Why Edge? Why Now?

How the Rise of Edge Computing will Reshape the Data Center Landscape

In the next decade, we will continue to see skyrocketing growth in the number of IP-connected mobile and machine-to-machine (M2M) devices, which will handle significant amounts of IP traffic. Tomorrow's consumers will demand better online experiences and faster, smoother Wi-Fi service and application delivery from online providers. Also, some M2M devices, such as autonomous vehicles (self-driving cars), will require real-time communications with local processing resources to guarantee safety and avoid accidents.

Unfortunately, today's IP networks are not set up to handle the high-speed data transmissions that tomorrow's connected devices will require. In a traditional IP architecture, data must often travel hundreds of miles over a network between end users or devices and cloud resources. This results in latency, or slow inefficient delivery of time sensitive data.

The solution to reducing latency lies in edge computing. By establishing IT deployments for cloud-based services in edge data centers located in localized areas or communities, we effectively bring IT resources closer to end users and devices. This helps us achieve efficient, high-speed delivery of applications and data. Edge data centers are typically located on the edge of a network, with connections back to a centralized cloud core.

Many new technologies will utilize and benefit from edge data centers, including fifth generation (5G) networks, Internet of things (IoT) and Industrial Internet of things (IIoT) devices, autonomous vehicles, virtual and augmented reality, artificial intelligence and machine learning, data analytics, and video streaming and surveillance. High-tech providers who invest in the deployment of edge data centers today will achieve a competitive edge for tomorrow, and will be able to offer faster, more reliable delivery of their services and applications.

SIEMON™

Introduction

As the connected world of the 21st century comes into existence, the next decade will give rise to a variety of technologies that will continue to transform and automate our everyday lives. This will include everything from 5G mobile networks, to self-driving cars, to smart cities and smart factories driven by IoT and IIoT devices. But to fully realize the potential of these technologies, we still need to overcome some key technical limitations.

A chief problem lies in the fact that today's IP networks are not set up to handle critical time sensitive application delivery and data processing. A traditional IP network relies on a centralized architecture, where IT cloud resources are housed in a data center that is often located hundreds or thousands of miles away from end users and devices. The distance and number of hops (i.e., switches and routers) that data must travel over the network between cloud resources and IP-connected devices often results in latency, or slow transmitting and processing. While the delays may only be a few milliseconds, they can add up to create unacceptable performance responses and an inability to support emerging technologies.

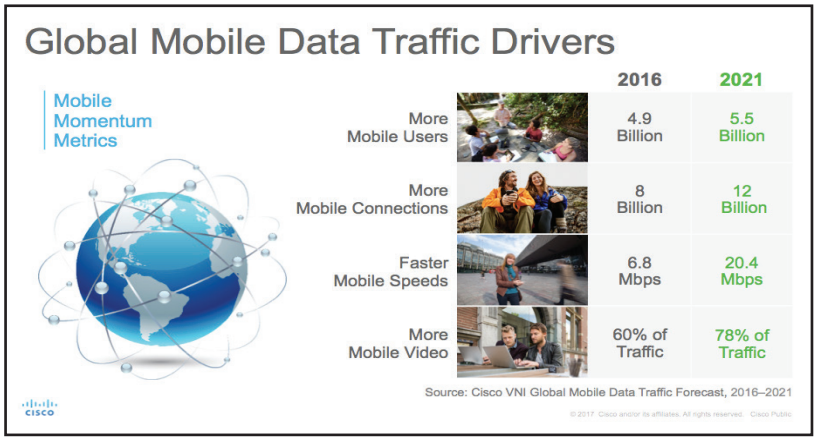
Under a centralized architecture, even tomorrow's 5G networks will not be fast enough to handle the real-time application delivery that end users will demand and IP-connected devices like self-driving cars will require. To support IoT, IIoT, and other next-generation technologies, we must expand data and network capacity, reduce latency, and achieve faster processing of data and applications.

So what's the solution? Instead of bringing the users and devices to the data center, we bring the power of the data center to the users and devices!

Edge computing relies on a distributed data center architecture, in which IT cloud servers housed in edge data centers are deployed on the outer edges of a network. By bringing IT resources closer to the end users and/or devices they serve, we can achieve high-speed, low-latency processing of applications and data. Companies and organizations that invest in edge computing today will have a competitive edge for tomorrow, and will be able to provide faster, smoother delivery of online services and better performance of online applications.

Edge Drivers

In the next few years, we will continue to see unprecedented growth, not just in the number of IP-connected mobile devices (smartphones, laptops, tablets, etc.), but also in the number of M2M devices where machines "talk" to each other by processing and exchanging data.

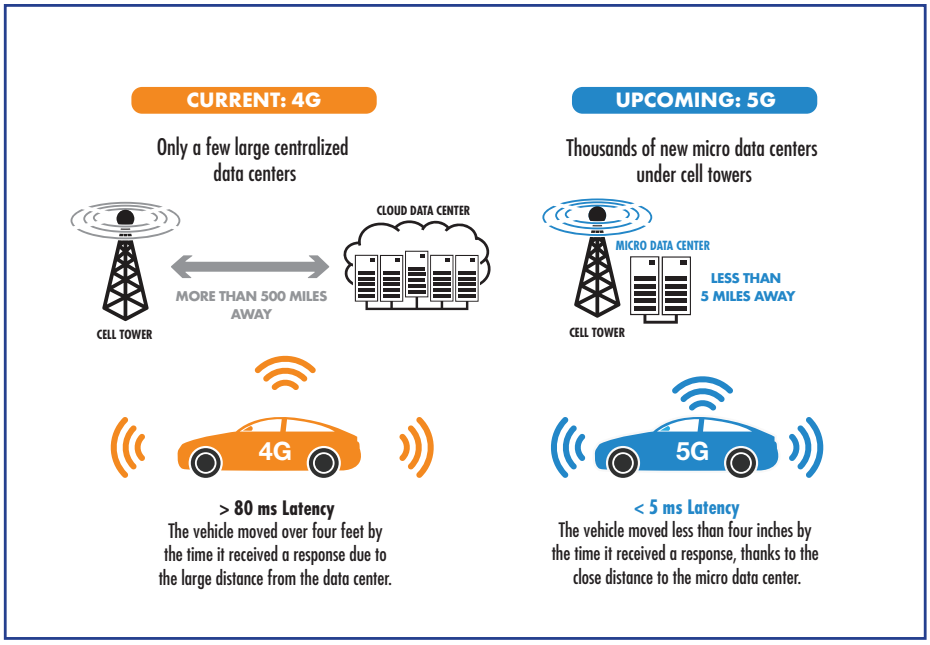


According to the Cisco Visual Networking Index report, by the year 2021:

- Cellular networks will carry 78% of mobile data traffic.
- Wireless and mobile devices will handle more than 63% of mobile IP traffic.
- Smartphone traffic will fully exceed PC traffic, and will account for 33% of total IP traffic, with an annual traffic growth rate of 49%.
- Tablets will have annual traffic growth rates of 29%.
- Internet TVs will have annual traffic growth rates of 21%.
- M2M modules will have annual traffic growth rates of 49%.

Cloud and content providers are now embracing edge computing as an efficient and cost-effective solution for supporting the high traffic levels of tomorrow's connected devices:

- An IDC report predicts that by the end of this decade, 45% of IoT-created data will be stored, processed, and analyzed at the edge of the network.
- A Gartner report predicts that within five years, 50% of enterprise-generated data will be created and processed outside of traditional centralized data centers or clouds.
- An ACG research study estimates that companies can save approximately $110 million in transmission costs over a 5-year period by caching content locally in a metro population.



The Problem of Latency

Latency is caused by a number of factors, including the physical distance that data transmissions must travel between centralized cloud servers and end user devices, the number of network hops between switches that the transmission has to make, and the amount of traffic on the network.

For consumers, latency is usually just an annoyance. It means a slow download of a movie to their laptop or tablet, or interference with the reaction time of the online video game they are playing on their smartphone. Occasionally, latency results in application failure due to transmission errors, and the application must be reloaded.

For IoT and M2M devices, network latency can be a major impediment, especially for devices that rely on guaranteed response times and real time processing of data and applications to handle life-or-death situations. A prime example is the autonomous vehicle, which will rely on interactions between internal and external devices to ensure the safety of the vehicle's passengers, other motorists, and pedestrians.

While autonomous vehicles will be required to carry very powerful internal safety systems via a combination of onboard computers and sensors, they cannot fully perform as independent, self-operating systems. They will always need to communicate with smart traffic signals and road sensors to receive information about current traffic conditions and possible safety threats.

For example, if an autonomous vehicle is approaching a blind intersection, the vehicle's sensors can't "see" around the corners of nearby buildings. Instead, a smart traffic light installed above the intersection would signal the vehicle as to whether it is safe to proceed. In this scenario, it would be impractical for the smart traffic light to send data over a long-distance network, back to a centralized cloud core due to latency.

What is Edge Computing?

Edge computing is the practice of deploying IT cloud resources, housed in edge data centers, in localized areas or communities. Edge computing operates on the proven principle that applications perform better and faster when processing is performed in close proximity to the end users and/or devices they serve. Bringing the data center to the end user or device results in the following advantages:

- Improved application performance and reliability, through localized computing, data storage, and data analytics.
- High-speed data processing and application delivery that overcomes network latency issues, and significantly reduces the possibility of data loss in transfer.
- A reduction in the number of network hops that data and applications must make between cloud servers and devices.
- Reduced data transportation costs.

Edge vs. Micro vs. Containerized Data Centers

A common misconception is that edge data centers and micro data centers are one and the same, or that "edge data center" is an interchangeable term for a micro or containerized data center. For example, if you have a standalone micro data center in a utility closet at your main office, you might mistakenly assume that it's also an edge data center. Or, if you place a containerized data center outside a manufacturing facility, you might assume it's an edge data center, because it sits on the outer edge of the building.

The term "edge" in "edge data centers" refers to the data center's location on the edge of a network, and/or its physical location in an area that is far removed from the main data center deployment. If it isn't part of a larger network that connects back to a central cloud core, it is not an edge data center.

It's more accurate to say that a modular, containerized, or micro data center can be used as an edge data center. The "edge" in "edge data centers" refers more to a data center's function or purpose, while the terms "modular," "containerized," or "micro" refer to the data center's physical construction or form factor.

In the previous example, an edge data center would be installed at the blind intersection to operate the smart traffic light. It might be a quickly-deployed enclosure on a nearby street corner, or even a small box at the base of the traffic light that holds a microprocessor. If sensors detect that another car is about to speed through a red light at the intersection, the servers in the edge data center would instantly process a warning application, directing the smart traffic light to signal the autonomous vehicle to activate its brakes. By reducing the distance that data must travel, you eliminate latency, allowing the application to operate in real-time to avoid a collision.

What is an Edge Data Center?

An edge data center is a self-functioning data center that holds localized IT deployments for cloud services, with compute, storage, and analytics resources for application processing and data caching.

It should be noted that an edge data center qualifies as an actual data center, as opposed to being just a Wi-Fi router or point of presence on the network. An edge data center includes the same power, cooling, connectivity, and security features that you find in a centralized data center, but on a smaller scale. Also, the IT deployments in an edge data center will handle processing of applications, data analytics, and data storage within the general vicinity of end users and devices that use those applications and data.
The following are examples of edge data centers:

- A containerized data center installed outside a Wi-Fi tower to bulk up local service on a 5G network.
- A Central Office Re-architected as a Data Center (CORD), where the servers provide enhanced Wi-Fi, streaming TV, and gaming services to local customers in a mid-sized city, small town or rural county.
- A modular, one-room data center that provides secure Wi-Fi service at an enterprise facility, such as a remote office park, medical complex, or branch college campus, with connections back to a larger, centralized data center at the enterprise's main headquarters or campus.
- A small data center that is placed inside or outside a manufacturing plant or other enterprise business to provide application and data processing for the facility's IIoT devices, with reporting of operational data to a central IT core located in an out-of-state data center.
- A micro data center (as previously mentioned) that is placed on a street corner to provide local service for connected traffic lights and autonomous vehicles.

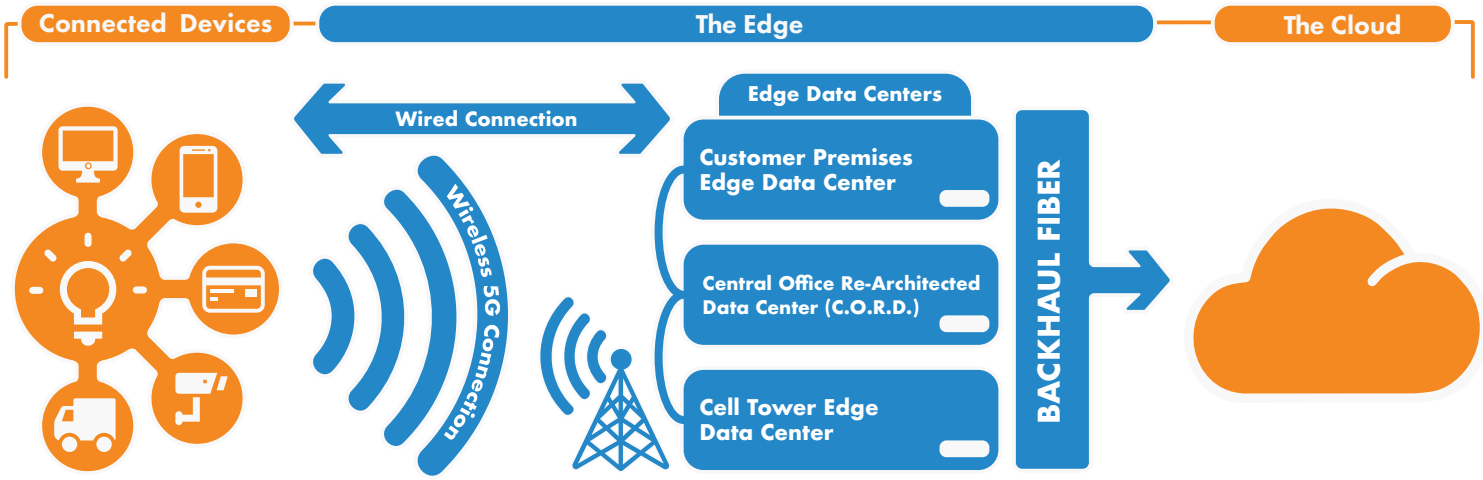
Edge data centers have several common characteristics that distinguish them from larger, centralized data centers:

Extensive Local Services – An edge data center provides extensive local services to users and/or devices within a certain geographic area. Depending on its purpose, the IT servers in the edge data center will typically provide local hosting and high-speed delivery of applications, data collection, caching, and storage.

Edge data centers are often deployed in a group, to provide local services to a larger area. For example, by placing edge data centers at 5-to-10 mile intervals in urban and suburban neighborhoods, you can ensure that there will be at least one data center located within a few miles of every end user or connected device.

Part of a Larger Network - An edge data center resides on the outer edges of an IP network. It connects back to a centralized cloud core in a data center that is usually located some distance away. Also, a group of edge data centers will often be connected to each other to form an aggregated edge or edge cloud which creates a shared pool of localized compute, storage, and network resources.

Smaller Size – Edge data centers tend to be smaller in size than regular data centers. For example, micro data centers may be used as edge data centers and may store only a few servers. Some larger edge data centers might be located in warehouses or office locations and might operate up to 50 cabinets.



Technologies Will Benefit

There are two types of technologies that will benefit from edge data centers – consumer technologies and industrial technologies. (In some of these categories, consumer and industrial uses for edge data centers may overlap.)

Edge Data Centers and Consumer Technologies

5G Networks – 5G mobile networks will offer very high bandwidth (up to 20 Gbps) and low latency (less than one millisecond). Edge data centers will need to be installed at the bases of 5G towers to improve localized service to mobile customers. (Due to the very high speeds that will be offered by 5G networks, consumers may be able to switch away from traditional in-home cable providers to their cellular carriers for entertainment needs.)

Autonomous Vehicles – As mentioned, establishing edge data centers on street corners and on the edges of highways will enable autonomous vehicles to receive real-time traffic information and emergency warnings from smart traffic signals and road sensors. Edge data centers will also allow self-driving cars to easily connect with centralized servers at their home company, to transmit automotive data, download application updates, schedule maintenance, etc.

Virtual Reality (VR) and Augmented Reality (AR) – Edge data centers will provide local computing resources for highly-complex VR and AR data processing, and high-speed delivery of applications (3-D rendering, machine vision) for services like VR meeting rooms, VR classrooms, and virtual city tours.

Artificial Intelligence (AI) and Machine Learning – Placing edge data centers in local communities will improve the performance and response times of AI and ML applications (i.e. GPS systems), allowing these applications to transmit, receive, and process data faster, and conduct interactions with users in real-time.

Online Video Games – Hosting video games on local servers housed in edge data centers will improve the gaming experience for players. It will increase game responsiveness to player commands, reduce latency and lag, and eliminate mid-game failures caused by application transmission errors.

Edge Data Centers and Industrial Technologies

Internet of Things (IoT) – Smart cities will feature large-scale deployments of millions of IoT devices, such as smart traffic lights, thermostats, and gas and water sensors. Local computing resources housed in edge data centers will enable operation of these IoT devices, running high-speed applications for public safety, data collection, and data processing.

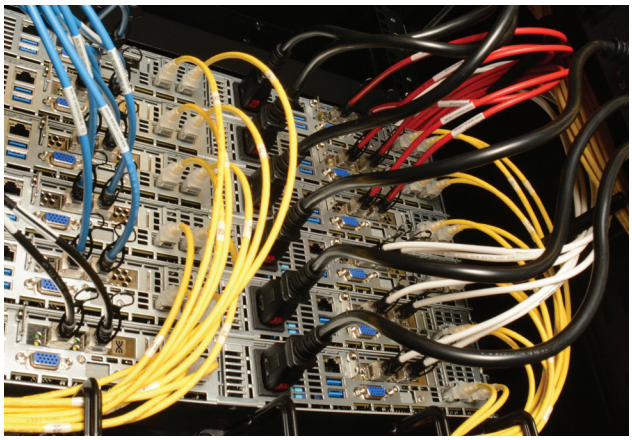
Industrial Internet of Things (IIoT) – Smart industrial facilities, such as manufacturing plants, power stations, oil platforms, and water and sewage treatment plants will have deployments of smart sensors and devices. Edge data centers will provide local computing resources, allowing IIoT devices to run industrial applications, perform automatic functions, and collect and process data.

Data Analytics and Big Data – Servers in edge data centers will provide localized data analytics, collecting and processing large amounts of data from IoT and IIoT devices (i.e. sensors in smart factories). Edge computing will also provide a faster means of collecting data from multiple locations and delivering it back to a centralized high-performance computing (HPC) cloud core for Big Data analysis.

Video Surveillance – Edge data centers will provide links to local security, law enforcement, or government databases, allowing smart video surveillance cameras to notify human operators of specific situations (i. e. a break-in or fight in progress).

Essential Infrastructure

Edge data centers take many forms – modular, containerized, micro, warehouse, or office-based – but all edge data centers require the same infrastructure elements that you find in larger, centralized data centers. Some of the essential elements include:



Fiber Optic Connectivity – All edge data centers will require high-density fiber optic connectivity solutions for high-speed, low latency transmissions. Fiber optic cabling provides the best option for connecting edge data centers back to cloud colo facilities, hyperscale data centers, and central offices to achieve speeds of 400 Gbps and beyond.

High-Speed Copper Cabling – Edge data centers use direct-attach, high-speed copper cabling such as category 6A to support 10 Gbps and category 8 to support 25 and 40 Gbps, to achieve direct connections from access switches to servers in cabinets.

Cable Management Solutions – These solutions are critical to protecting, routing, and managing cables inside any kind of edge data center. Cable management options might include cable tray systems for overhead and underfloor cable pathways, as well as horizontal and vertical cable management to protect critical copper and fiber cords.

Automated Infrastructure Management (AIM) Tools – Many edge data centers will be unmanned or limited access sites. Service providers will require infrastructure management tools that allow them to manage edge data centers in remote locations. This might include remote management and monitoring of copper and fiber connections and security locks, with real-time alerts of network and security events (i.e. cable connects/disconnects or opening of a cabinet door) to help prevent downtime or unauthorized access. It may also include monitoring of power usage at the outlet level, as well as cabinet-level environmental monitoring via data center infrastructure management (DCIM).

Intelligent PDUs – All edge data centers require power distribution units (PDUs) to distribute power to active equipment. Intelligent PDUs offer outlet-level power usage monitoring and switching control, sensors for cabinet-level environmental monitoring (i.e. temperature, humidity), and intuitive web interfaces that enable remote management and monitoring of PDUs in edge facilities from a centralized cloud location.

Racks and Cabinets – Depending on their size, edge data centers may require up to 50 racks and/or cabinets with aisle containment. The best strategy is to utilize pre-configured cabinets that can be easily rolled into place inside an edge data center and installed with active IT equipment. Additionally, edge data centers will require thermal management solutions such as blanking panels and brush guards, to maintain separation of hot and cold air.

Summary

While expanding rapidly, edge computing is still in its infancy. The infrastructure elements (i.e. power, cooling, connectivity) and form factors (i.e. rugged containers, micro data centers) of edge data centers are fairly well-developed, but we are just now starting to see the actual widespread deployment of distributed data center architectures to provide localized cloud resources to end users and devices. As new technologies like 5G networks, smart cities, and autonomous vehicles are further developed, they will integrate with, operate on, and be more reliant on edge computing resources.

What is clear at this point is that enterprises and organizations that invest in edge data centers today will achieve a competitive edge for the future. Their specific competitive edge will depend on the company, their target customer, and their technology, but here are some examples:

- Consumers will continue to demand better online experiences, and will flock to providers who offer faster, smoother delivery of Wi-Fi services and applications (i.e. two-way streaming content, VR-enhanced online games) to their smartphones, tablets, and other devices. Internet and Wi-Fi network providers and online application companies that utilize edge computing will be in a better position to provide consumers with high-speed, low-latency services, applications, and content.

- Enterprises that utilize edge data centers in an industrial setting will be able to provide local computing resources for IIoT devices, which will make their facilities faster and more productive in manufacturing and/or delivering goods to customers.

- Automotive companies will need to address safety issues regarding autonomous vehicles if the public is ever going to accept them, much less buy them. Establishing a network of edge data centers with cloud resources to direct and control autonomous vehicles will help automotive companies to guarantee the safety of their vehicles to customers, and also help them to keep in compliance with government safety regulations.

- Financial institutions that utilize edge data centers will be able to offer local computing resources to enhance services such as high-frequency trading. Additionally, edge data centers will support the use of AI for data and market analysis to help traders get better results in the financial sector.

- City and state governments that utilize edge data centers will be able to ensure the reliability of IoT devices like smart traffic lights, and thus guarantee the safety of citizens in regard to those devices. Also, establishing a network of IoT infrastructure devices (i.e. power, water, and environmental sensors) will aid local governments in managing resources, city planning, budget analysis, etc. A network of edge data centers will help local governments to manage, monitor, and collect data from IoT devices in multiple locations.

It's time to start thinking about investing in edge computing, and Siemon offers a wide range of essential infrastructure solutions that emerging edge data centers will rely on. Call your Siemon representative today to find out more about our data center infrastructure solutions and how we can help ensure you achieve a competitive edge for the future.

WP_EDGE_RevA 2/19