Edge Computing 2023

Joe Madden

MOBILE EXPERTS | July 2023

Abstract

Predictions on the future evolution of Edge Cloud deployment for enterprise applications in multiple vertical markets, including Oil & Gas, Industrial, Manufacturing, Retail, Gaming, and Media. The preferred business models between Telco, Cloud, and Neutral Host are explored and forecasted. The five-year forecast for deployment includes expectations for data centers as well as revenue from software and Edge Cloud services.

TABLE OF CONTENTS

CHARTS

FIGURES

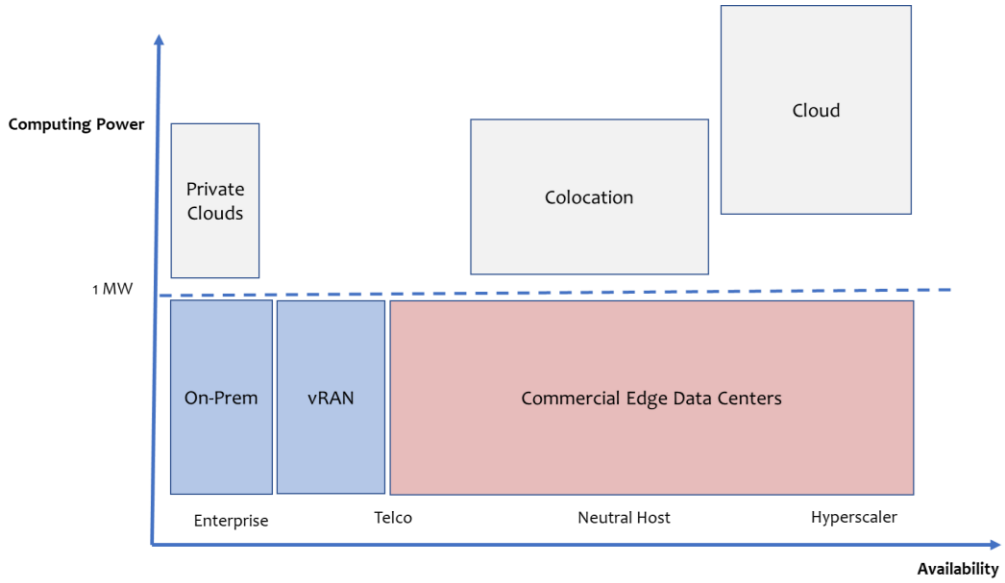
DEFINITIONS

Definition of Edge Computing

Every player in the industry has a different definition of the Edge. The concept of the "Edge" can range from regional data centers down to handheld devices, and of course there will be computing at every level because of the vast variety of use cases in the world.

The purpose of this report is to investigate the potential for network-based Edge Computing infrastructure and services. To fully understand this segment, we focus on investments by network operators, hyperscalers, and neutral hosts in regional and local data centers that are shared among multiple applications. We also cover on-premises data centers deployed for specific enterprises because these investments directly compete with the network-based Edge concept.

For brevity, we do not examine computing in mobile devices, cars, or other 'hyper-local' use cases.



Source: Mobile Experts. This report covers the colored areas.

Figure 1. *Edge Computing Segmentation*

One common thread is that the Edge constitutes a cloud platform smaller than a hyperscale data center, which is positioned as close to the user as necessary for the application to function properly. In other words, we see the essential nature of Edge Computing as based on a small data center, handling local traffic. By contrast, a cloud data center is typically larger in size (hundreds of MW of power) and handles any global computing load.

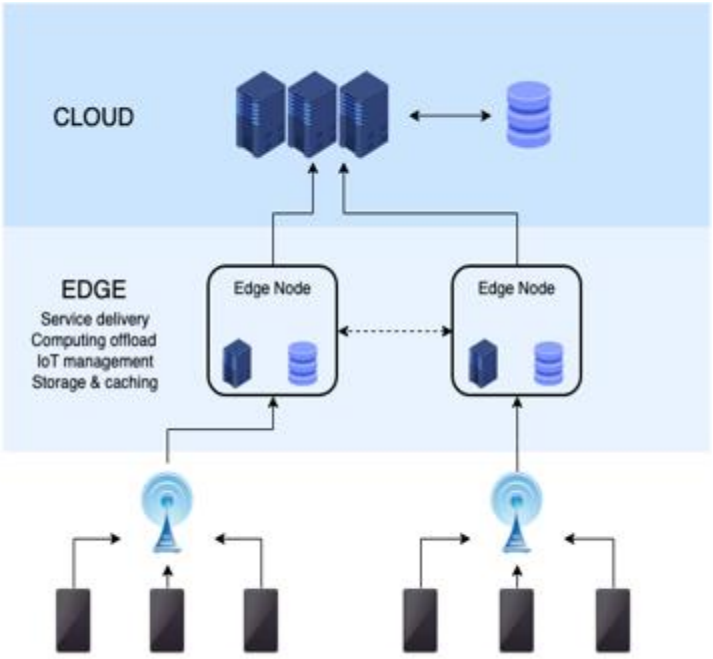
Another way to say this: Hyperscale data centers are optimized for cost with some concessions to latency and location. Edge data centers sacrifice on economy of scale to provide better latency or transport costs. In short, the priorities are different for each case.

In order to be crisp and clean in our forecasting model, we have decided to distinguish between Edge Computing and Cloud Computing using the size of the data center. Hyperscale Cloud providers and their colocation partners build data centers with power in the level of 100 MW and above. These can be located thousands of miles away (typical hyperscale data center) or they can be nearby (colocation cloud data centers), so distance and latency are not precise ways to distinguish between Cloud and Edge.

Edge data centers can range from 100W to a few megawatts. We currently use 1 MW as our defining point, but we anticipate that this will rise in the future. For our definition as a Commercial Edge Data Center, the facility must be available for multiple client applications. This report tracks On-Prem Edge Data Centers which are placed for and by an enterprise for a specific application. However, our main focus rests on the Commercial Edge Data Centers that are created as infrastructure in the hope of building up multiple applications.

Technical Background

Edge Computing, or more formally known as Multi-Access Edge Computing (MEC), is a network architecture concept that enables cloud computing, or IT service delivery environment, closer to the edge of a mobile or fixed network. The basic premise of MEC is that by running applications closer to end users or devices, IT services delivered via the cloud can yield much better user experiences. Simply put, MEC is a cloud-computing environment at the network edge.



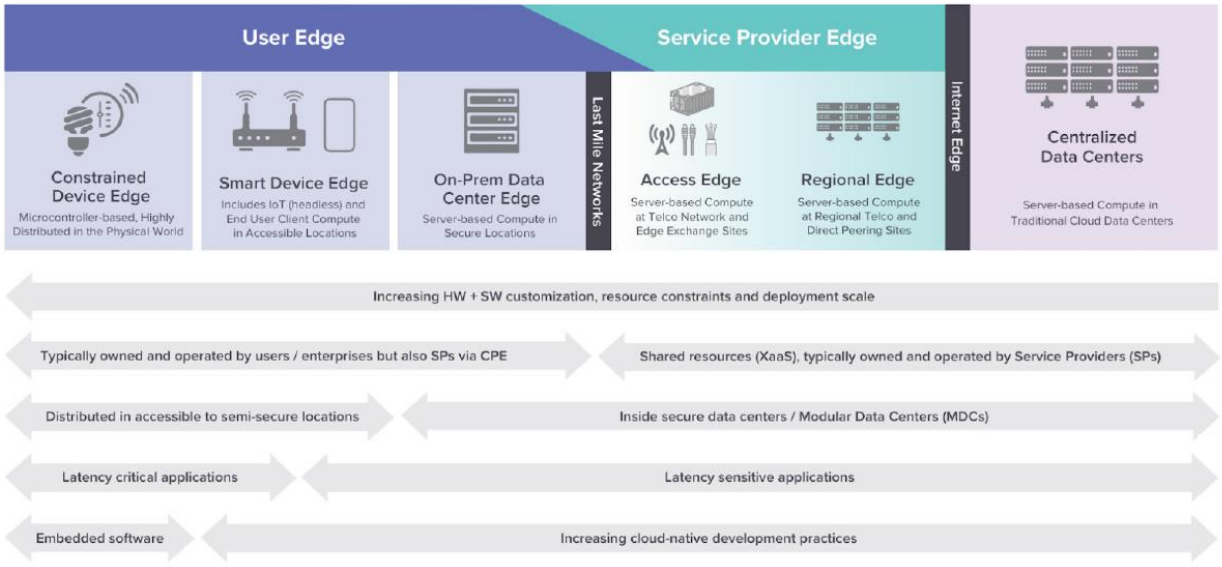
Source: Thin-nology

Figure 2. *Edge Computing Concept Diagram*

Where is the edge? Well, it's not a single fixed position… it's different in each vertical market and each application. The Edge Computing node could be regional, covering hundreds of kilometers, local (covering 10-20 kilometers), or On-Premises, depending on the application. In general, economic pressure in the up-front investment will push the nodes to be more centralized, while latency, throughput costs, and reliability concerns will push the edge node closer to the end user.

Segmentation of MEC: Access Edge, Regional Edge, On-Prem, and vRAN

In this report, we classify edge clouds into three different categories, which in the past we have referred to as: *Near Edge*, *Far Edge*, and *On-Prem*. The Linux Foundation *State of the Edge* report has tried to harmonize the terminology, so we have decided to change our terms to match up, shifting to Regional Edge, Access Edge, and On-Prem. These categories are based on our expectation of where the communication service providers or enterprises will likely deploy MEC infrastructure in typical network architecture, as shown below.



Source: Linux Foundation SOTE report 2021

Figure 3. *Edge Computing Continuum*

In this report, we make one change to distinguish between different types of On-Prem edge data centers. Many of these are in a customer's building, such as a manufacturing plant. But the telecom vendors alone represent one major vertical market. A large proportion of the "On-Prem edge data centers" deployed in 2021 were, in fact, vRAN applications running in a server at the bottom of a radio tower. To keep track of these, we broke out the telecom application as a significant On-Prem market.

The primary purpose of this report is to identify the market for open and available Edge Computing resources that will build the open ecosystem. So, while we break out the On-Prem data centers and vRAN as interesting deployments, we don't really consider these to be "Commercial Edge Computing." Throughout our forecast, we will identify the "Commercial Edge Computing" market as the area where local resources are set up to be available for new enterprises and applications to grow, not just the market for a single enterprise to perform its local computing.

We recognize that graphic processing power in a smartphone is excellent, and can be considered Edge Computing, but we also see a major distinction, as end-user devices are generally not used for computing tasks by other nearby customers… they're used for the customer who owns the device. The purpose of this report is to evaluate the potential for Commercial Edge Data Centers that serve multiple clients.

Regional Edge

As shown in the above diagram, Regional Edge refers to an Edge Cloud deployed closer to hyperscale public cloud data centers. At this level, MEC is typically deployed

in a regional sense, with between 30 and 80 near edge data centers for a country like the United States. In contrast to the Access Edge, the Regional Edge handles a higher number of end users or devices and focuses on the economies of scale instead of latency. For example, video CDN applications for the consumer market would be a good fit for Regional Edge deployment.

Access Edge

Access Edge refers to MEC infrastructure that is deployed closer to the end user or devices, with more emphasis on latency and local services than cost. In other words, an Access Edge Data Center might be located only a few kilometers from the customer— and as a result, Access Edge may be more tailored to a specific anchor tenant or application. One example is the Local Zone that AWS set up in Hollywood to handle multiple movie studios. In this case, the memory and bandwidth were tailored for heavy use of computer graphics software for the local customers.

Keep in mind that to qualify as a MEC data center, we want to see the deployment controlled by a company like AWS/Azure/GCP, by a telco, or by a neutral host.These players intend to offer services to any new customers that arise in that neighborhood, whereas the same computing power might be deployed on a single customer's premises and closed off to other customers.

On-Prem

On-Premises edge is clearly the largest segment, with hundreds of "data centers" deployed in a wide variety of customer buildings and facilities. We've broken out the vRAN application as a special kind of "on-prem" data center… this is essentially a telecom company that is using edge computing in its own operations and therefore should be distinguished from commercial Edge. Tens of thousands of vRAN sites are deployed now, so the numbers are much higher than the spotty data center deployment in various manufacturing facilities, schools, or other buildings.

The On-Prem cloud stack could be similar to the Access Edge, using something like Azure Edge or AWS Outpost. In the On-Prem category, we specifically refer to edge infrastructure that is deployed physically at an enterprise location.

vRAN Sites

We've broken out the vRAN sites this year because of the heavy deployment in tens of thousands of sites for DISH and Verizon. Where Rakuten could be seen as a "Commerical Edge Computing" play, DISH is using AWS for its vCU sites and its vDU is located at the radio tower, without capacity or capability to handle MEC applications.

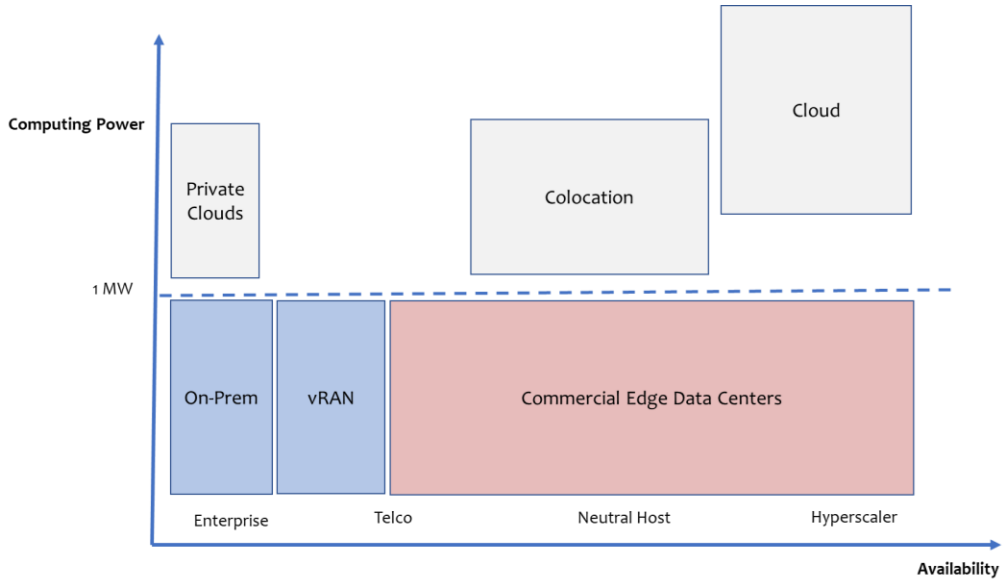
Verizon's deployment is similar, where the vCU is currently in 19 different AWS sites but the vDU is On-Prem at the tower in a form factor that doesn't lend itself to expansion for MEC customers.

METHODOLOGY

To create estimates and forecasts for Edge service revenue and data center deployments, Mobile Experts relied on direct input from more than 30 industry sources, with communication service providers, edge compute software vendors, server hardware vendors, and some enterprise IT companies. Mobile Experts built a "bottom-up" list of edge data centers based on interviews and public announcements from mobile operators and edge infrastructure companies. Chinese data center numbers were taken from government ministry reports and financial reports from the operators, and we estimated the numbers in on-prem, campus, and other locations from our interviews with Chinese suppliers.

For MEC service revenue forecast, Mobile Experts built a "top-down" forecast based on trends in various use cases identified and general enterprise adoption trend of cloud computing model for technology consumption.

Mobile Experts has broken the Cloud Computing market up, segmenting the market by the size of the data center and by the availability of the data center for new customers. The colored segments are covered in this report, and our main focus is on the red colored segment which is available for Edge Computing services commercially.



Source: Mobile Experts

Figure 4. Segmentation of Edge Computing Market

Cloud Computing	Computing based in locations with more than 200 kW of power such as hyperscale data centers (100 MW +) and co-location data centers (typically 1-10 MW). Each data center is open to traffic from any location
Edge Computing	Computing based as close to the user as practical, in data centers below about 1-2 MW, dedicated to local traffic
Infrastructure Cloud Service	Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) combined. As hyperscale cloud service providers increasingly gain scale, traditionally separated IaaS and PaaS services are offered as an underlying infratructure cloud service.
Application Cloud Service	Applications consumed as a service. Often referred to as "Software as a Service" (SaaS).

Source: Mobile Experts

Figure 5. Public Cloud Service Definitions

Video / CDN	Video processing, caching, or analytics of higher-resolution video streams and delivery. Online video streaming CDN by fixedline providers like cable operators are included.
Cloud Gaming	Gaming delivered over the cloud so that users can play online games on different devices. Enterprise opportunity arises from working with gaming or hyperscale cloud providers like AWS, Azure and Google Cloud Platform to deliver to real-time mobile games at scale
Industrial IoT	Industrial automation applications in smart factory, utilities, mining and construction
Retail / E-commerce	E-commerce use cases to expedite e-commerce via mobile and fixed devices
Automotive (Infotainment/ Driving)	Automotive applications related to safety, vehicle infotainment, HD mapping, navigation, and autonomous driving
AR / VR	Augmented and virtual reality applicartions via goggles for enterprise applications like training, real-time diagnostics, and AI-powered assistance.
Aerial / Drones	Drone and other aerial applications for navigation or autonomous flying
Virtual RAN	Baseband processing for telecom Radio Access Network such as 4G or 5G communications, including Baseband Unit (BBU), Distributed Unit (DU), and Centralized Unit (CU) processing tasks. Some of these tasks must be performed within about 20 km of the radio site due to realtime processing considerations.

Source: Mobile Experts

Figure 6. Edge Use Case Definitions

Regional Edge Data Center	MEC edge cloud location located in metro area (20-100 servers per site), planned for 200-300 km proximity to customers (commonly referred to "near edge")
Access Edge Data Center	MEC edge cloud location closer to RAN aggregation sites (4-8 servers per site), planned for 20-100 km proximity to customers (commonly called "far edge")
On-Prem Data Center	MEC edge cloud location resides on-site at enterprise locations (1-5 servers per site)

Source: Mobile Experts

Figure 7. Edge Data Center (Location) Definitions

North America:	USA and Canada
Latin America:	Mexico through South America, including Caribbean
Europe:	Western and Eastern Europe, including Russia
China:	China, including Tibet and Hong Kong
Asia Pacific:	India through Australia/Micronesia, excluding China
Middle East/Africa:	Pakistan and Turkey through Africa

Source: Mobile Experts

Figure 8. Geographic Regions