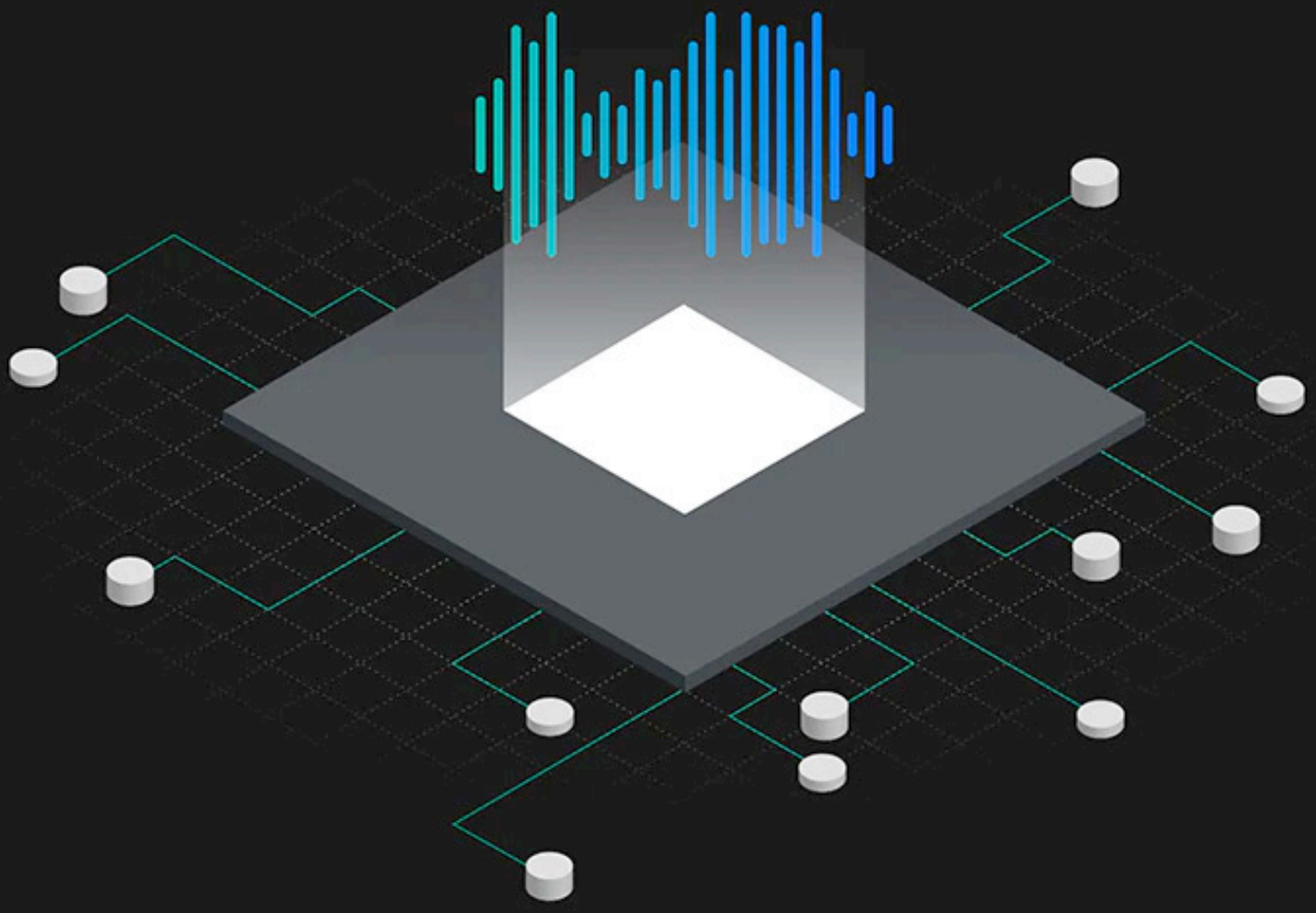
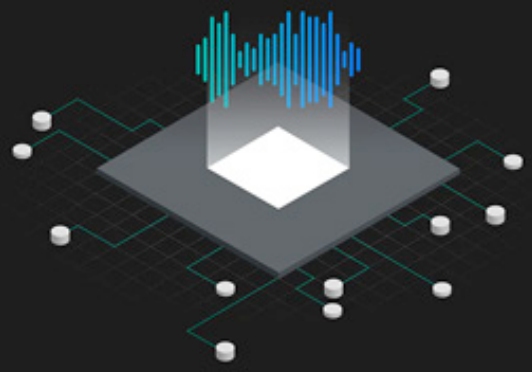
IBM Cloud

IBM Edge Computing

Rob High
IBM Fellow, Vice President and CTO,
IBM Edge Computing



Executive summary



From cars to manufacturing equipment to ATMs to mining equipment, intelligent devices are being infused into the very fabric of the tools we use to conduct business. Their compute capacity creates new opportunities to bring analytics to where data is first created and actions are taken. Innovations around edge computing can fuel improved quality, enhance performance, and drive deeper, more meaningful user interactions.

Edge computing can:

- **Solve new business problems with AI** - Modern devices at the point of data capture have their own analytic capabilities. Move compute resources closer to where data originates and leverage AI to solve new business problems with less latency and reduced data transmission.

- **Increase capacity and resiliency** - Moving compute and data analysis to edge devices increases the system's overall analytics capacity. Edge devices can run container technology natively, maximizing the cloud-native programming skills of your company's developers.

- **Gain better security and privacy protections** - By processing data closer to its source, less data is transmitted across networks, reducing the potential attack surface and easing the ability to enforce enterprise policies at the source of data creation.

- **Leverage 5G network's reduced latency** - With the adoption of 5G networks, business processes can take advantage of localized data analytics to feed automated decisions through centralized AI.

Forward-looking companies want to unlock the potential of untapped data created by the growth of connected devices, enabling new business opportunities, increasing operational efficiency and improving customer experiences. Edge computing brings enterprise applications closer to where the data is created and actions need to be taken, allowing enterprises to leverage AI and analyze their data in near-real time.

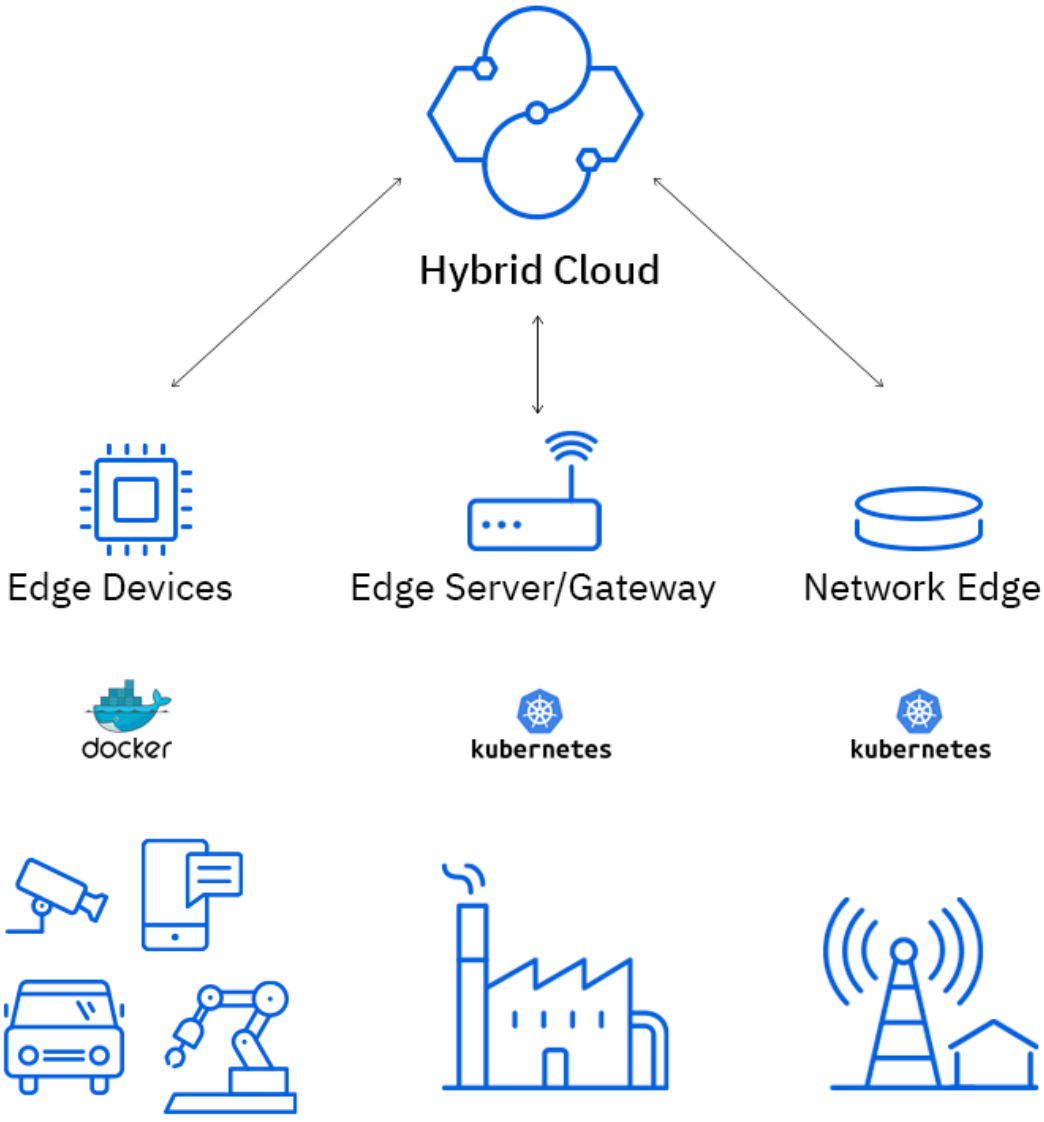
Imagine that you're a manufacturer and you're trying to deal with the potential cost for outages in your production line–costs that might occur if any of your equipment goes down and stops the production process. Or imagine that you are an automobile industry manufacturer trying to engage your clients, your drivers, and your passengers with a better driving experience. These examples can be enhanced and extended using analytics and AI to improve the overall experience.

In the past, the promise of cloud and AI was to automate and speed innovation by driving actionable insight from data. However, the scale and complexity of data created by connected devices may strain network and infrastructure capabilities. IDC reports that by 2025, each connected person will have at least one data interaction every 18 seconds[1]; this underscores the potential of meaningful insights created from data captured closer to its source.

Edge Computing Benefits

Edge computing helps solve these issues of speed and scale. By leveraging the computational capacity of edge devices, gateways, and networks, you retain the principles of dynamic allocation of resources and continuous delivery that are inherent to cloud computing. With edge computing, today's businesses have the potential to virtualize the cloud beyond the four walls of their data center. Workloads created in the cloud, including some of the more modern forms of AI and analytics, can now be migrated out towards the edge; where appropriate, data generated at the edge can be cleansed and optimized and brought back to the cloud.

IBM has a view of edge computing that spans many industries and multiple tiers optimized with open technologies and standards like Docker and Kubernetes.



The IBM edge computing platform spans private cloud and enterprise environments, network compute spaces and even further to on-premise gateways, controllers and servers, and, finally, intelligent devices such as robots, connect vehicles, etc.)

Centrally, the hyper-scale public clouds, hybrid clouds, co-located managed data centers and traditional enterprise IT data centers will continue to serve as an aggregation point for data, analytics and back-end data processing.

Public, private and content-delivery networks are transforming from simple pipes to higher-value hosting environments for applications–a form of edge network cloud.

Edge Computing risks and responsibilities

As discussed in the prior sections, while the introduction of edge computing creates unique opportunities, it also presents some challenges. First, it breaks down the neat physical boundaries of the cloud data center–forcing us to think about issues of security, addressability, management, ownership, and compliance. More importantly, it multiplies the scaling issues of cloud-based management techniques.

Edge networks increase the number of compute nodes by an order of magnitude. Edge gateways increase that by another order of magnitude. Edge devices increase that magnitude even more. If DevOps (continuous delivery/continuous deployment) is critical to managing a hyper-scale cloud infrastructure, then zero-ops (that is, operations without *any* human intervention) is critical to managing at the massive scale that edge computing represents.

At that scale, change is constant and explosive. Networks are being reconfigured to work around congestion on a continuous basis. Edge gateways are being updated with new features and processes. Edge devices are being moved around, changing ownership, and being re-purposed or re-prioritized. This level of churn is practically business as usual. Moreover, for intelligent equipment (edge devices) deployed in today's businesses, equipment is often shared, so a user cannot be expected to manage the compute available on it.

What's involved in edge computing?

Hybrid Cloud Computing – Traditional hyper-scale, public clouds such as IBM Cloud and other cloud providers like Microsoft, Amazon, and Google as well as private clouds deployed in co-locations and on-premises IT data centers.

5G Network – During the transition to 5G, many public network providers are expanding their infrastructures to include general-purpose computing services. The edge network itself is potentially multi-tiered–composed of regional data centers, central offices and hub micro-data centers. Telcos are transforming these tiers in their core network to host application workloads using cloud technologies within the network edge.

Edge Servers – Servers, gateways and controllers acting as edge servers are often deployed in factories, warehouses, hotels, and retail stores to provide local compute capacity for operations. These resources may or may not be clustered, but still support critical business processes.

Edge Devices – The number of devices that contain enough computational capacity to do work is growing rapidly [2]. These devices commonly have sufficient CPU power, RAM and local storage to run a Linux operating system.

IoT Devices – Most traditional IoT devices are closed, fixed-function devices. They are typically integrated with sensors for collecting data that is transmitted upstream to other aggregation points–traditionally the cloud.

Mobile Devices – Mobile devices play an important role in edge networks. They are distinct from other edge devices because they typically belong to an individual who assumes personal responsibility for it, and the mobile device runs iOS or Android operating systems, which may refuse to run container software that was not acquired through their app stores.

Need a primer? Watch the video,
What is edge computing?

It must be possible to deploy, update, monitor and recover the edge compute space without human intervention. All activities and processes must be fully automated, able to make decisions on their own about what work needs to be placed where and able to recognize and recover from changing conditions without intervention. All placement activities should to be secure, traceable and defensible. The system must have a deep awareness of the nature, location and purpose of different devices with different capabilities and different uses and be able to use that awareness to make informed, policy-driven decisions.

These are all issues that need to be considered and addressed when enabling the other advantages of edge computing. IBM is addressing these capabilities with its introduction of IBM Edge Application Manager.

Extending multicloud deployments to the Edge

Last year, IBM announced Cloud Pak for Multicloud Management, which unifies cloud platforms from multiple vendors into a consistent dashboard from on-premises to the edge. IBM Edge Application Manager is a natural extension, enabling the distribution and management of workloads beyond the edge network – out to edge gateways and edge devices.

Of course, an edge platform is only as useful as the extent of the ecosystem it supports. That's why IBM Edge Application Manager recognizes workloads from enterprise applications with edge components, private and hybrid cloud environments, plus public cloud where edge computing provides a new execution environment for distributed AI to reach key and time-critical data sources. Applications developed on Amazon, Microsoft, Google and other cloud providers supporting container technology are now candidates for IBM Edge Computing workloads.

In addition, IBM brings AI tools for accelerated deep learning, visual and speech recognition, and video and acoustic analytics, enabling inferencing on many resolutions and formats of video and audio, conversation services and discovery to advance the creation of sophisticated enterprise applications. IBM also brings deep domain expertise and industry-leading solutions, such as asset performance management, public safety, intelligent locations and mobility, financial services and retail. These solutions incorporate advances in AI and analytics, as well as leverage the benefits of the distributed edge computing topology to maximize their performance, utility and protection of private user and enterprise data.

Resources

What is edge computing? blog and video explainer (10:39)

IBM Edge Application Manager
and video What is IBM Edge Computing? (2:36)

The Importance of Effective Operations in Unlocking Edge IT Value (IDC)

IBM Cloud Paks

5G Edge Computing Whitepaper
(FCC Technical Advisory Council)

Summary

Unlock the potential of your untapped data created by the growing number of connected devices, enabling new business opportunities, increasing operational efficiency and improving customer experiences. IBM Edge Computing and its companion edge computing solutions can help your company:

- Enable transformation in the telecommunication, manufacturing, retail, automobile and many other industries.
- Enable deployment of AI and analytics to edge devices, gateways, operations controllers and other compute locations.
- Facilitate the emergence of 5G and position players to capture higher value within their network infrastructure through the virtualization of network functions and through the creation of new compute opportunities for enterprise solutions.

IBM is your trusted partner, offering the industry expertise to deliver an open and intelligent edge solution that enables enterprises to build, distribute and manage applications at scale.

IBM Hybrid Cloud
IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
February 2020