reliable access to cloud resources and applications as part of the scaling efforts. Other cloud connectivity benefits include:

- **Security:** A dedicated and closed network helps avoid vulnerabilities associated with traffic that traverses the public internet.
- **Performance:** Bypassing the congested links and exchange points to the public internet can result in reduced packet loss and lower latency.
- **Cost Savings:** Traffic back to applications and end users that is sent over a direct link is usually either at no cost or lower cost than egress to the public internet.

Overall, IT decision-makers seeing cloud connectivity as a priority item have climbed in the last few years. The flexibility offered by cloud solutions to scale as demand increases — or decreases — applies to networks as well. In 2023, as enterprises settled on their hybrid cloud architecture and repatriated some workloads, the requirement for direct connections slid back to 27.1%, compared to 31.6% the year prior. Those who were "not sure" if they required a direct connection increased to 45.8%, pointing to an area for Trusted Advisors to step in and provide guidance.

FIGURE 6

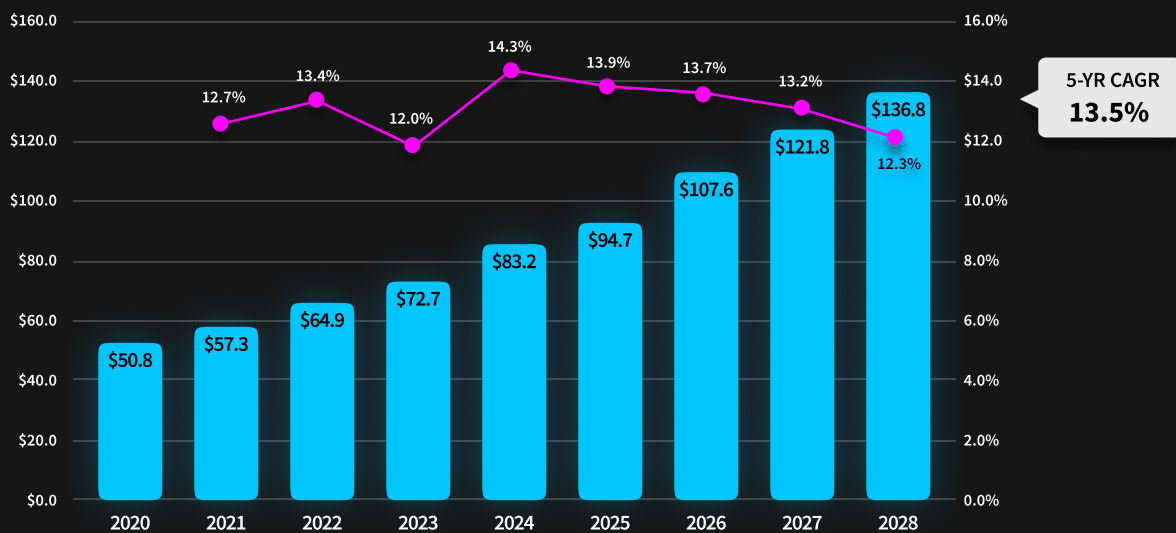## Do You Require a Direct Connection to Your Cloud Provider?

Source: AVANT Analytics
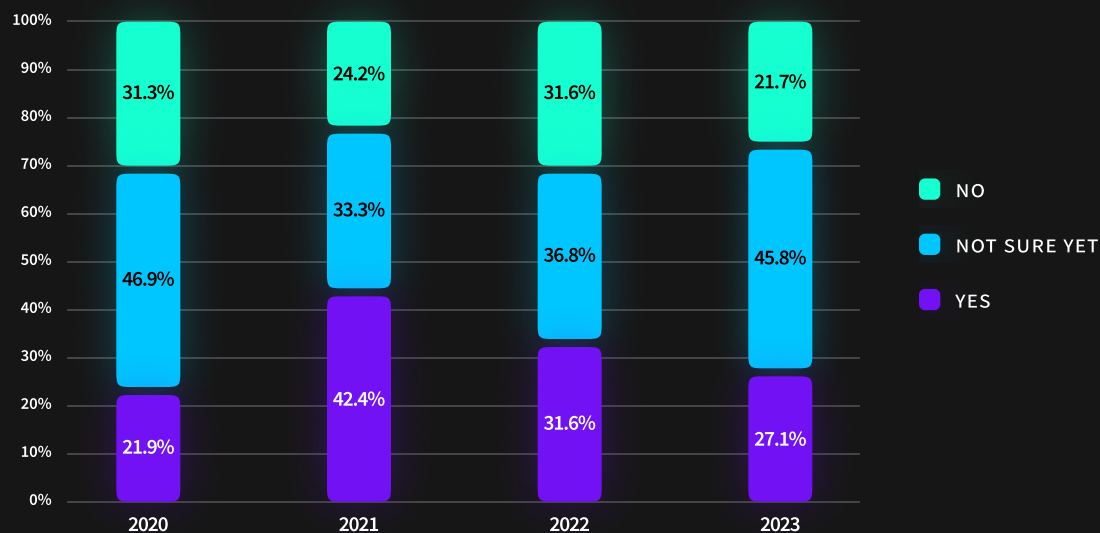
Global Market Data

According to Structure Research, the global data center colocation market reached $72.7 billion in revenue in 2023 and is expected to reach $83.2 billion in 2024 (Fig. 7).

FIGURE 7

Global Data Center Colocation Market Size (in $ Billions)
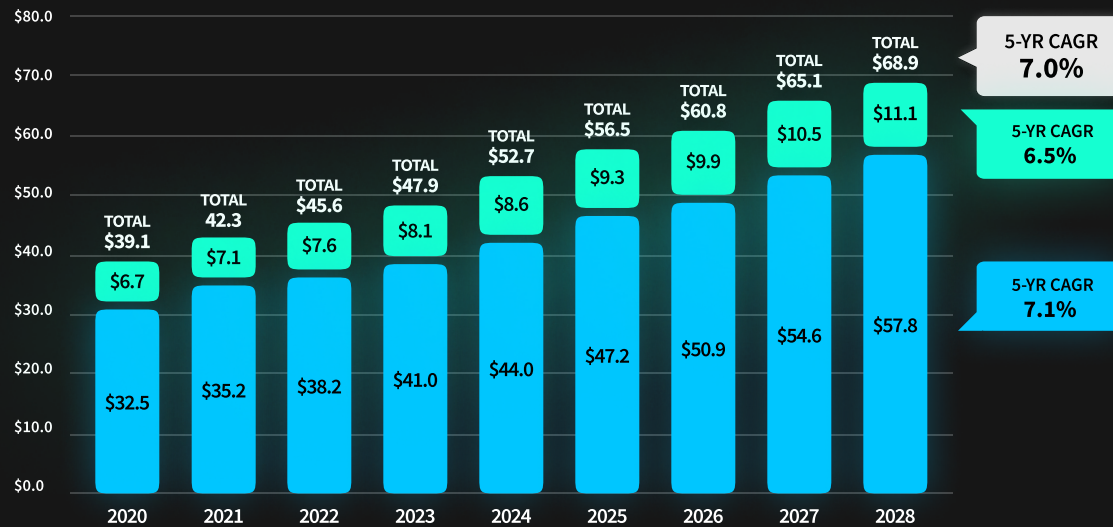
Source: Structure Research

Over the five-year forecast period, the compound annual growth rate is expected to be 13.5%. This growth number includes all facility types.

Looking at the global enterprise colocation market in Figure 8 provides further insight into colocation demand.

FIGURE 8

Global Enterprise Colocation Market Size (in $ Billions)

Source: Structure Research

Structure Research defines the "enterprise" and "enterprise plus" categories for the data center colocation market based on the scale of the colocation requirements. The "enterprise" category typically refers to the smaller-scale colocation needs of individual enterprises, while the "enterprise plus" category encompasses larger-scale colocation requirements, often associated with hyperscale cloud providers.

Global revenue from the enterprise colocation segment reached $8.1 billion in 2023 and is forecast to top $8.6 billion in 2024, with a five-year growth rate of 7.1%.
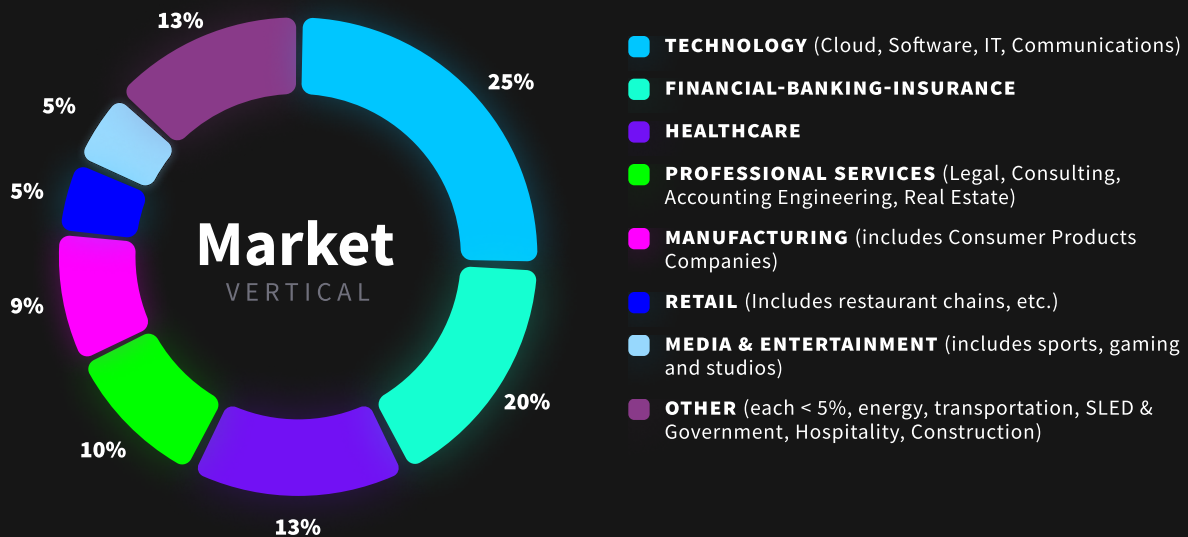
Choosing a Solution

Several industries are increasingly opting for colocation services. A review of AVANT data from 2023 shows growth based on the sales of colocation services to **new** customers in different industries:

- **Technology (Cloud, Software, IT, Communications):** The technology sector accounted for 25% of new customer sales and led in colocation service deals driven by AI services and chip companies.
- **Financial/Banking/Insurance:** Following closely behind, financial services represented 20% of colocation services sold in 2023.
- **Healthcare:** With a 13% share, the healthcare industry was also a significant player in the colocation services market.

FIGURE 9

AVANT Colocation Service Orders in 2023, by Market Vertical

Source: AVANT Analytics

- **TECHNOLOGY** (Cloud, Software, IT, Communications)
- **FINANCIAL-BANKING-INSURANCE**
- **HEALTHCARE**
- **PROFESSIONAL SERVICES** (Legal, Consulting, Accounting Engineering, Real Estate)
- **MANUFACTURING** (includes Consumer Products Companies)
- **RETAIL** (Includes restaurant chains, etc.)
- **MEDIA & ENTERTAINMENT** (includes sports, gaming and studios)
- **OTHER** (each < 5%, energy, transportation, SLED & Government, Hospitality, Construction)

Outside of the technology vertical, AVANT and Trusted Advisors saw tremendous activity from banks, credit unions, investment companies, mortgage companies, and insurance companies in the financial vertical. For the healthcare industry, high activity was seen not only in the hospital systems but also in many multi-site healthcare organizations in the clinical, small practice, and specialty medical space.

Two common themes played a role in the sales to the financial and healthcare verticals, according to Hoisington:

- Compliance and security drove a need for space in monitored facilities.
- Where employees work is resulting in changes in the corporate real estate footprint. Fewer people in the office means less office space; therefore, IT equipment can be relocated to a less expensive colocation facility nearby.

In the Media and Entertainment segment, new customers in online gaming and online betting were also busy expanding their colocation footprint.

Choosing Colo Services: The Need for Power and Resilience

"Utilities cannot keep up with power demand in big cities. We can expect to see more brownouts, especially in periods of peak demand, such as the heat waves that took place in the summer of 2023," notes Hoisington.

An example of the issues this could cause can be seen in an outage one CDN (Content Delivery Network) experienced last year. A preliminary analysis suggested that an 8-hour service interruption started when a line test by the utility caused a ground fault that took the data center generators offline. In fact, all customers at the facility experienced service interruption because UPS batteries ran out of power due to the generators not being fully restarted. Additionally, the electric breakers connecting to the CDN vendor's equipment were found faulty when the generators were restored.

Mid-market and larger enterprises must have a second facility that is outside of a particular power grid for disaster recovery, AVANT's Hoisington advises. On the East Coast, that means at least going to Pittsburgh or Ohio; on the West Coast, it is advisable to put a second data center east of the San Andreas fault, which typically means Denver or Las Vegas. Customers in the Midwest will want to avoid having a second facility in hurricane- or tornado-prone locations such as southern Texas.

Conclusion

This evolving technology landscape underscores the necessity for scalable and flexible IT infrastructure, but also highlights the diversity of solutions available to organizations. While cloud services offer remarkable benefits in terms of accessibility and integration, the recent concerns regarding performance, security, and cost have prompted a strategic shift towards colocation services.

AVANT data shows increased adoption of colocation services across various sectors, from regional banks and credit unions to healthcare providers, demonstrating that colocation is no longer the exclusive domain of large enterprises. As businesses strive to enhance efficiency and agility, understanding the value and potential of both cloud and colocation solutions will be pivotal in ensuring that organizations can harness the full potential of their data and technology investments.

Checklist

When considering colocation service providers for IT needs, enterprises should conduct thorough due diligence to ensure they select a provider that aligns with their specific requirements and objectives. The following detailed checklist encompasses critical factors that these organizations should evaluate:

POWER AND COOLING

- Have a clear understanding of power and cooling requirements, both current and projected. (Stated in kW load per cabinet. For example, 5kW, 8.6kW, 17kW+)

- Assess the provider's ability to deliver the required power – including support for high-density workloads like AI. Are there significant surcharges for those capabilities?

- What is the backup power delivery system and is there N+1 or 2N redundancy? What are the associated procedures for testing?

- What uptime guarantees are offered?

- Consider the overall efficiency of the facility (PUE rating) and cooling systems. Is there an option for liquid cooling (such as rear door heat exchange, direct to chip, and submersion) for high-density workloads?

NETWORK ACCESS

- Ensure the provider is carrier-neutral and offers connectivity to a variety of network carriers. This will help with finding competitive pricing for networking services.

- Evaluate the provider's interconnect options for diversity. For maximum resilience, are there multiple ingress/egress options for the facility?

- How much bandwidth for internet connectivity is needed?

- How many cross-connects are needed?

- Is cloud connectivity needed? (AWS Direct Connect, Azure ExpressRoute, Google Cloud Interconnect, Megaport, etc.)

- Are Layer 2 point-to-point private lines needed?

LOCATION

- Decide on the location of the colocation site based on strategic needs, such as disaster recovery or market presence.

- Consider site accessibility under adverse conditions (weather, traffic, etc.) and the potential need for remote management.

CONTRACT AND SLA FLEXIBILITY

- Review the provider's primary service agreement and SLA with legal, business, and IT teams to address mission-critical items such as availability and incident response time.

SECURITY

- Physical Security Measures: Does the facility have 24/7 on-site security personnel, biometric access controls, and video surveillance?
- What are the provider's cybersecurity measures, including firewalls and intrusion detection systems?
- Compliance and Certifications: Ensure the provider complies with relevant standards. (e.g., SSAE 18, ISO 27001, HIPAA, PCI DSS)

SERVICES

- Evaluate the provider's local staff availability and the range of managed services offered, including remote hands for equipment management when local IT staff are unavailable.
- Carefully review the SLAs to understand the scope of support services and response times.

COMPLIANCE

- Verify the provider's compliance with relevant industry-recognized data center certifications. Ensure third-party audits are allowed for validation.
- Verify the provider's disaster recovery and business continuity plans.

FOR CONSIDERATION: SUSTAINABILITY

- Check if the provider uses renewable energy sources and their percentage of total energy use.
- Sustainability Certifications: Look for providers with sustainability certifications such as LEED or Energy Star.
- Inquire about the provider's initiatives to reduce carbon emissions and achieve carbon neutrality.

When it comes to buying colocation services, several key job roles or titles are typically involved in the decision-making and procurement process. These roles often span different departments within an organization, reflecting the multi-disciplinary nature of IT infrastructure management. In addition to the final say from C-level executives, here are some of the typical job roles:

IT AND FACILITIES MANAGEMENT
- IT Director
- IT Infrastructure Manager
- IT Project Manager
- Data Center Manager
- Data Center Services Procurement Manager
- Facilities Director/Manager

NETWORK
- Network Architect
- Network Engineer
- Network Administrator

SECURITY
- Chief Security Officer (CSO)
- IT Security Manager

SUSTAINABILITY
- Chief Sustainability Officer (CSO)
- Sustainability Manager

Acknowledgements