



DATA CENTER

Frontier Special Report

How Colocation Can Be the Nervous System of Digital Business

Written by Paul Gillin



brought to you by



Contents

Introduction.....	2
Real-time business.....	3
The colocation solution	3
Colocation myths	4
The business drivers of colocation	5
The technical drivers of colocation	6
Understanding interconnection	7
Types of interconnections	7
Why use interconnection services from a colocation provider?.....	8

Introduction

How much does your company spend on data communications each year? Chances are it's a lot and growing. [Gartner](#) estimates that spending on communication services will make up 35% of the average IT budget this year.

For organizations that are geographically distributed or moving operations to the cloud, the network is increasingly the lifeblood of the business. And as more business moves online, latency is not just unacceptable, it's a threat to the health of the business.

Cloud computing is driving a new focus on network bandwidth and quality as processing expands beyond the on-premises data center. Cisco [predicts](#) that 94 percent of enterprise workloads will be processed by cloud data centers by 2021. A [survey](#) of more than 300 IT professionals by Enterprise Management Associates found that 84% expect their cloud traffic to grow over the next three years.

And customers are using more clouds than ever. Rather than settling on a single provider, they use a combination of platforms, software-as-a-service (SaaS) providers, colocation facilities and managed services. Rightscale has [reported](#) that 81% of enterprises are using multiple clouds and 45% are using hybrid cloud. Mordor Intelligence [valued](#) the global hybrid cloud market at \$40.6 billion in 2017, growing 22% annually to nearly \$139 billion by 2023.

A survey of more than 300 IT professionals by Enterprise Management Associates found that 84% expect their cloud traffic to grow over the next three years.

At the same time, the distributed processing phenomenon called edge computing is creating new network demand at the corners of the organization. International Data Corp. projects that [the world will create 175 zettabytes of data](#) in 2025, or about 10 times as much as in 2016. The research firm also expects that one-quarter of that data will be created in real time, with “internet of things” (IoT) devices contributing 95% of that volume. Data growth will challenge organizations’ ability to maintain adequate performance, even as customers demand more responsive online experiences.

Real-time business

The pace of business no longer tolerates delay.

- ▶ The [\\$384 billion](#) digital advertising market is based on automated split-second decisions about where to place promotions based upon auctions that take place in near-real-time.
- ▶ Ride-sharing services use stream processing to monitor the precise location of millions of vehicles simultaneously.
- ▶ Telecommunications providers keep constant track of the status of millions of customers and share relevant data instantaneously with business partners.
- ▶ Transactions of all kinds are increasingly taking place on mobile apps, where a few seconds of delay can drive a customer to a competitor.
- ▶ In the financial services industry, the difference between profit and loss on a trade can be measured in milliseconds. Firms that engage in high-frequency trading go so far as to locate their computers and facilities physically adjacent to stock exchanges to gain a few microseconds of performance advantage on their competitors.

For companies that depend upon high-speed communications to conduct business, the public Internet is no longer an option.

Sending large volumes of time-sensitive data over standard IP connections is becoming increasingly impractical. For companies that depend upon high-speed communications to conduct business, the public Internet is no longer an option. High-speed leased lines have historically been the best alternative, but those connections are expensive and too fixed for the constantly changing needs of digital businesses who need the flexibility to connect to multiple clouds and business partners instantaneously and as needed without long and costly setup times.

The colocation solution

Colocation facilities can be valuable partners in reducing IT costs, improving flexibility and creating robust, highly available infrastructure that connects seamlessly to multiple clouds. The colocation

The interconnection market is growing even faster than the colocation market, with revenues expected to more than double from [\\$3.48 billion in 2019](#) to reach [\\$7.65 billion in 2025](#).

concept emerged in the 1990s as businesses flocked to the Internet looking for robust, reliable and secure places to house their computing infrastructure. Colocation facilities provided not only rapid setup in professionally managed data centers but also reliable on-ramps to the Internet through shared high-speed pipes.

With the rise of cloud computing a decade ago, many organizations began to divest themselves of servers and storage entirely as they shifted infrastructure responsibility to cloud providers. Some people said cloud was a threat to the colocation industry because hyperscale computing would obviate the need for businesses to own equipment. Yet ironically, cloud is fueling a rebirth and a resurgence of colocation in a new form.

The global colocation market is expected to reach [\\$90 billion in 2024](#), up from [\\$35 billion in 2017](#), [according to](#) Global Market Insights. A principal driver of this growth is interconnection, or private data exchange between businesses. Interconnection is best done in a vendor-neutral data center where a large ecosystem of vendors, customers and cloud providers are clustered together. The interconnection market is growing even faster than the colocation market, with revenues expected to more than double from [\\$3.48 billion in 2019](#) to reach [\\$7.65 billion in 2025](#), [according to Mordor Intelligence](#).

Colocation myths

Because colocation has evolved rapidly in a constantly shifting landscape, many myths have developed about the quality and reliability of colocation services as well as the level of control customers retain. But the reality is that cloud infrastructure and colocation not only have a lot in common but complement each other.

There are five common misperceptions about colocation in particular.

PERCEPTION #1

The customer has little or no control over infrastructure.

Reality: Customers own the equipment they place in a colocation center and have access to it 24 hours a day. The provider is responsible for providing power, network and cooling infrastructure and often offers availability service level agreements for customers who value that peace of mind. The provider doesn't manage customer workloads and has zero visibility into the programs they run. While the responsibility is on the customer to maintain software and balance workloads, the same would be true in an on-premises data center. Customers in regulated industries, in particular, value having that level of security and control. In the meantime, sharing the costs of power, floor space and common infrastructure such as backup provides economies of scale that save customers money. Small companies can enjoy the same robust data center features as giant enterprises.

PERCEPTION #2

Equipment is less secure when moved off site.

Fact: The opposite is true; only the largest enterprises come close to matching the security features of colocation facilities. Physical security controls typically include security cameras, biometric authentication, double door "mantrap" entry and locked cages and cabinets. Employees at the leading colocation providers are trained in best-in-class security procedures, must pass certification exams and receive rigorous ongoing training. Security systems and protocols are tested regularly, and facilities comply with leading standards such as ISO 27001, SOC HIPAA and PCI DSS. Maintaining such a high level of security is all but unaffordable for most customers.

PERCEPTION #3

There are no availability guarantees in a colocation arrangement.

Fact: Any reputable colocation provider offers service level agreements that guarantee availability levels and provide penalties if they aren't met. But fulfilling SLAs is rarely a problem. Leading colocation facilities employ a full suite of high-availability technology like redundant power supplies, backup generators, fire suppression systems, disaster recovery facilities and highly available multi-region cloud architectures. Few customers can afford to install and maintain such complete protections on their own premises. This enables CoreSite, for example, to offer a 100% uptime SLA.

PERCEPTION #4

Colocation extracts a performance penalty because equipment is offsite.

Fact: Performance is usually better than that of equipment housed on-premises because colocation providers use state-of-the-art server racks and cooling equipment. The performance benefit is particularly pronounced when cloud connectivity is required because colocation centers maintain high-speed and dedicated connections to cloud on-ramps that are much faster than public Internet connections.

PERCEPTION #5

Colocation is expensive.

Fact: Most customers save money compared to buying and maintaining their equipment on premises. That doesn't include the intangible benefits of having access to faster bandwidth, better security controls, superior data protection, experienced staff and other byproducts of the economics of the multi-tenant model.

The business drivers of colocation

Businesses increasingly depend upon secure, reliable availability and high-speed connectivity as they pursue the goal of digital transformation. Downtime is no longer a luxury. [One recent study](#) found that the cost of an hour of downtime exceeds \$300,000 in most cases and can run over \$5 million for some businesses. Poor performance also carries a penalty. [Another study reported](#) that a one-second delay in webpage loading time results in 11% fewer page views, a 16% drop in customer satisfaction ratings and 7% fewer conversions. Customers said they expect a webpage to load within two seconds on average before they consider abandoning a site.

IT is also no longer exclusively a back-office function. Many companies now conduct large amounts of business online, meaning that the speed and availability of their production systems directly impact relationships with suppliers and customers. With e-commerce [expected to be](#) a \$740 billion market by 2023, the need for high-performing and available servers is a cost of doing business.

Digital business places new demands on IT infrastructure in other ways as well.

- ▶ Website performance and availability is critical to establishing customer confidence as well as articulating brand value and mission. In the early days of the Internet customers learned to tolerate occasional outages to get the information and services they needed. Today they simply leave.
- ▶ Organizations trade data with business partners as part of the management of their supply chains. Data must frequently be exchanged in near real time to expedite deliveries, track shipments and identify fraudulent transactions. Latency and downtime are big problems.
- ▶ Marketing organizations subscribe to a multitude of data services that provide information that can be used for customer profiling and personalization. The data volumes can be very large and multiple streams often have to be merged into a single database.
- ▶ Online advertising requires moving and analyzing massive amounts of data for split-second programmatic decision-making.
- ▶ Fleet management, package tracking and logistics systems often require up-to-the-second data about the location of assets to enable operational efficiency.
- ▶ In some industries, like financial services, latency is the difference between success and bankruptcy.

Digital business is also redefining relationships that were once a patchwork of point connections into an integrated mesh in which organizations may transact business with each other constantly, occasionally or sporadically. Connections are unpredictable and opportunistic. Arbitrage systems compare prices and availability from multiple suppliers to enable transactions to be performed at the lowest cost. Applications built on microservices or serverless platforms must access data and services from a multitude of sources to work properly. Cloud workloads and data needs to be shifted seamlessly between platforms to give customers the best balance of performance and cost. These use cases and many others require fast, fluid connections that can be set up and torn down in seconds.

Organizations that adopt SaaS recognize that installing, patching and upgrading software isn't a core competency and are eagerly outsourcing that function to specialty vendors.

At the same time, digital transformation is prompting organizations to carefully assess core competencies and choose which operations to keep in house and which to outsource. SaaS, which at [\\$85 billion](#) is by far the largest segment of the overall cloud market, is emblematic of this trend. Organizations that adopt SaaS recognize that installing, patching and upgrading software isn't a core competency and are eagerly outsourcing that function to specialty vendors. The custom applications that enterprises create are increasingly collections of SaaS services woven together in unique ways. Reliable, high-bandwidth connections are essential to supporting this new functionality.

The technical drivers of colocation

As noted earlier, the IT world is increasingly multi-cloud and hybrid cloud. IT leaders want maximum flexibility to deploy workloads where they make sense, to provision additional infrastructure for peak load periods and to shift workloads easily between on-premises and multiple cloud platforms.

Point-to-point connections between entities over the public Internet or privately provisioned lines are erratic and expensive. Shared connections are vulnerable to traffic bursts that may impact performance for everyone as well as anomalous events such as denial of service attacks. Cost is also an issue. The faster and more reliable the connection needs to be, the higher the price tag. Latency is inherent in connections that traverse long distances, making the public Internet a risky place to do business where time is a factor.

In many cases it is physically impossible to reliably transfer multi-terabyte-sized files over a standard Internet connection at the speed the business demands.

In some applications, conventional Internet connections can never provide the reliability and performance that is required. For example, advertising networks thrive on high-speed communications. In many cases, brokers can deliver ads on a page before the rest of the content even loads. The faster the throughput, the more quickly ads can be matched to available inventory and served on a timely basis.

Another throughput-intensive application is the transfer and processing of media such as video and high-resolution images. Film editors, for example work with multi-terabyte-sized files that may need to be shared with partners in other parts of the world to enable “follow the sun” editing. In many cases it is physically impossible to reliably transfer those files over a standard Internet connection at the speed the business demands. Architectural firms, ad agencies, photo bureaus and mapping firms are among the many other types of businesses that work with extremely large file sizes.

A network architecture that best reflects the structure of modern business uses a distributed model in which data is processed close to the point of origin or decision. A good analogy is commercial aviation. There are more than 5,000 airports

Each hub can house multiple customers who can connect with each other directly as well as with business partners in other locations over the high-speed backbone.

serving the general public in the U.S. alone. If all travel was point-to-point between them, the cost and complexity would make flying prohibitively expensive. The airline industry long ago adopted a network model based on large hubs that connect to smaller airports. While it may take longer for travelers to get to their destination, they have more scheduling options, greater flexibility to re-route around adverse weather conditions and much lower costs.

Fortunately, distributed networks can have the best of both worlds. High-speed backbones between network hubs ensure that traffic flows smoothly and quickly. Each hub can house multiple customers who can connect with each other directly as well as with business partners in other locations over the high-speed backbone. Messages routed through central hubs can be directed more efficiently and less expensively to their destinations because all traffic is managed centrally. However, customers aren’t precluded from connecting to each other independently of the hub. They can move processing to the edge and use the high-speed backbone only when needed.

As a result, costs are lower because traffic doesn’t have to traverse leased lines or the public Internet. A hub enables customers to connect to each other, to value-added service providers and to major cloud platforms. The result is a cost-effective, secure and high-performance alternative to long-distance dedicated connections from one on-premises data center to another.

Understanding interconnection

The architecture just described is a strategy used in interconnection, and it is the fastest growing component of the colocation industry. Interconnection refers to physical and virtual data connections between companies that enable the rapid exchange of data. Colocation and interconnection have been inseparable since the early days of the industry, but the interconnection component is growing as connectivity becomes an important driver of the colocation decision for cost, security and scalability reasons.

Colocation provides a central meeting place for networks, clouds and enterprises to host their physical infrastructure and interconnection enables them to efficiently exchange traffic with one another. Selected IT services from business partners can be physically located in a colocation facility for maximum performance.

This approach takes advantage of the fact that colocation facilities are optimized to house multiple customers. Those customers can connect with each other over many different mediums including Ethernet to enable data exchange at extremely high speeds and at low cost without the latency penalty of public networks or the cost of private leased lines from an on-premises data center. Accessing the cloud via dedicated connections within a colocation facility can cut bandwidth costs by 60% compared to using the public Internet.

Interconnection has grown in importance as cloud usage has expanded and diversified. Companies now

[use an average](#) of 2.3 infrastructure-as-a-service providers and two platform-as-a-service providers. A [Skyhigh Networks analysis](#) of anonymized data from over 600 enterprises in 2016 found that the number of cloud-based applications in use had nearly tripled over a three-year period to an average of 1,427, a figure that is no doubt higher today.

All cloud and many SaaS providers use “on-ramps” such as Amazon Web Services’ [Direct Connect](#) or [Microsoft Azure ExpressRoute](#) to enable customers to establish dedicated connections between their networks and the cloud provider’s. Most of these on-ramps are actually routed through colocation facilities in order to reach the largest number of customers. Colocation providers partner with the big cloud platform providers to enable customers to establish private, low-latency network links between their dedicated infrastructure and cloud infrastructure providers for optimal performance of a hybrid IT environment. Customers benefit from reduced data transfer costs, improved security and optimal network performance along with the ability to connect to multiple cloud providers through a single interconnection point.

For example, the [CoreSite Open Cloud Exchange®](#) enables direct, private, virtual connectivity into AWS, Microsoft Azure, Google Cloud Platform and Oracle Cloud. Customers have the flexibility to select speeds and connectivity options that meets their needs and to scale bandwidth as needed, all from a central console.

Types of interconnections

There are several different categories of interconnections.

▶ PEERING EXCHANGE

In a peering exchange, two networks connect and exchange traffic with each other without having to pay a third party such as a telecom operator or Internet service provider to carry that traffic across the Internet. The operator can adjust routing to avoid bottlenecks and optimize performance. Peering essentially keeps traffic local to avoid the latency caused by multiple Internet hops.

▶ CROSS CONNECT

A cross connect is the equivalent of running a fiber or copper cable between each company’s servers. A physical, hardwired cable is provisioned between two different termination locations within the colocation data center, enabling high-performance, dedicated connectivity, excellent reliability and minimal latency.

INTER-SITE CONNECTIVITY

Inter-site connectivity provides communications between campuses in the colocation provider's network. This enables customers to access all the providers they need in an interconnection-dense facility and to grow without constraints within campus facilities. A variety of carrier-grade transport options are provided as well as common-carrier access to other regional interconnection hubs that are ideal for remote locations.

BLENDED IP

A blended IP service provides the convenience of the colocation provider working with a variety

of upstream carriers and ISPs to create a highly reliable, SLA-backed solution based on a fully redundant network architecture that provides the best performance across providers.

In addition, virtualized services like the **CoreSite Open Cloud Exchange®** provide a platform for accessing multiple public cloud providers and connecting distributed deployments across multiple markets. Using inter-site connectivity, customers can build multi-region cloud architectures from a single port with no long-term commitments.

Why use interconnection services from a colocation provider?

One of the most revolutionary concepts introduced by the cloud has been pay-as-you-go pricing. Today's colocation services enable the same model to be applied to interconnection services so that IT resources can be provisioned and scaled up and down in a cloud-like manner with no long-term commitments.

Major colocation providers operate multiple large data centers with high-speed pipes between them. For latency-sensitive applications that require processing to be as close to the point of delivery as possible, such as edge computing and IoT environments, this distributed architecture offers customers more choices about how to deploy infrastructure. Instead of isolating IT resources in a single data center, compute and storage can be distributed across multiple colocation facilities for maximum speed and redundancy.

Interconnection enables data exchange between two or more entities or partners at the fastest possible speed by combining high-performance networks with physical proximity. Customers avoid having to work around the unpredictability of the public Internet and data communication costs are minimal.

Customers using multiple cloud providers can enjoy access speeds equivalent to that of dedicated on-ramps along with the flexibility to switch sessions easily between cloud providers and transfer data with reduced egress charges, which can run as high as 12 cents per gigabyte.

There are other reasons to consider interconnection services from a colocation provider:

- ▶ Customers have access to a wide range of third-party services that in the same facility that can be accessed directly through interconnection.
- ▶ Software-defined networking enables flexible control of connections, automated provisioning, network segmentation and security for distributed IT infrastructure via private SDN connections that bypass the Internet.
- ▶ Customers with multi-cloud and hybrid cloud architectures and/or multi-site needs can manage their cloud providers through a single console as well as consolidate billing.
- ▶ Many colocation interconnection services are connected directly to undersea cables, enabling the fastest possible data exchange with partners in other countries.
- ▶ Distributed colocation infrastructure is the ideal environment for organizations that are experimenting with edge computing, which moves processing close to endpoints and applies intelligence to manage traffic between endpoints, local processing resources and the cloud. Customers who begin using colocation services now will be better equipped to take advantage of these decentralized networks.

Discover more about colocation and interconnection solutions at coresite.com.