Ready for 5G?

Considerations for Building Unified Networks for Edge and Data Centers

To gain value from edge computing, organizations need to consider how and from where they're harvesting data.

By 2023 more than half of new IT infrastructure will be deployed at the edge, according to technology research firm IDC. Similarly, the majority of enterprise data will be generated and processed outside of centralized data centers and the cloud by 2025, according to Gartner.

Fifth-generation (5G) wireless networks can play a crucial role in these edge deployments, but that will require major changes at the device and data center levels. Edge applications will, in most cases, rely on substantial volumes of data being sent to and from centralized data hubs. This will require businesses to weave together edge and data center capabilities so they can manage data to be processed locally while efficiently transmitting the subset that must be centrally processed.

Edge is the new frontier

Locating processing resources closer to sources of data is crucial to harvest the potential value of Internet of Things (IoT) and edge devices and to deploy and leverage artificial intelligence (AI) and machine learning (ML) technologies. IDC predicts that by 2025 there will be 55.7 billion connected devices worldwide, with IoT generating 73.1 ZB of data.

Transmitting such huge volumes of data to and from central hubs is costly and imposes latency issues that are unacceptable for near-real-time decision-making and execution. In that context, according to the IEEE: "Traditional cloud computing has serious disadvantages, including data security threats, performance issues, and growing operational costs. Because most data saved in the cloud has little significance and is rarely used, it becomes a waste of resources and storage space."

By 2025 there will be 55.7 billion connected devices worldwide.

Ultimately, businesses will require localized, low-latency computing resources for a large portion of edge data processing and storage to meet their goals. They need to limit the data that must be transmitted to the cloud or on-premises data centers for intensive compute tasks.

"As operation teams began experimenting with machine learning and AI in the cloud, they realized that latency or lag in response times was prohibitive for real-time applications," writes IDC analyst Dave McCarthy. "Edge computing solves this problem by training models in the cloud and running them on local infrastructure."

In 2018 McKinsey analysts documented more than 100 use cases for edge computing, spanning virtually every industry. Some examples: condition-based maintenance in aircraft, satellite telecommunications and weather analysis, home energy conservation, use of drones, optimizing operations at mines and oil rigs, improving fitness and wellness, inventory optimization in retail, and autonomous vehicles.

Yet latency and communication speed issues are major hurdles to overcome. Many edge decisions, such as preventing expensive industrial equipment from overheating, must be made in real time. In time, AI technologies will make it possible for devices to automatically make crucial decisions such as ensuring that an automobile stops for a pedestrian crossing the road. Until then some tasks and decisions will still rely on communications with data centers. That's where 5G comes in.

"5G and edge computing are two inextricably linked technologies," according to research and consulting firm STL Partners. "They are both poised to significantly improve the performance of applications and enable huge amounts of data to be processed in real-time. 5G increases speeds by up to ten times that of 4G, whereas mobile edge computing reduces latency by bringing compute capabilities into the network, closer to the end user."

Retooling data centers for edge

Edge computing can move data processing closer to the source, but many applications still will need to stream to data centers for critical operations—such as oversight by remote human operators, more-resource-intensive analysis, and development and training of ML models.

The data center will require major changes to accommodate and enable 5G devices. "Moving more compute and storage resources to the edge has implications for current data center infrastructures and requires special planning," says Matt Ritter, director of engineering at Silicon Mechanics, a leading provider of servers, storage, high-performance computing (HPC), and AI technologies.

"It's important that the data center has a high-speed network backbone in place to facilitate node-to-node communication between compute and storage as well as communication out to users and sensors producing and consuming the data," says Ritter. "The increased power and cooling load of these additional resources also needs to be taken into account, especially for deployments in more remote locations."

"It's important that the data center has a high-speed network backbone in place to facilitate node-to-node communication between compute and storage as well as communication out to users and sensors producing and consuming the data."

— Matt Ritter, director of engineering at Silicon Mechanics

Heavy compute workloads require high-speed, high-capacity networks, with an advanced switch layer feeding multiple servers working on the same data and a high-speed connection combining the central processing unit (CPU), graphics processing unit (GPU), and storage. Some applications may require new technologies such as composable infrastructure, which abstracts hardware resources from their physical location and can be managed by software over the network fabric to apply those assets where needed at any given time.

These new edge infrastructures will also need multilevel security solutions to eliminate redundant copies of data or resource silos, facilitate secure data sharing and efficient collaboration between the edge and the data center, and consolidate security auditing for greater accountability.

Privacy and other regulatory concerns will require new protocols to protect data. A group of researchers warns that "edge nodes are close to the users, and as a result can potentially receive large amounts of privacy-sensitive data. If the data from an edge node is leaked, the consequences can be serious."

The medical industry is expected to be prominent in developing edge use cases, such as patient monitoring, so compliance with the Health Insurance Portability and Accountability Act (HIPAA) is a major issue. Another example: Military edge applications will require encryption while data is in flight or at rest.

5G and the extended edge

Some edge devices already process and provide limited storage of data on the edge, even in harsh environmental conditions that require ruggedization. But others will need to act as complex gateways to the data center, so devices equipped with 5G or other advanced telecommunications will be essential.

"5G has massive systems capacity to connect many sensors and other equipment, but these devices have to meet not only the requirements for remote and often harsh environmental conditions but also safety and regulatory requirements," says Parag Shah, chief technology officer of Comark, which has been designing and manufacturing customized rugged edge computing solutions for 40 years.

The deployment of 5G will require new generations of industrial mobile computing devices. This will include new thought given to:

- **RF antennas**
- **Power requirements**
- **New sets of hardware and firmware**
- **Testing and certification**
- **Cybersecurity defenses**

Much will depend on carrier requirements that edge device providers may have to meet when designing and manufacturing client solutions. "In order to incorporate 5G into edge devices, you have to strive for low power usage, integrity of the radio frequency signal, and safety," Shah says. "It is still unclear whether carriers such as AT&T, Verizon, and T-Mobile will have specific certifications for edge devices. Large carriers generally don't focus on mobile industrial applications."

Some industrial companies may bypass carriers—or utilize them as managed services providers—for their own private wide- and local-area networks for more efficient communications in locales such as factory floors, mines, and shipyards. That will require development of provisioning and activation processes.

Realizing the promise of the edge

The 5G-enabled future will enable the creation of billions of IoT and other edge devices. These will come in all shapes and sizes, configured to enable a vast range of new applications, including a slew of new industrial-grade (or rugged) devices, meaning that new servers, storage, and high-performance computing technologies will be required.

New generations of 5G edge devices hold great potential for improving processes and creating new revenue opportunities.

However, businesses need assurance that both their device manufacturers and their data center partners have the expertise and technology to help safeguard this transition.

It is critical that organizations looking at 5G devices work with technology partners that understand the practical aspects. Some firms are well positioned to do so, because they have already successfully navigated the migration to 4G—although that had less impact on systems than 5G will—and know what to be aware of. Others have organically gained experience in the relevant areas through related client work.

This partnership factor means that some organizations will reap the benefits of 5G while others, mired in complexity and revision, will waste time and resources while competitors leap ahead. A considered approach to 5G deployment, however, will change the way businesses, governments, and individuals interact with data in ways that previous generations of wireless networks could not offer.

Learn more about Comark here.
Learn more about Silicon Mechanics here.

Artificial Intelligence at the Edge

Autonomous vehicles, just-in-time maintenance, and real-time image processing. These applications and many more are the goal of deploying AI technologies such as machine learning and deep learning at the edge.

Machine learning and deep learning rely on huge volumes of data that must be stored and processed. Using these technologies at the edge generally requires a tiered processing system in which data is uploaded to cloud infrastructure or a data center for processing. The improved bandwidth and associated low latency of 5G will make this easier and lead to the proliferation of AI on the edge.

In the cloud or a data center, algorithms can be further refined and downloaded to the edge device, which, in turn, becomes better able to act or output in the desired fashion without having to wait for the data center to execute a decision, such as stopping an automobile when a ball rolls into its path.

Once the AI training process is done, edge devices can potentially also process data for inference more effectively with 5G, thanks again to the large bandwidth (ideal for images) and low latency. However, this requires technology configuration considerations. Options include GPU-accelerated and field-programmable gate array (FPGA)–accelerated devices.

As more-powerful and less-power-intensive devices are deployed at the edge, those devices become less dependent on real-time interaction with the cloud or a data center.