EDGE COMPUTING: AN INTRODUCTION WITH EXAMPLES

Edge computing is a relatively new paradigm that aims to bring computational power in close proximity of IoT sensors, smartphones, and connected technologies.

Edge computing has emerged as a promising technology thanks to the proliferation of smart IoT applications in autonomous vehicles and other computation-sensitive industrial use cases that require low-latency data processing. Gartner predicts that 50% of enterprise-generated data will be created and processed beyond centralized cloud data centers via edge computing by the year 2022. Other research finds that, by 2025, the global IoT installed base will reach over 75.4 billion devices.

So, let's explore edge computing through some real world use cases.

What is edge computing?

Edge computing is the computational processing of sensor data away from the centralized nodes and close to the logical edge of the network, toward individual sources of data. It may be referred to as a distributed IT network architecture that enables mobile computing for data produced locally. Instead of sending the data to cloud data centers, edge computing decentralizes processing power to ensure real-time processing without latency while reducing bandwidth and storage requirements on the network.

The concept dates back to the 1990s, when Akamai solved the challenge of Web traffic congestion by introducing Content Delivery Network (CDN) solutions. The technology involved network nodes storing static cached media information at locations closer to end-users.

Today, edge computing takes this concept further, introducing computational capabilities into nodes at the network edge to process information and deliver services.

Examples of edge computing

Edge computing offers a range of value propositions for smart IoT applications and use cases across a variety of industries. Some of the most popular use cases that will depend on edge computing to deliver improved performance, security and productivity for enterprises include:

Autonomous vehicles

For autonomous driving technologies to replace human drivers, cars must be capable of reacting to road incidents in real-time. On average, it may take 100 milliseconds for data transmission between vehicle sensors and backend cloud datacenters. In terms of driving decisions, this delay can have significant impact on the reaction of self-driving vehicles.

Toyota predicts that the amount of data transmitted between vehicles and the cloud could reach 10 exabytes per month by the year 2025. If network capacity fails to accommodate the necessary network traffic, vendors of autonomous vehicle technologies may be forced to limit self-driving capabilities of the cars.

In addition to the data growth and existing network limitations, technologies such as 5G connectivity and Artificial Intelligence are paving the way for edge computing.

- 5G will help deploy computing capabilities closer to the logical edge of the network in the form of distributed cellular towers. The technology will be capable of greater data aggregation and processing while maintaining high speed data transmission between vehicles and communication towers.
- AI will further facilitate intelligent decision-making capabilities in real-time, allowing cars to react faster than humans in response to abrupt changes in traffic flows.

Fleet management

Logistics service providers leverage IoT telematics data to realize effective fleet management operations. Drivers rely on vehicle-to-vehicle communication as well as information from backend control towers to make better decisions. Locations of low connectivity and signal strength are limited in terms of the speed and volume of data that can be transmitted between vehicles and backend cloud networks.

With the advent of autonomous vehicle technologies that rely on real-time computation and data analysis capabilities, fleet vendors will seek efficient means of network transmission to maximize the value potential of fleet telematics data for vehicles travelling to distant locations.

By drawing computation capabilities in close proximity of fleet vehicles, vendors can reduce the impact of communication dead zones as the data will not be required to send all the way back to centralized cloud data centers. Effective vehicle-to-vehicle communication will enable coordinated traffic flows between fleet platoons, as AI-enabled sensor systems deployed at the network edges will communicate insightful analytics information instead of raw data as needed.

Predictive maintenance

The manufacturing industry heavily relies on the performance and uptime of automated machines. In 2006, the cost of manufacturing downtime in the automotive industry was estimated at $1.3 million per hour. A decade later, the rising financial investment toward vehicle technologies and the growing profitability in the market make unexpected service interruptions more expensive in multiple orders of magnitude.

With edge computing, IoT sensors can monitor machine health and identify signs of time-sensitive maintenance issues in real-time. The data is analyzed on the manufacturing premises and analytics results are uploaded to centralized cloud data centers for reporting or further analysis.

- Analyzing anomalies can allow the workforce to perform corrective measures or predictive maintenance earlier, before the issue escalates and impacts the production line.
- Analyzing the most impactful machine health metrics can allow organizations to prolong the useful life of manufacturing machines.

As a result, manufacturing organizations can lower the cost of maintenance, improve operational effectiveness of the machines, and realize higher return on assets.

Voice assistance

Voice assistance technologies such as Amazon Echo, Google Home, and Apple Siri, among others, are pushing the boundaries of AI. An estimated 56.3 million smart voice assistant devices will be shipped globally in 2018. Gartner predicts that 30 percent of consumer interactions with the technology will take place via voice by the year 2020. The fast-growing consumer technology segment requires advanced AI processing and low-latency response time to deliver effective interactions with end-users.

Particularly for use cases that involve AI voice assistance capabilities, the technology needs go beyond computational power and data transmission speed. The long-term success of voice assistance depends on consumer privacy and data security capabilities of the technology. Sensitive personal information is a treasure trove for underground cybercrime rings and potential network vulnerabilities in voice assistance systems could pose unprecedented security and privacy risks to end-users.

To address this challenge, vendors such as Amazon are enhancing their AI capabilities and deploying the technology closer to the edge, so that voice data doesn't need to move across the network. Amazon is reportedly working to develop its own AI chip for the Amazon Echo devices.

Prevalence of edge computing in the voice assistance segment will hold equal importance for enterprise users as employees working in the field or on the manufacturing line will be able to access and analyze useful information without interrupting manual work operations.

The future of edge computing

According to the Gartner Hype Cycle 2017, edge computing is drawing closer to the Peak of Inflated Expectations and will likely reach the Plateau of Productivity in 2-5 years. Considering the ongoing research and developments in AI and 5G connectivity technologies, and the rising demands of smart industrial IoT applications, Edge Computing may reach maturity faster than expected.

Additional resources

For more on this topic and other emerging technologies, explore the BMC Business of IT Blog or read these articles:

- What is the "Intelligent Edge"?
- The Empowered Edge: An Introduction
- AWS Snowball Edge: An Introduction
- An Introduction to Distributed Clouds