What's the Difference Between Edge Computing and Cloud Computing?

Author: Tiffany Yeung

Public cloud computing platforms allow enterprises to supplement their private data centers with global servers that extend their infrastructure to any location and allow them to scale computational resources up and down as needed. These hybrid public-private clouds offer unprecedented flexibility, value and security for enterprise computing applications.

However, AI applications running in real time throughout the world can require significant local processing power, often in remote locations too far from centralized cloud servers. And some workloads need to remain on premises or in a specific location due to low latency or data-residency requirements.

This is why many enterprises deploy their AI applications using edge computing , which refers to processing that happens where data is produced. Instead of cloud processing doing the work in a distant, centralized data reserve, edge computing handles and stores data locally in an edge device. And instead of being dependent on an internet connection, the device can operate as a standalone network node.

Cloud and edge computing have a variety of benefits and use cases, and can work together.

According to research firm Gartner, "cloud computing is a style of computing in which scalable and elastic-IT-enabled capabilities are delivered as a service using Internet technologies ."

There are many benefits when it comes to cloud computing. According to Harvard Business Review's "The State of Cloud-Driven Transformation" report, 83 percent of respondents say that the cloud is very or extremely important to their organization's future strategy and growth.

Cloud computing adoption is only increasing. Here's why enterprises have implemented cloud infrastructure and will continue to do so:

Lower upfront cost – The capital expense of buying hardware, software, IT management and round-the-clock electricity for power and cooling is eliminated. Cloud computing allows organizations to get applications to market quickly, with a low financial barrier to entry.

Flexible pricing – Enterprises only pay for computing resources used, allowing for more control over costs and fewer surprises.

Limitless compute on demand – Cloud services can react and adapt to changing demands instantly by automatically provisioning and deprovisioning resources. This can lower costs and increase the overall efficiency of organizations.

Simplified IT management – Cloud providers provide their customers with access to IT management experts, allowing employees to focus on their business's core needs.

Easy updates – The latest hardware, software and services can be accessed with one click.

Reliability – Data backup, disaster recovery and business continuity are easier and less expensive because data can be mirrored at multiple redundant sites on the cloud provider's network.

Save time – Enterprises can lose time configuring private servers and networks. With cloud infrastructure on demand, they can deploy applications in a fraction of the time and get to market sooner.

Edge computing is the practice of moving compute power physically closer to where data is generated, usually an Internet of Things device or sensor. Named for the way compute power is brought to the edge of the network or device, edge computing allows for faster data processing, increased bandwidth and ensured data sovereignty.

By processing data at a network's edge, edge computing reduces the need for large amounts of data to travel among servers, the cloud and devices or edge locations to get processed. This is particularly important for modern applications such as data science and AI.

According to Gartner, "Enterprises that have deployed edge use cases in production will grow from about 5 percent in 2019 to about 40 percent in 2024 ." Many high compute applications such as deep learning and inference, data processing and analysis, simulation and video streaming have become pillars for modern life . As enterprises increasingly realize that these applications are powered by edge computing, the number of edge use cases in production should increase.

Enterprises are investing in edge technologies to reap the following benefits:

Lower latency : Data processing at the edge results in eliminated or reduced data travel. This can accelerate insights for use cases with complex AI models that require low latency, such as fully autonomous vehicles and augmented reality.

Reduced cost : Using the local area network for data processing grants organizations higher bandwidth and storage at lower costs compared to cloud computing. Additionally, because processing happens at the edge, less data needs to be sent to the cloud or data center for further processing. This results in a decrease in the amount of data that needs to travel, and in the cost as well.

Model accuracy : AI relies on high-accuracy models, especially for edge use cases that require real-time response. When a network's bandwidth is too low, it is typically alleviated by lowering the size of data fed into a model. This results in reduced image sizes, skipped frames in video and reduced sample rates in audio. When deployed at the edge, data feedback loops can be used to improve AI model accuracy and multiple models can be run simultaneously.

Wider reach : Internet access is a must for traditional cloud computing. But edge computing can process data locally, without the need for internet access. This extends the range of computing to previously inaccessible or remote locations.

Data sovereignty : When data is processed at the location it is collected, edge computing allows organizations to keep all of their sensitive data and compute inside the local area network and company firewall. This results in reduced exposure to cybersecurity attacks in the cloud, and better compliance with strict and ever-changing data laws.

Both edge and cloud computing can take advantage of containerized applications. Containers are easy-to-deploy software packages that can run applications on any operating system. The software packages are abstracted from the host operating system so they can be run across any platform or cloud.

The main difference between cloud and edge containers is the location. Edge containers are located at the edge of a network, closer to the data source, while cloud containers operate in a data center.

Organizations that have already implemented containerized cloud solutions can easily deploy them at the edge.

Often, organizations turn to cloud-native technology to manage their edge AI data centers . This is because edge AI data centers frequently have servers in 10,000 locations where there is no physical security or trained staff. Consequently, edge AI servers must be secure, resilient and easy to manage at scale.

Learn more about the difference between developing AI on premises rather than the cloud .

Edge and cloud computing have distinct features and most organizations will end up using both. Here are some considerations when looking at where to deploy different workloads.

An example of a situation where edge computing is preferable over cloud computing is medical robotics, where surgeons need access to real-time data. These systems incorporate a great deal of software that could be executed in the cloud, but the smart analytics and robotic controls increasingly found in operating rooms cannot tolerate latency, network reliability issues or bandwidth constraints. In this example, edge computing offers life-or-death benefits to the patient.

Discover more about what to consider when deploying AI at the edge .

For many organizations, the convergence of the cloud and edge is necessary. Organizations centralize when they can and distribute when they have to. A hybrid cloud architecture allows enterprises to take advantage of the security and manageability of on-premises systems while also leveraging public cloud resources from a service provider.

A hybrid cloud solution means different things for different organizations. It can mean training in the cloud and deploying at the edge, training in the data center and using cloud management tools at the edge, or training at the edge and using the cloud to centralize models for federated learning. There are limitless opportunities to bring the cloud and edge together.

Learn more about NVIDIA's accelerated compute platform , which is built to run irrespective of where an application is — in the cloud, at the edge and everywhere in between.

Dive deeper into edge computing on the NVIDIA Technical Blog .

Original URL: https://blogs.nvidia.com/blog/2022/01/05/difference-between-cloud-and-edge-computing/