Rearchitecting the Internet

By: Phillip Marangella

It's an invigorating time to be in the data center industry, particularly the edge data center industry, as global, regional and local businesses enter a new phase of the digital economy characterized by a host of next-generation technologies. According to a forecast from Global Market Insights, the edge data center market, valued at roughly $4.5 billion in 2018, is projected to quadruple at a compound annual growth rate (CAGR) of 20.5 percent to reach more than $16 billion by 2025. The market research firm points to the rise in data traffic and the increasing adoption of IoT technology as key drivers for the demand of edge infrastructure.

As predicted by Cisco, global IP traffic will reach 4.8 zettabytes per annum by 2022. That same year, the number of connected IoT sensors and devices is expected to exceed 50 billion, up from an estimated 21 billion in 2018, according to Juniper Research. Let's put it in perspective: to accommodate this surge in data and "things," the infrastructure we are building today, including edge infrastructure, is as essential to the new age of our digital economy as the development of the interstate highway system was in the middle of the last century.

Now, more than ever, it's clear that the volume, velocity and variety of data traffic flows are mandating a dramatic rearchitecting of the Internet to make it more customer-centric than ever before. The Edge is the lowest latency point of demarcation between service delivery and consumption. By processing data away from the cloud and closer to the source, the Edge increases both the scalability and security of IoT-based applications.

The first and second waves of demand for edge infrastructure were driven by the exponential growth in Over-the-Top (OTT) content and the mainstreaming of cloud services, respectively. Along with the growth of consumer and industrial IoT applications as our homes, buildings, streets, factories and entire communities become "smarter" and ever more connected, the next wave will see a range of Edge form factors proliferate globally in support of such advanced technologies as 5G networks, autonomous vehicles, Artificial Intelligence (AI), and virtual and augmented reality (AV/VR) applications.

So, let's take a look at each of these next-gen technology trends, and how the edge and a rearchitecting of the Internet will solve many of the key challenges ahead.

5G Wireless

The initial standards for 5G, which is shorthand for fifth-generation cellular wireless, were established at the end of 2017. 5G is truly an investment, however, for the next decade and beyond. As disruptive as previous generations of cellular wireless technology were in comparison to earlier iterations, 5G brings three significant advances to the world of always-on, mobile connectivity. 5G provides greater speed to move more data, lower latency to be more responsive, and the ability to connect many more devices simultaneously for IoT-enabled sensors and smart devices.

The coming data tsunami will see a range of edge solutions in service of 5G networks, from local reach edge data centers of 2MW to 20MW, to hyperlocal MicroEdge data centers of 100kW to 1MW. As a recent report from Market Watch observed, the deployment and commercialization of 5G technology is having a significant impact on

edge data center market growth, since edge facilities enable the processing of large volumes of data in proximity to base stations in 5G infrastructure.

Autonomous Vehicles

One of the sexiest and most-often cited use cases for 5G is autonomous vehicles. While the first generation of driverless cars will be self-contained, future autonomous vehicles will be able to interact with other intelligent vehicles, smart roads and buildings embedded with IoT-enabled sensors that will improve safety and manage traffic.

For the autonomous vehicle, 5G's sub-one-millisecond latency is essential. Even so, processing the high volumes of data these IoT-enabled sensors will generate in real-time will drive the need for more and more edge data centers, which facilitate data processing at or near the source of data generation, and overcome cloud overhead in latency and bandwidth. By 2030, Forbes estimates that one in four cars will be self-driven. So, it comes as no surprise that technologists consider the autonomous vehicle to be both the ultimate IoT device and an emissary for the edge.

While estimates vary, some experts believe that a self-driving test vehicle can produce as much as 30 terabytes of data in a single day of driving, and that data will need to run through powerful analytics programs to produce actionable information. In this instance, edge data centers will prioritize what data needs to remain on the edge to be processed by the vehicle’s onboard computing power, and what data should be relayed back to centralized data centers or the cloud for analysis.

Artificial Intelligence

According to a survey by McKinsey & Company, one out of five of C-level executives report that they are using AI as a core part of their business. Indeed, AI and an associated technology, Deep Learning, are already at work helping financial institutions to detect fraudulent transactions, assisting retailers in creating personalized marketing, guiding streaming video content providers to make suggestions for our next Netflix binge-watch, and even assisting healthcare providers in making medical diagnoses.

AI and Deep Learning technologies promise to save organizations billions of dollars over the next few decades. However, as companies turn to new AI applications driven by high-performance computing (HPC), they are finding that traditional computing platforms and legacy data centers are not equipped to handle these new demands.

Traditional data centers were built to deliver 5kW to 10kW of power per rack, which was sufficient for servers running standard CPUs that run hot and fail to accommodate large numbers of processors or servers in a single rack or cabinet. Conversely, edge data centers that can offer 30kW to 35kW or more per rack allow more computing power to occupy a smaller physical footprint. This power density also allows customers to take advantage of new, more energy-efficient processors and servers that require less power to perform even more processing-intensive work, such as AI and Deep Learning applications.

Virtual and Augmented Reality

Tractica forecasts that worldwide enterprise VR hardware and software revenue will increase from $1 billion in 2018 to $12.6 billion annually by 2025. VR/AR applications are presently at work in such industries as automotive, energy, utilities, manufacturing, travel and transportation, and healthcare.

Creating entirely virtual worlds—or superimposing digital images and graphics on top of the real world in an immersive way—requires a lot of processing power. Both VR and AR also require high-bandwidth and ultra-low latency. Here again, as the lowest latency point of demarcation between service delivery and consumption, and by processing data away from the cloud and closer to the source, the Edge delivers the ultra-low latency data transport required by VR/AR applications.

Demystifying Data Traffic Flows

The Internet just wasn’t constructed to handle the traffic flows of today; it was built for download-centric traffic. So much content is now being created at the edge and it’s creating huge bottlenecks.

Hence, infrastructure providers must work with cloud partners and service providers to alleviate bottlenecks at the edge and ultimately help enterprises demystify their business-critical data traffic flows. Does traffic need to go back to the core? Does data need to remain at the edge?

As the IoT, autonomous vehicles, AI, VR/AR applications and other latency-sensitive and network-critical trends become more mainstream, computing and networking at the edge will become essential to alleviate global network congestion issues. Enterprise organizations and service providers will have to collaborate to solve for these challenges and create multi-directional traffic flows, and that means they need edge data centers in all their various form factors, and sited wherever customers demand.