Analyst Brief, sponsored by:
Penguin Solutions

Edge Computing: Critical to Driving Digital Operations

Written by:

Dave McCarthy
Research Vice President, Cloud and Edge Infrastructure Services

Published:

June 2025

Edge computing is a critical element of powering digital business and operations. New demands for performance and resilience have made the physical location of compute resources an important factor in IT decisions.

Introduction

The term edge computing is firmly entrenched in the vocabulary of CIOs, technology providers, and service providers. The drive to automate processes and extract value from data has created a need for compute infrastructure and connectivity to many new locations. Industry conferences frequently feature servers and management platforms to address the challenges of highly distributed operations and often harsh environments. SKUs for purpose-built edge products are available on technology provider websites. End users are embracing hybrid and multicloud in their strategic planning. They are increasingly gaining awareness of the challenges of deploying and managing IT resources outside of the datacenter.

Several years ago, the market buzz started about edge computing and was quickly met with a healthy degree of skepticism. CIOs had a burning question: How would edge computing enable or support business objectives and deliver new business value? The answer lies in how edge computing improves overall digital operations.

End users today classify three-quarters of their edge workloads as highly critical. If these edge resources experience downtime, the consequences range from severe business interruptions to safety risks. This signals the increased reliance on edge resources and the need to ensure edge resilience. Edge infrastructure has evolved from supporting remote office and back-office applications to supporting digital business and operations.

Across all industries, organizations recognize that a centralized computing model is not effective in all situations. Despite datacenter modernization and cloud migration efforts, there are situations where it is desirable to move processing closer to where the data is being generated. This is the role of edge computing.

At a glance

Key takeaways

Edge computing is the foundation for digital businesses and operations. It leverages many recent advancements in core and cloud computing by extending those capabilities into remote offices and field locations.

Agility, resiliency, and predictability are key elements of creating data-driven operations. Edge-native technologies deliver these qualities by supporting multiple platforms, integrating with OT systems, and anticipating the unique environmental requirements at the edge.

What is edge computing?

Edge computing is a distributed computing paradigm that deploys processing and storage resources outside centralized datacenter and cloud infrastructure.

In many ways, edge computing is the evolution of the hybrid computing model. Just as IT departments consider which workloads to run on premises versus in the cloud, edge computing challenges them with the concept of distributing applications to remote locations. It cements the belief that hybrid infrastructure is not a temporary state. Instead, it will be a part of systems architecture for the foreseeable future. A mix of core and edge infrastructure can unlock new capabilities for the enterprise more effectively than a centralized approach alone.

Driving factors

The motivation for edge computing can be attributed to several factors. While not mutually exclusive, each of the following factors represents a problem that core computing cannot solve alone:

- **Latency-sensitive applications:** Whether introduced by the network or the number of hops between the endpoint and server, latency represents a delay. Many use cases exist where the round trip to the cloud or datacenter provides an unacceptable response time.
- **Assets with limited connectivity:** Internet of Things (IoT) devices or other connected equipment can be mobile. In these situations, the assets may move in and out of coverage areas or have limited bandwidth.
- **Cost:** As the amount of data generated in remote locations increases, the costs associated with transmitting it to a central data store also grow. In the case of wireless, this could result in extra charges for data transferred or having to opt for more expensive connections over low-bandwidth alternatives. The same is true for wired connections that must be upgraded to handle the additional data flow.
- **Scalability:** IoT initiatives are quickly moving from proof-of-concept trials to production deployments, and the influx of data can overwhelm core infrastructure. This is especially relevant for telecommunications providers deploying 5G, which increases the speed at the network's edge. As applications take advantage of faster access, potential bottlenecks can emerge.
- **Security and compliance:** Whether due to government regulation or corporate governance, there can be restrictions to where data may reside. As governments continue to pursue data sovereignty legislation, businesses are challenged with compliance. Cloud or datacenter infrastructure might be located outside the local jurisdiction. The ability to classify data and apply data management policies is critical.

These factors can often overlap. In the case of predictive maintenance, analytics models are derived from historical information on previous failure modes. Real-time telemetry data is evaluated against those models to detect anomalies or events.

While the models are trained with high-fidelity data in a central location, the real-time inference is typically performed on a subset of the available data in the field. This is due to cost and scalability concerns, which result in only a fraction of the available data being collected and transmitted to the cloud or centralized datacenter. This compromise can affect the quality of decision-making. Combine that with the latency inherent in wide-area communications, there is a good chance the system will not achieve the desired outcome.

By taking a different approach and incorporating edge computing into this scenario, it is possible to mitigate these problems. Computing resources deployed locally to the equipment can access all the available data without incurring communication costs and achieve near-real-time results. It is then possible to find conditions and enact remediation without the delay of communicating with the core.

Elements of an edge solution

Edge solutions can vary based on capability and their intended function. However, there are common elements.

Hardware

This serves as the foundation for computing and storage resources. On the high end of capability, there are enterprise-class rack-based systems that consist of standalone, converged, or hyperconverged infrastructure. These systems are often chosen for environments with a high volume and velocity of data, running data analytics, or AI workloads.

There are also embedded-class systems, often called IoT gateways, although their function is not limited to IoT use cases. These are typically non-rackmount devices that run embedded operating systems with limited configurations. These devices can run applications to evaluate data from multiple endpoints. Sometimes, they are designed specifically for a vertical industry, supporting specialized protocols or data formats.

In either case, hardware designed for the edge can include features that support harsh operating environments in terms of temperature, dust, and vibration. This is important for situations where the physical environment does not share the same level of environmental control as a core datacenter.

Software

Several categories of software can be present in an edge solution. Depending on the hardware, it will be an operating system or operating environment that provides basic functions for running applications and connectivity.

Edge computing requires a strong emphasis on remote management and security. Since the equipment is often deployed in locations with little to no IT staff, it is important to have a framework that allows for low- or zero-touch provisioning of new resources, configuration changes, and software updates. It is also beneficial to consider automated orchestration of management tasks and proactive security measures.

Applications are another aspect of edge software. Both IT and operational technology (OT) ISVs include edge components that work in conjunction with core software. This is especially true for data analytics, real-time dashboards, and operational solutions, where there is a benefit to distributing some application functions to the edge. This software may be pre-validated to ensure compatibility with edge hardware.

Consulting and systems integrator services

Many enterprises are new to edge computing, and solutions integrators and consulting organizations often play a role in developing a strategy, designing an architecture, and implementing the overall solution. They can also be involved in post-installation managed services and support.

Considerations for edge-native digital operations

Organizations continue to rapidly invest in digital infrastructure that resides closer to the edge. Across several industries, edge technologies are reshaping customer interactions, improving product quality, increasing process efficiency, improving worker safety, and reducing equipment downtime, to highlight a few use cases.

Edge computing infrastructure and platforms are increasingly purpose-built for highly distributed deployments that span a variety of location types. IDC refers to this emerging set of capabilities as edge native, in which applications and platforms are designed and built to understand and overcome the challenges of deploying, managing, and securing large numbers of highly distributed and diverse infrastructure and applications at the edge.

When organizations begin to deploy edge-native solutions to all locations where data needs to be gathered, analyzed, and protected, they face many challenges, including the following:

- Ensuring the frictionless flow and use of data across core, cloud, and edge datacenters
- Securing data and infrastructure
- Optimizing applications to run on heterogeneous platforms and connectivity types
- Enabling remote visibility and management
- Managing containers and virtual machines (VMs) across cloud, core, and edge locations
- Deploying and maintaining digital infrastructure in remote locations with no IT staff

- Understanding and complying with regional data security and privacy regulations
- Protecting digital infrastructure from the surrounding environment

Edge native also encompasses the connectivity, resilience, and security requirements that are inherent in and unique to highly distributed architectures. Networking must be designed to work with various hardware types and connectivity modes. Resilience must be created with autonomous and self-healing infrastructure.

Edge-native applications are optimized for the challenges at the edge. They are maximized for field use, adapted to ensure resilience, and designed to support mobility. Edge-native applications run on highly orchestrated container and virtual machine platforms, leveraging both zero trust security and zero-touch operational models.

Conclusion

Edge computing is the foundation for digital businesses and operations. It leverages recent advancements in core and cloud computing by extending those capabilities into remote offices and field locations. Edge computing's distributed nature creates more flexibility in deployment architectures, enables faster response times to rapidly changing conditions, and addresses many scalability problems associated with IoT use cases.

IDC expects investments in edge computing will continue to grow for years to come.

Edge computing is complementary to other technology trends such as IoT, AI, and 5G. In fact, most use cases require several of those technologies working together. IoT enables the collection of data from sensors and connected equipment in the field. AI running on edge computing infrastructure can evaluate that data, identify conditions, and automate actions. 5G increases the speed that events and data are communicated back to the enterprise. Each adds value and enables businesses to make quicker, more accurate decisions based on real-time information.

As with any technology initiative, it is imperative to understand and clearly articulate the business objective for edge computing and a set of measurable success criteria. It is then possible to map out the required components to achieve it, which will likely consist of a mix of core and edge computing resources. The exact architecture will depend on the number of endpoints, amount of data generated, required processing capability, and a determination of where infrastructure will be deployed.

Most hardware and software suppliers have invested in enhancing their product portfolios with features designed for edge computing. Those solutions are pre-integrated and pre-validated and typically include a common management interface for managing core and edge resources.

All industries are adopting digitally infused or augmented processes to streamline operations, increase efficiency, and improve outcomes. These industries are in the process of extending digitally infused processes to all locations. They seek to enable the frictionless flow of data among all platforms and locations — including cloud, core, and edge locations. To accomplish this, they will need edge-native platforms that make the digital journey more secure, scalable, and resilient.

Agility, resiliency, and predictability are key elements of creating data-driven operations. Edge-native technologies deliver these qualities by supporting multiple platforms, integrating with OT systems, and anticipating the unique environmental requirements at the edge.

IDC expects investments in edge computing will continue to grow for years to come. Distributing workloads between core, edge, and intelligent endpoints creates a responsive architecture that reduces bottlenecks and can rapidly adapt to changing business requirements.

About the IDC analyst

Dave McCarthy
Program Vice President, Global Lead, Cloud and Infrastructure Services

Dave McCarthy leads IDC's cloud and infrastructure services global research subdomain with two primary focuses. The first is cloud infrastructure and its related adoption strategies: public, private, hybrid, multicloud, distributed, sovereign and edge. The second is infrastructure services: flexible consumption, deployment, support, and the circular economy. Benefiting both technology suppliers and IT decision makers, Dave's insights delve into ways in cloud and infrastructure services provide the foundation for both general purpose and AI workloads, enabling organizations to innovate faster, create new revenue streams, and achieve competitive advantages. His research is available via syndicated research programs (subscription services), data products (IDC Trackers) and custom engagements.

More about Dave McCarthy

Message from the sponsor

About Penguin Solutions, Inc.

As trusted advisors accelerating time to value and delivering peace of mind to our valued customers, Penguin Solutions has over two decades of experience providing end-to-end solutions that solve complex challenges in computing and memory.

We deliver high performance and high availability compute infrastructure solutions and services. We are experts in the infrastructure required to successfully deploy and run data intensive and critical workloads from edge to core to cloud — most notably artificial intelligence (AI), high-performance compute (HPC), fault-tolerant (FT), and edge computing infrastructure.

Our Stratus high availability and fault tolerant computing platforms ensure the continuous availability of our customers' critical applications and data in remote data centers and edge locations. For more information, visit https://www.penguinsolutions.com.

IDC Custom Solutions

IDC

IDC Research, Inc.

One Beacon Street, Suite 33100, Boston, MA 02108, USA

T +1 508 872 8200

blogs.idc.com | LinkedIn @IDC | www.idc.com