Stratecast FROST & SULLIVAN

April 2019

In Search of the Edge

Demystifying Edge Compute Infrastructure

Stratecast Analysis by

Lynda Stadtmueller

Cloud Computing (CC)

Volume 9, Number 3

In Search of the Edge

Demystifying Edge Compute Infrastructure

Introduction[1]

The edge appears to be everywhere these days. Industry pundits and vendors talk about edge computing and edge networking, edge devices, and edge security.

But what are we all talking about? Definitions of "the edge" bring to mind the fable about the blind men describing an elephant. Each description reflects the provider's limited perspective. All are correct; none is complete.

The problem with the conflicting definitions and vantage points is that enterprises are left without guideposts to assess the new technology. Do they need an edge strategy? If so, how should they go about buying, building, or implementing one?

In this report, we explore the edge landscape. We parse the myriad definitions, and show how each component is relevant to a total solution. We also step back and show how the principles behind "edge compute" are consistent with best practices for any workload deployment in a hybrid cloud environment.

Defining the Edge

As in the English language, the word "edge" is relative, necessarily referring to the outer limits of *something* (usually a provider's network or cloud). In the Information & Communications Technology industry, the "edge" can refer variously to a device, a location, or a configuration that combines compute and network resources.

But in common usage, "edge" usually means "external to the cloud." Like Saul Steinberg's illustration of a New Yorker's view of the world,[2] the common depiction is of a cloud-centric universe with edge locations or workloads dotting the perimeter like satellites. While this image has some usefulness, it is limiting: in fact, neither the public cloud nor the internet is necessary for some edge workloads.

Some veteran IT leaders recognize edge computing as the next wave in an ongoing cycle of centralization/distribution of compute resources that has characterized compute technologies from the early days. Consider the successive eras of business computing:

- Mainframe (centralized)
- Personal computers (distributed)
- Cloud (centralized)
- Edge (distributed)

[1] In preparing this report, Stratecast conducted interviews with the following companies: Akamai, CenturyLink, Comcast, Dell/EMC, Equinix, Microsoft, TenFour, Verizon, and vXchnge.

Please note that the insights and opinions expressed in this assessment are those of Stratecast, and have been developed through the Stratecast research and analysis process. These expressed insights and opinions do not necessarily reflect the views of the company executives interviewed.

[2] View of the World from 9th Avenue - Wikipedia

However, this is not to imply that the public cloud will give way to local edge computing as a dominant architecture. Instead, edge compute is finding its way into enterprise hybrid IT environments, comprising premises-based and hosted cloud infrastructure options. Today, even cloud evangelists who believe that all modern workloads belong in the public cloud, graciously accept edge workloads as the narrow exception.

Why Edge? Why Now?

Why is edge computing having its moment in the sun? The business drivers include:

- **Increased thirst for data:** Businesses recognize data as a valuable asset—one that can be leveraged for operational efficiencies, new revenue opportunities, customer satisfaction, even industry disruption. Fifty-one percent of business leaders surveyed by Frost & Sullivan said that investigating and utilizing data is a top business priority.[3] To support the goal, businesses are seeking to collect data points to support every area of the business. This includes data collection and processing far from traditional data centers, such as in an Internet of Things (IoT) configuration.
- **Growth of analytics and artificial intelligence:** Unless the data yields insight, collection is a waste of time and resources. Organizations are aggregating and exposing more data to new, increasingly intelligent analytics tools, including machine learning and AI, as a way to improve decision-making. In the Frost & Sullivan survey, 47% of business decision-makers said that data analytics was more important to their organization than other technologies.
- **Application performance requirements:** As they transform to digital businesses, enterprises are keeping a close eye on application availability and performance. Seventy-two percent of IT decision-makers surveyed by Frost & Sullivan cited concerns about poor or inconsistent application performance as a key driver for IT investments. For latency-sensitive applications—for example, a factory-based predictive analytics tool that responds in real-time when a manufacturing component is at risk of imminent failure—network delay may be intolerable; thus, driving businesses to place applications and data close to the users or machines they serve.
- **Application and data security:** Compliance requirements, as well as the need to protect proprietary data assets, mean businesses must ensure that any solution includes adequate end-to-end security and control points. This includes data sovereignty regulations that specify that certain information (e.g., personal data) must remain in-country. For sensitive applications, businesses are increasingly choosing to keep data out of the shared public cloud. Eighty percent of IT leaders say that security concerns have driven a decision not to deploy applications in the public cloud; 75% say compliance is a factor in the decision. Furthermore, "inability to meet compliance requirements" was rated as a top reason for repatriating workloads from the public cloud to the premises data center.

[3] Unless otherwise cited, all survey statistics in this report are from the following Frost & Sullivan reports:

- Big Data and Analytics Survey, June 2018
- An End User Perspective on Navigating Digital Transformation, Global, October 2017
- Frost & Sullivan Cloud User Survey, 2018

- **Cost efficiencies:** Businesses are looking to drive costs out of the business. Cost reduction is cited as a strategic priority by 78% of business leaders; 72% say they are willing to invest in technologies that will help reduce spending. Achieving this goal may mean deploying multiple platforms to process data. For example, some organizations maintain separate Security Information and Event Management platforms to take advantage of varied vendor pricing schemes. They choose a lower-cost platform for first-tier processing and analysis; and a higher-cost platform for more sophisticated processing and analysis on a smaller, more refined data set (or conversely, on a diverse data set that aggregates data from multiple sources).

To support these strategic business priorities, most IT leaders are implementing hybrid cloud strategies that are flexible enough to meet diverse application needs. In particular, as they consider architectures for deploying a host of new data-intensive applications, they must address unique technology considerations, including:

- **Management of sensor-enabled devices:** OEM manufacturers have stepped up to the data collection challenge by producing a variety of sensor-equipped devices embedded with increasingly powerful and tiny microprocessors and network gateways. Such sensors fuel machine-to-machine communications and Internet of Things (IoT) applications, performing operational self-monitoring of the device and/or collecting and transmitting environmental data. IT organizations must be prepared to procure, deploy, and maintain the devices.
- **Network latency and bandwidth constraints:** Many sensor-enabled devices are not large enough to host analytics and application software, or to store and process the influx of data. Therefore, the solutions rely on a remote data center (private or cloud) to exchange data and perform those functions. But transmitting massive amounts of data from sensor to cloud (and back again, for certain applications) can tax the ability of available networks—especially when the sensor-equipped devices are located far from traditional data centers. Furthermore, network latency may be unacceptable for some use cases, delivering inconsistent quality and performance. Network reliability also is a consideration: enterprises must weigh the risks associated with a single network link (single point of failure) versus diverse network links, which add to direct and overhead costs.
- **Need to reduce hardware/software maintenance:** Local deployments—whether in a branch office, factory, or field—make sense where network constraints come into play. However, remote data collection sites are rarely staffed with IT professionals. To minimize deployment of technical resources, local solutions should be designed to operate with minimal on-site technical knowledge. Such solutions are generally monitored and managed from a central data center, with the devices themselves often designed to be "plug and play."

The Hybrid Edge Solution

To support the business and technical needs for data insights, IT organizations are building hybrid solutions that comprise:

- Data collection and some processing close to the source, via low-maintenance equipment and remotely-managed software.

- Transmission of data or metadata to a data center (private or cloud—or both), via secure, low-latency network links.
- Cloud-based processing and storage. This may include aggregation of data from multiple sources, integration with the latest analytics and machine-learning software, access by multiple users and applications, secure and cost-effective storage and backup, and instructions transmitted back to the devices.

This is what we call an edge solution.

Where is the Edge?

As noted, the "edge" is not in the cloud; nor is it in a traditional enterprise data center or branch server room. But where exactly *is* the edge? Depending on the vendor and/or workload, the edge can define any number of places.

- To device manufacturers and IoT solution providers, the edge may refer to devices equipped with sensors and network access.
- To some network service providers, the edge is the network Point of Presence: the interface between the provider's core network and the local access network (whether broadband, wired, or wireless), or interconnect point to the cloud.
- To other WAN and broadband service providers, the edge may be the customer premises or business branch office, where (for example) an SD-WAN or Virtual Network Function appliance is deployed.
- To wireless carriers, it may refer to a cell tower that has been enhanced with private compute facilities—or to the customer premises where a 5G device is deployed.

The edge may also be a moving target. To providers of "smart car" and "smart city" technology, the edge may be an automobile. To marketers or mobile app providers, the edge may be a user's smartphone or other personal mobile devices (e.g., wearables).

Despite the disparate perspectives, a common view of edge computing emerges: **the edge is the most logical place, outside a traditional data center or cloud, to bring together data and data-intensive applications for optimal price-performance**. The edge location (where the application software is hosted and data is ingested) should be as close as necessary to the user (human or machine) to minimize the volume and distance that data must be transmitted.

Configuring the Edge

Edge solutions—regardless of the application they support—require a specialized configuration of IT infrastructure, network, management platform, and security.

IT Infrastructure

Sensor-equipped edge device: The outermost rim of the "edge" is some sort of sensor or device that contains some sort of processing and networking capability. The device collects data, and does one or more of three things with it: processes it, stores it, and/or transmits it elsewhere for additional action. For many use cases, the sensors themselves are not physically large enough to hold the components required to do much processing. Instead, the devices may subject data points to a yes/no condition (e.g., is the temperature above x degrees?), and forward anomalous data to a control server for action. Or they may tally data points (for example, research measuring dust particles), and forward results. Or capture all data and forward immediately to a local server (e. g., measuring seismic activity in the ocean).

Edge server: The processing components that yield insight are often too bulky to be supported in sensor-embedded devices. As a result, most edge solutions place a local server within a short range, to handle the first layer of data processing. While the edge server could potentially be a standard server in a typical data center or server room, most edge solutions are remote from standard IT infrastructure. Compared with purpose-built edge servers, traditional servers are like hothouse flowers, requiring certain temperatures and conditions to operate efficiently. In contrast, the specialized edge server is likely to be hardened to withstand weather (e. g., on a truck farm, a traffic intersection, or the base of a cell tower); or shaking (e. g., on a ship, airplane, or factory floor). Dell/EMC and HPE are among providers that offer a range of purpose-built edge servers that are configured to specific application needs—for example, high capacity compute or data throughput.

This close-to-the-source compute layer is sometimes called "fog computing," to distinguish it from the "cloud computing" layer.[4]

Cloud or private data center: The next tier of the edge solution is the enterprise or cloud data center that stores and processes the collected data (or metadata). To yield valuable insight, the data center or cloud destination must be scalable to handle the volume and velocity of data, aggregate data from multiple sources, expose the data to AI or analytics software, and be accessible to applications and users. The configuration may include a co-location or interconnect facility, which may be a final terminus for data processing and storage, or a way station that links to a cloud.

Network

A few years ago, the best-effort internet was expected to handle the demands of IoT and edge applications, transmitting data from remote locations to the cloud. Today, IoT and edge devices are increasingly configured with network gateways that route traffic over private networks, where quality and performance can be managed to meet the needs of the application. Equinix reports that private and direct interconnection bandwidth will reach 8200Tbps, or 32ZB, by 2021—a compound annual growth rate of 48%.[5] The private network interconnection growth rate exceeds the estimated 26% growth in internet (IP) traffic reported by Cisco.[6]

Management Platform

Edge solutions require one or more platforms to tally, process, and direct the data as needed for the application. Certain management functions, especially those associated with IoT applications, are usually hosted in the cloud or data center; these include connectivity management, service enablement, big data analytics, and storage. Other functions—particularly Operations Technology functions that monitor and control physical equipment or processes—remain close to the edge, in the form of a thin client. Functions found at the edge may include device management, data

[4] See Stratecast report, The Fog Rolls In: Network Architectures for IoT and Edge Computing (June 2016).

[5] Equinix, 2018 Global Interconnection Index

[6] Cisco, Global Visual Networking Index, 2018

processing, and real-time incident management. Edge management platforms generally fall under providers' Internet of Things portfolios.

At the top of the technology stack is the application software that accesses or leverages the data. The application software varies by use case and is beyond the scope of this report.

Security

Securing an edge solution can be complex and challenging, requiring protection of the device itself, as well as the data. Since edge devices are located external to a traditional data center facility, users cannot rely on the physical security measures that offer some access and data protection within their own data center (e.g., biometric access controls; fire suppression equipment; environmental monitoring). Furthermore, traditional endpoint security solutions (physical or virtual) may not be appropriate for capacity- or size-limited local devices. In addition, some specialized devices (such as medical devices) have closed or proprietary software environments that tightly control access and do not permit overlay software.

For businesses deploying an edge solution, security challenges include:

- How to ensure that ingested data is from legitimate, authorized sources—and not rogue sensors or noise?
- How to ensure the edge device is not subject to physical tampering or destruction by environmental factors (e.g., weather)?
- How to enforce privileged access controls, and supervise change management of the edge management platform?
- How to thwart cyberattacks?
- How to ensure data integrity, at the edge and in transit?
- How to minimize lost or corrupted data?

Enterprises may find few market-ready, end-to-end security solutions for their edge applications—instead, cobbling together a range of data and endpoint security solutions.

Working Off-Line: the Cloudless Edge

As noted, platforms that support IoT applications are generally hosted in the cloud, with gateways linking to the edge servers or devices. However, providers of some IoT solutions, including Microsoft Azure IoT, recognize that edge functions may need to work offline—for example, an oil rig in the middle of the ocean. To meet that need, the Azure IoT platform can be installed directly on the edge server, or in the cloud or data center, or both, for seamless transfer and processing of data. To support the need for AI processing at the edge, Microsoft has recently introduced Azure Databox Edge—a physical storage device for deployment at the edge data center. Using machine learning, the Databox Edge pre-processes data at the edge, and then transfers data to the cloud via a virtual gateway.

In some cases, the application does not require immediate in-depth analysis of data collected at the edge. For that purpose, AWS has built its Snowball Edge—the latest in its "Snowball" branded line of physical storage devices. Snowball Edge is a hardened device that ingests data from sensors, keeping it secure from tampering. The physical device is shipped off to an AWS data center for data migration into the AWS cloud. The simplicity of the solution is such that collecting and shipping the device does not require a knowledgeable technician; this makes it easier to be deployed in areas remote from the company data center.

Next Generation Edge: 5G will Make a Difference

It's possible that our ideas about the "edge" vs. the core, and edge as a place, will be completely disrupted by 5G wireless technology. 5G promises to eliminate the constraints associated with current wireless networks, such as limited coverage area per cell tower (which require clear lines-of-sight); the high cost and challenges involved in deploying physical infrastructure (such as small cells) in public areas; inconsistent performance (weather-related); network-induced latency; and traffic-related congestion that impacts the ability to connect to a cell tower consistently.

According to Verizon, 5G (which stands for "fifth generation" cell network technology) will provide ubiquitous high-speed bandwidth sufficient to handle high traffic volumes, with high levels of security and minimal latency. The service, which Verizon recently rolled out in select cities in the US, has the potential to redefine the edge by minimizing network-related constraints.

Consider the promise of 5G:

- Devices can transmit data longer distances with minimal network latency—perhaps directly to a cell tower, and then to the cloud, negating the need to deploy an edge processor.
- Large data volumes can be transmitted and processed in near-real-time, without significant bottlenecks.
- Data from any source, in any volume, can be collected and transmitted securely and consistently, ensuring data integrity.

In theory, at least, 5G brings the cloud (limitless processing and storage capacity) to the edge, streamlining architectures and simplifying deployments.

The Higher Goal: Configuring Compute, Storage, and Network Resources Optimally for the Workload

In the digital era, perhaps the best way to think about edge is not as a distinct architecture, but as one of many options for optimal IT deployments. Rather than discussing whether a workload should be in the cloud or at the edge, it may make more sense to ask, "What is the best way to deploy IT resources to meet needs?"—for every workload.

Equinix takes this approach with its series of Interconnection Oriented Architecture playbooks.[7] Like a "service oriented architecture" in application design, an IOA offers a framework for building a flexible, fully meshed hybrid environment in which people, locations, clouds, and data can be connected securely, efficiently, and cost-effectively. The framework applies not only to IoT applications, but also to many of the new ways data is utilized and distributed—including mobile apps, location-based services, machine-to-machine transmissions, remote workforce, and split workloads.

[7] Equinix, Architecting for the Digital Edge: IOA Playbook

In developing IOA, Equinix studied over 400 discrete customer use cases representing a variety of industries, data types and sensitivities, and geographies. From these, the company identified four standard use cases describing different types of networking challenges.

- **People** – The employees, customers, and partners that use the enterprise data or applications may be located anywhere, with different access options and devices. How can enterprises ensure optimal experiences?
- **Locations** – Whether the enterprise is expanding into emerging markets or collecting IoT data far from metro centers, location still matters. When does it make sense to regionalize data, and when should it be centralized?
- **Clouds** – In a hybrid or multi-cloud environment, workloads will be split among cloud services, to maximize efficiency and reduce costs. How can the enterprise share and transfer data without compromising performance or security?
- **Data** – Analytics software yields valuable business intelligence, but only if enterprise users have real-time visibility into diverse sources and types of data. How can the business connect data sources so that users and applications maximize data value, while also maintaining compliance with data sovereignty regulations?

With this approach, "edge" and "core" are not the only architectural options. Instead, enterprises will assess all applications and data, and deploy the optimal architecture for each.

Stratecast
The Last Word

As data becomes more valuable to business operations, organizations will seek to collect, process, analyze, and store it efficiently. For data that is collected far from the traditional data center—for example, in an IoT application—this may require an "edge" architecture, in which data is collected from sensors, minimally processed locally, and deployed to the cloud or data center for aggregation, further analysis, and storage.

But IoT is not the only use case that calls for an edge configuration. Enterprises are deploying hybrid environments—comprising public cloud, on-premises data center, and co-location or other third-party data center facilities—to ensure optimal cost-performance for all their applications. Each workload should be assessed and placed in an optimal IT environment according to requirements such as scaling; data access; throughput and processing speed; consistency; latency-tolerance; integration with other apps and functions; and cost.

Even today, enterprises are not constrained to choosing "edge" or "cloud" deployments. Their IT environments may comprise premises and public cloud; centralized and distributed (branch) deployments; content distribution networks; hybrid workloads split across multiple environments; and third-party hosted services—with high degrees of integration and mobility across environments.

The challenge for enterprises will be to design, implement, and maintain their far-flung edge deployments, without technical resources at the edge sites. While providers are introducing low-maintenance hardware and remote management platforms, service providers have yet to introduce fully-managed end-to-end edge solutions that run from device to cloud.

But change is coming. Devices are becoming smarter, with more processing and storage capacity and embedded analytics functionality; this lessens the dependency on transmitting data to the cloud for processing. Conversely, 5G wireless networking may remove the constraints of the network link, enabling more data to go directly to the cloud for processing and storage without introducing delay. Each advance will enable enterprises to derive more value from their data, fast and affordably.

Lynda Stadtmueller

Vice President– Network, Data Center & Cloud
Stratecast | Frost & Sullivan
lynda. stadtmueller@frost. com

Stratecast FROST & SULLIVAN

About Stratecast

Stratecast collaborates with our clients to reach smart business decisions in the rapidly evolving and hyper-competitive Information and Communications Technology markets. Leveraging a mix of action-oriented subscription research and customized consulting engagements, Stratecast delivers knowledge and perspective that is only attainable through years of real-world experience in an industry where customers are collaborators; today's partners are tomorrow's competitors; and agility and innovation are essential elements for success. Contact your Stratecast Account Executive to engage our experience to assist you in attaining your growth objectives.

About Frost & Sullivan

Frost & Sullivan, the Growth Partnership Company, works in collaboration with clients to leverage visionary innovation that addresses the global challenges and related growth opportunities that will make or break today's market participants. For more than 50 years, we have been developing growth strategies for the Global 1000, emerging businesses, the public sector and the investment community. Is your organization prepared for the next profound wave of industry convergence, disruptive technologies, increasing competitive intensity, Mega Trends, breakthrough best practices, changing customer dynamics and emerging economies? For more information about Frost & Sullivan's Growth Partnership Services, visit http://www.frost.com.

CONTACT US

For more information, visit www.stratecast.com, dial 877-463-7678, or email inquiries@stratecast.com.