

## INTELLIGENCE UPDATE

# Retail vs wholesale: finding the right colocation pricing model



Owen Rogers

15 Jul 2025

Colocation providers may offer two pricing and packaging models to sell similar products and capabilities. In both models, customers purchase space, power and services. However, the method of purchase differs.

In a retail model, customers purchase a small quantity of space and power, usually by the rack or a fraction of a rack. The colocation provider standardizes contracts, pricing and capabilities — the cost and complexity of delivering to a customer's precise requirements are not justified, considering the relatively small contract value.

In a wholesale model, customers purchase a significantly larger quantity of space and power, typically at least a dedicated, enclosed suite of white space. Due to the size of these contracts, colocation providers need to be flexible in meeting customer needs, even potentially building new facilities to accommodate their requirements. The colocation provider negotiates price and terms, and customers often prefer to pay for actual power consumption rather than be billed on maximum capacity. A metered model allows the customer to scale power usage in response to changing demands.

A colocation provider may focus on a particular market by offering only a retail or wholesale model, or the provider may offer both to broaden its appeal. The terms "wholesale" and "retail" colocation more accurately describe the pricing and packaging models used by colocation providers rather than the type of customer.

Table 1 Key differences between retail and wholesale colocation providers

	Retail colocation	Wholesale colocation
Typical customer	Small- to medium-sized businesses, startups	Large enterprises, hyperscalers (e.g., cloud providers)
Engagement	Direct, online, via agency/broker	Agency broker, or direct through an existing relationship. RFPs issued.
Space allocation	Shared space (e.g., 1 unit, quarter rack, half rack, full rack)	Dedicated space (e.g., private cages, suites, or entire rooms)
Power requirements	Typically <100 kW, but varies	Typically >100 kW to MW, but varies
Deployment time	Often immediate or within days	Weeks to months (may involve custom build-outs)
Customization	Limited	High (e.g., cooling, power, security)
Hands-on support	Colocation provider will perform simple tasks such as racking servers	Less day-to-day support: customer often manages infrastructure
Use case examples	Hosting a few web servers, backup infrastructure and small apps	Data centers for cloud platforms, scalable software as a service, content delivery networks, AI training

Retail colocation deals typically have higher gross margins in percentage terms, but the volume of sales is lower. Most colocation providers would rather sell wholesale contracts because they offer higher revenues through larger volumes of sales, despite having lower gross margins. As wholesale colocations are better prospects, retail customers are more likely to experience cost rises at renewal than wholesale customers.

## Retail colocation pricing model

Retail terms are designed to be simple and predictable. Customers are typically charged a fixed fee based on the maximum power capacity supplied to equipment and the space used. This fee covers both the repayment of fixed costs, and the variable costs associated with IT power and cooling. The fixed fee bundles all these elements together, so customers have no visibility into these individual components — but they benefit from predictable pricing.

In retail colocation, the facilities are already available, so capital costs are recovered across all retail customers through standard pricing. If a customer exceeds their allotted maximum power capacity, they risk triggering a breaker and potentially powering down their IT equipment. Some colocation providers monitor for overages and warn customers that they need to increase their capacity before an outage occurs.

Customers are likely to purchase more power capacity than they need to prevent these outages. As a result, some colocation providers may deliberately oversubscribe power consumption to reduce their power costs and increase their profit margins. There are operational and reputational risks if oversubscription causes service degradation or outages.

Some colocation providers also meter power, charging a fee based on IT usage, which factors in the repayment of capital, IT and cooling costs, as well as a profit margin. Those with metering enabled may charge customers for usage exceeding maximum capacity, typically at a higher rate.

Can a colocation provider increase prices during a contract term? Occasionally, but only as a last resort — such as if power costs increase significantly. This possibility will be stipulated in the contract as an emergency or force majeure measure.

Usually, an internet connection is included. However, data transfer over that connection may be metered or bundled into a fixed cost package. Customers have the option to purchase cross-connects linking their infrastructure to third-party communications providers, including on-ramps to cloud providers.

## Wholesale colocation pricing model

Wholesale colocation pricing is designed to offer customers the flexibility to utilize their capacity as they choose. Because terms are customized, pricing models will vary from customer to customer.

Some customers may prefer to pay for a fixed capacity of total power, regardless of whether the power is used or not. In this model, both IT power and cooling costs are factored into the price.

Other customers may prefer a more granular approach, with multiple charging components:

- Fixed fee per unit of space/rack based on maximum power capacity and is designed to cover the colocation provider's fixed costs, while including a profit margin.
- Variable IT power costs are passed directly from the electricity supplier to the customer, metered in kilowatts (kW). Customers bear the full cost of price fluctuations, which can change rapidly depending on grid conditions.
- To account for variable cooling costs, power costs may be calculated by multiplying actual power usage by an agreed design PUE to create an "additional power" fee. This figure may also be multiplied by a "utilization factor" to reflect cases where a customer is using only a small fraction of the data hall (and therefore impacting overall efficiency).

Some customers may prefer a blended model of both a fixed element for baseline capacity and a variable charge for consumption above the baseline. Redundant feeds are also likely to impact cost. If new data halls need to be constructed, these costs may be passed on to the customers directly, or some capital may be recovered through a higher fixed rack fee.

Alternatively, for long-term deployments, customers may opt for either a "build-to-suit" or "powered shell" arrangement. In a build-to-suit model, the colocation provider designs and constructs the facility—including power, cooling and layout—to the customer's exact specifications. The space is then leased back to the customer, typically under a long-term agreement exceeding a decade.

In a powered shell setup, the provider delivers a completed exterior building with core infrastructure, such as utility power and network access. The customer is then responsible for outfitting the interior (racks, cooling, electrical systems) to suit their operational needs.

Most customers using wholesale colocation providers will need to implement cross-connects to third-party connectivity and network providers hosted in meet-me rooms. They may also need to arrange the construction of new capacity into the facility with the colocation provider and suppliers.

Hyperscalers are an excellent prospect for wholesale colocation, given their significant scale. However, their limited numbers and strong market power enable them to negotiate lower margins from colocation providers.

Table 2 Pricing models used in retail and wholesale colocation

	Retail colocation pricing	Wholesale colocation pricing
Billing unit	Per rack, rack unit or kW	Per MW, kW block or entire data hall/suite
Power pricing	Flat rate: per kW capacity (maximum power available)	Metered usage (actual power consumed)
Space pricing	Per rack or partial rack	Per square meter/feet or cage/suite
Cooling	Included in the base price	Factored in through PUE multiplier
Cross-connect fees	Often charged per connection (monthly plus setup)	Negotiated; lower per-connection cost due to scale
Remote hands	Pay-as-you-go or limited included hours	Typically charged separately; the customer may provide their own team
Setup fees	Low or none; standard setup	Higher, may include custom build-out, engineering, etc.
Contract length	Shorter terms (monthly to 1-2 years common)	Long-term contracts (5-20 years)
Price flexibility	Less negotiable; standard pricing tiers, standard contract/agreement	Highly negotiable based on scale and commitment, custom contract/agreement
Inclusions	Often includes IP transit, monitoring and basic support	Bare infrastructure; customer-managed

## The Uptime Intelligence View

In a retail colocation engagement, the customer has limited negotiating power — with little scale, they generally have minimal flexibility on pricing, terms and customization. In a wholesale engagement, the opposite is true, and the arrangement favors the customer. Colocation providers want the scale and sales volume, so are willing to cut prices and accommodate additional requirements. They are also willing to offer flexible pricing in response to customers'

rapidly changing requirements.

Hyperscalers have the strongest market power to dictate contracts and prices. With so few players, it is unlikely that many hyperscalers will be bidding for the same space, which would push up prices. However, colocation providers still want their business, because of the volume it brings. They would prefer to reduce gross margins to ensure a win, rather than risk losing a customer with such unmatched scale.

## ABOUT THE AUTHOR

---



### Owen Rogers

Dr. Owen Rogers is Uptime Institute's Senior Research Director of Cloud Computing. Dr. Rogers has been analyzing the economics of cloud for over a decade as a chartered engineer, product manager and industry analyst. Rogers covers all areas of cloud, including AI, FinOps, sustainability, hybrid infrastructure and quantum computing.

[orogers@uptimeinstitute.com](mailto:orogers@uptimeinstitute.com)

## **About Uptime Institute**

Uptime Institute is the Global Digital Infrastructure Authority. Its Tier Standard is the IT industry's most trusted and adopted global standard for the proper design, construction, and operation of data centers – the backbone of the digital economy. For over 25 years, the company has served as the standard for data center reliability, sustainability, and efficiency, providing customers assurance that their digital infrastructure can perform at a level that is consistent with their business needs across a wide array of operating conditions.

With its data center Tier Standard & Certifications, Management & Operations reviews, broad range of related risk and performance assessments, and accredited educational curriculum completed by over 10,000 data center professionals, Uptime Institute has helped thousands of companies, in over 100 countries to optimize critical IT assets while managing costs, resources, and efficiency.