

## What and where are edge data centres?



Although it is still early days for significant edge data centre investments, there is increasing interest and demand for the benefits they bring. However, there is much confusion for what constitutes an edge data centre. In this article we will run through the different types of edge data centres, beginning at the nearest point to the end-user.

Matt Bamforth, Consultant

Edge data centres are located closer to the end-user than the public cloud. They enable a wide range of next-generation applications and provide their customers with the benefit of lower latency, higher security, and greater control over their data. The public cloud – dominated by the hyperscalers (Google, Amazon, Microsoft) – has a large number of benefits and carries the weight of the internet on its back. However, as more applications are developed which will benefit from edge computing, edge data centres will become a large area of investment.

The term edge data centre sparks much debate; edge computing can exist at various locations and these data centres can take many different forms. Here we try and provide clarity on the different types of edge data centres.

**Figure 1: Types of edge data centres**

Type of edge	Data centre	Location	Number of DCs per 10m population	Latency (average)	Size	
<b>On premises edge</b>	Enterprise site	Businesses	NA	2-5 ms	1 rack max.	
<b>Network (mobile)</b>	Tower edge	Tower	Nationwide	3000	10 ms	2 racks max.
	Outer edge	Aggregation points	Town	150	30 ms	2-6 racks
	Inner edge	Core	Major city	10	40 ms	10+ racks
<b>Regional edge</b>	Regional	Major city	100	50 ms	100+ racks	
<b>Not edge</b>	Hyperscale	State/national	1	60+ ms	5000+ racks	

Source: STL Partners

## On premise

On-premise data centres have existed for some time already and refer to data centres located at the enterprise’s own facilities and are 100% owned by themselves. Historically, their core purposes has been to keep within complete control of the organisation. They can run a private cloud which enables some of the scalability of the public cloud, while being far more secure. Enterprises can also benefit from more customisable infrastructure that is tailored towards the needs of their applications.

The on-premises edge is slightly different. Here we are referring to micro-data centres at the customer site. These can take many shape or forms. We are seeing edge being deployed as one or two servers, as a rack co-located in the existing on-premise data centre, or as a mini-data centre enclosure which could be up to a single rack. These are designed to be as close to the end-user as possible, where latency is lowest and 2-5ms roundtrip is achievable.

## Network – tower edge

Towers are located at the ‘last mile’ – the final step of the telecoms networks to end customers. There are thousands located over a region, and crucially all over cities where demand for edge can be highest, including applications such as connected cars, cloud gaming, and AR/VR. Towers are the perfect location for micro data

### What and where are edge data centres?

centres as real estate in cities is expensive and limited, hence it is very hard to build larger data centres here. Plus, they are already equipped with connectivity and power – two critical factors to enable data centres.

Tower companies (and mobile operators who still own their towers) can capitalise on this opportunity. Many next-generation applications will need compute to be brought to the last mile to achieve the necessary latency. Meanwhile, carriers are virtualising their RAN, which will accelerate investment in data centre-like facilities at towers and consolidate network equipment at these premises. IT infrastructure for edge applications can be co-located at these premises, which optimises investments, while also opening up new revenue opportunities from delivering edge cloud services.

## Network – outer edge (aggregation points)

This is traditionally where all of the telephone lines in a region meet and are electronically directed to their destination. Privately owned by the telco, 'central offices' are found in every locality, and are regarded as being at the outer edge of the mobile network. While they formerly served fixed line connections, they are being repurposed by telcos into more digital facilities and consequently have lots of empty space. This makes them good candidates for edge cloud data centres. Located between the RAN and the core, central offices are close to the end-user and data centres here can achieve latency of ~30ms.

## Network – inner edge (mobile core)

The mobile core is known as the inner edge of the mobile network, and similarly to the outer edge (central office) is privately owned by the telco. This is where the core/interfaces nodes are located and different services are provided to customers who are connected to the access network. These sites need to be near the peering points and hence are located near major cities, this makes them another good candidate for edge cloud data centres. Inner edge data centres are much bigger than outer edge data centres, and are more reminiscent of typical data centres.

## Regional

Regional data centres are the largest types of edge data centres. These colocation data centres host hundreds of different types of customers so will usually be either in or near a major city, serving a significant population. Retail customers (e.g. cloud providers, large enterprises, application developers) can rent cabinets holding server racks, whole cages filled with cabinets, or indeed the whole data centre. On average, a latency of ~50ms can be achieved here which is an improvement on the ~60ms achieved at the hyperscale data centres. However, this can vary significantly by market. In the United States, the latency from highly centralised networks and hyperscale data centres is driving investment in regional data centres in Tier-2 cities as big as Austin, Texas.

The advantage of regional data centres over tower colocation or on-premise data centres is in cost. Economies of scale can be achieved with cooling and power, and any other operations or maintenance expenses. Many regional colocation centres may host an Internet Exchange but private peering occurs too. Some telcos own private data centres which host private peering at the regional edge as well.

Much of the large **edge investments today** are happening at the regional edge and driven by infrastructure companies such as DataBank and Compass Datacenters. However, we anticipate that we will see further investments trickle down into smaller edge data centres as these build up over time.

## Hyperscale

Hyperscale data centres are not on the edge – in fact, the edge begins where the hyperscale data centres end. These are massive facilities located in the middle of nowhere, usually defined as exceeding 5,000 servers and 10,000 sq. ft with the number of servers running into the millions. However, the term 'hyperscale' does not

### What and where are edge data centres?

refer to size, but rather an ability to scale in response to increased demand. The size of data centres that these companies such as Google, Microsoft, and Amazon rely upon mean that leasing racks in a colocation facility or using out of the box hardware from vendors is not a possibility – they must operate and exclusively control a massive data centre of their own. These hyperscale data centres are where the public cloud is located.

**Matt Bamforth is a Consultant at STL Partners, specialising in edge computing, 5G, and private networks.**

Get in touch with the author to learn more

[matt.bamforth@stlpartners.com](mailto:matt.bamforth@stlpartners.com)

Or visit STL Partners' Edge Hub

[www.stlpartners.com/edge-computing](http://www.stlpartners.com/edge-computing)