A seminar Report

on

EDGE COMPUTING

Abstract

With Digital Transformation and emerging technologies that will enable “smart” everything – cities, agriculture, cars, health, etc – in the future require the massive deployment of Internet of Things (IoT) sensors while edge computing will drive the implementations. The proliferation of Internet of Things (IoT) and the success of rich cloud services have pushed the horizon of a new computing paradigm, edge computing, which calls for processing the data at the edge of the network. Edge computing has the potential to address the concerns of response time requirement, battery life constraint, bandwidth cost saving, as well as data safety and privacy. In this paper, we introduce the deﬁnition of edge computing, followed by several case studies, ranging from cloud ofﬂoading to smart home and city, as well as collaborative edge to materialize the concept of edge computing. Finally, we present several challenges and opportunities in the ﬁeld of edge computing, and hope this paper will gain attention from the community and inspire more research in this direction.

**Introduction**

CLOUD computing has tremendously changed the way we live, work, and study since its inception around 2005. For example, software as a service (SaaS) instances, such as Google Apps, Twitter, Facebook, and Flickr, have been widely used in our daily life. Moreover, scalable infrastructures as well as processing engines developed to support cloud service are also signiﬁcantly inﬂuencing the way of running business, for instance, Google File System, MapReduce, Apache Hadoop, Apache Spark, and so on.

Internet of Things (IoT) was ﬁrst introduced to the community in 1999 for supply chain management, and then the concept of “making a computer sense information without the aid of human intervention” was widely adapted to other ﬁelds such as healthcare, home, environment, and transports. Now with IoT, we will arrive in the post-cloud era, where there will be a large quality of data generated by things that are immersed in our daily life, and a lot of applications will also be deployed at the edge to consume these data. By 2019, data produced by people, machines, and things will reach 500 zettabytes, as estimated by Cisco Global Cloud Index, however, the global data center IP trafﬁc will only reach 10.4 zettabytes by that time. By 2019, 45% of IoT-created data will be stored, processed, analyzed, and acted upon close to, or at the edge of, the network. There will be 50 billion things connected to the Internet by 2020, as predicted by Cisco Internet Business Solutions Group. Some IoT applications might require very short response time, some might involve private data, and some might produce a large quantity of data which could be a heavy load for networks. Cloud computing is not efﬁcient enough to support these applications. With the push from cloud services and pull from IoT, we envision that the edge of the network is changing from data consumer to data producer as well as data consumer. In this paper, we attempt to contribute the concept of edge computing. We start from the analysis of why we need edge computing, then we give our deﬁnition and vision of edge computing.

**What is Edge Computing?**

Edge computing is a “mesh network of micro data centers that process or store critical data locally and push all received data to a central data center or cloud storage repository, in a footprint of less than 100 square feet,” according to research firm IDC.

Edge computing is a method of optimizing cloud computing systems by performing data processing at the edge of the network, near the source of the data. This reduces the communications bandwidth needed between sensors and the central datacenter by performing analytics and knowledge generation at or near the source of the data. This approach requires leveraging resources that may not be continuously connected to a network such as laptops, smartphones, tablets and sensors.

Edge computing covers a wide range of technologies including wireless sensor networks, mobile data acquisition, mobile signature analysis, cooperative distributed peer-to-peer ad hoc networking and processing also classifiable as local cloud/fog computing and grid/mesh computing, dew computing, mobile edge computing, cloudlet, distributed data storage and retrieval, autonomic self-healing networks, remote cloud services, augmented reality, and more.

It is typically referred to in IoT use cases, where edge devices would collect data – sometimes massive amounts of it – and send it all to a data center or cloud for processing. Edge computing triages the data locally so some of it is processed locally, reducing the backhaul traffic to the central repository.

Typically, this is done by the IoT devices transferring the data to a local device that includes compute, storage and network connectivity in a small form factor. Data is processed at the edge, and all or a portion of it is sent to the central processing or storage repository in a corporate data center, co-location facility or IaaS cloud.

In Industrial Internet of Things (IIoT), applications such as power production, smart traffic lights, or manufacturing, the edge devices capture streaming data that can be used to prevent a part from failing, reroute traffic, optimize production, and prevent product defects.

In the context of IIoT, 'edge' refers to the computing infrastructure that exists close to the sources of data, for example, industrial machines (e.g. wind turbine, magnetic resonance (MR) scanner, undersea blowout preventers), industrial controllers such as SCADA systems, and time series databases aggregating data from a variety of equipment and sensors. These devices typically reside away from the centralize computing available in the cloud.

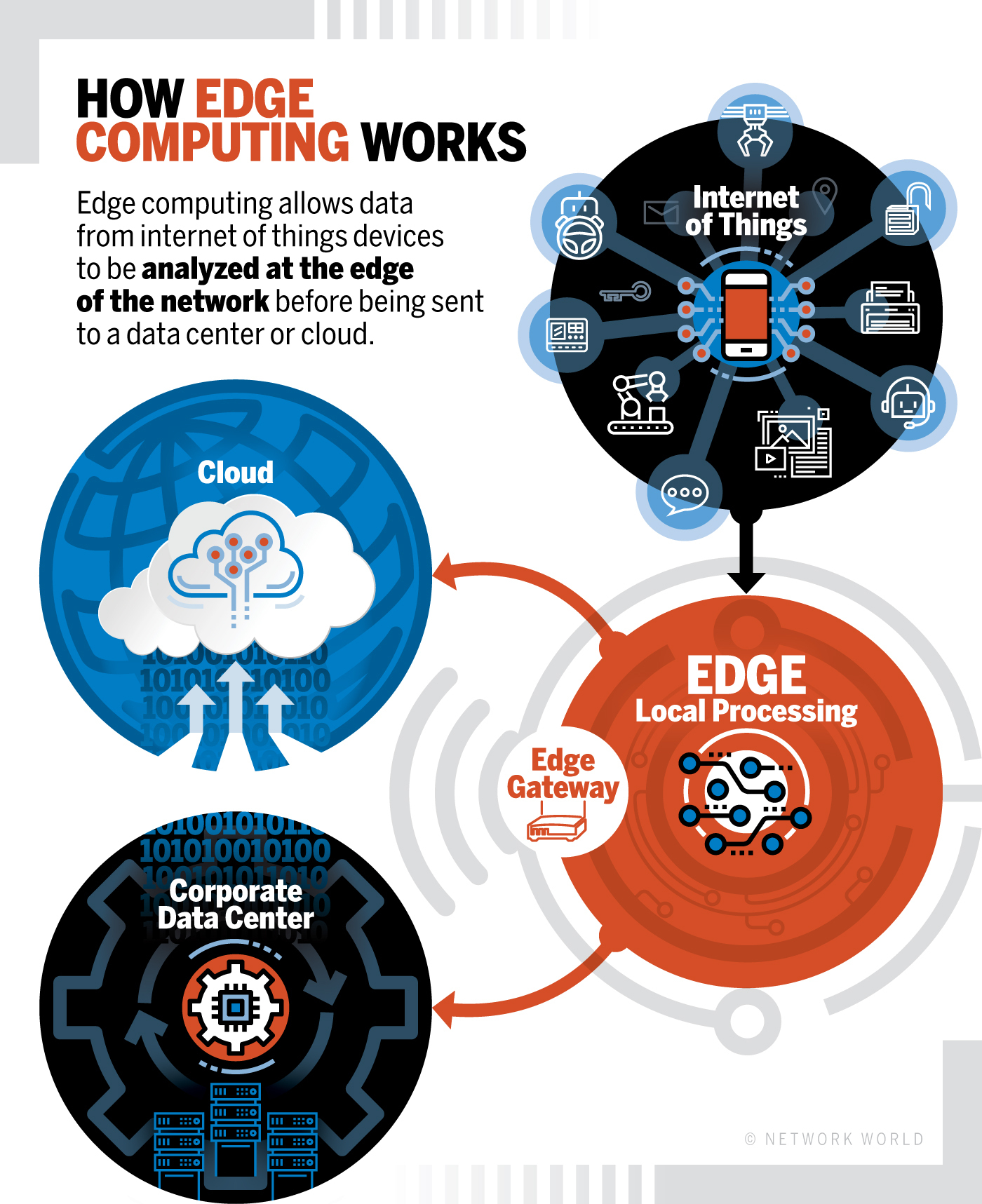
The role of edge computing to date has mostly been used to ingest, store, filter, and send data to cloud systems. We are at a point in time, however, where these computing systems are packing more compute, storage, and analytic power to consume and act on the data at the machine location. This capability will be more than valuable to industrial organizations—it will be indispensable.

**How Edge Computing works?**

Edge computing pushes applications, data and computing power (services) away from centralized points to the logical extremes of a network. Edge computing replicates fragments of information across distributed networks of web servers, which may spread over a vast area. As a technological paradigm, edge computing is also referred to as mesh computing, peer-to-peer computing, autonomic (self-healing) computing, grid computing, and by other names implying non-centralized, nodeless availability.

To ensure acceptable performance of widely dispersed distributed services, large organizations typically implement edge computing by deploying Web server farms with clustering. Previously available only to very large corporate and government organizations, edge computing has utilized technology advances and cost reductions for large-scale implementations have made the technology available to small and medium-sized businesses.

The target end-user is any Internet client making use of commercial Internet application services. Edge computing imposes certain limitations on the choices of technology platforms, applications or services, all of which need to be specifically developed or configured for edge computing.



**Why Do We Need Edge Computing?**

1. **Push From Cloud Services**: Putting all the computing tasks on the cloud has been proved to be an efﬁcient way for data processing since the computing power on the cloud outclasses the capability of the things at the edge. However, compared to the fast developing data processing speed, the bandwidth of the network has come to a standstill. With the growing quantity of data generated at the edge, speed of data, transportation is becoming the bottleneck for the cloud-based computing paradigm. For example, about 5 Gigabyte data will be generated by a Boeing 787 every second, but the bandwidth between the airplane and either satellite or base station on the ground is not large enough for data transmission. Consider an autonomous vehicle as another example. One Gigabyte data will be generated by the car every second and it requires real-time processing for the vehicle to make correct decisions. If all the data needs to be sent to the cloud for processing, the response time would be too long. Not to mention that current network bandwidth and reliability would be challenged for its capability of supporting a large number of vehicles in one area. In this case, the data needs to be processed at the edge for shorter response time, more efﬁcient processing and smaller network pressure.
2. **Pull from IoT**: Almost all kinds of electrical devices will become part of IoT, and they will play the role of data producers as well as consumers, such as air quality sensors, LED bars, streetlights and even an Internet-connected microwave oven. It is safe to infer that the number of things at the edge of the network will develop to more than billions in a few years. Thus, raw data produced by them will be enormous, making conventional cloud computing not efﬁcient enough to handle all these data. This means most of the data produced by IoT will never be transmitted to the cloud, instead it will be consumed at the edge of the network. Fig. 1 shows the conventional cloud computing structure. Data producers generate raw data and transfer it to cloud, and data consumers send request for consuming data to cloud, as noted by the blue solid line. The red dotted line indicates the request for consuming data being sent from data consumers to cloud, and the result from cloud is represented by the green dotted line. However, this structure is not sufﬁcient for IoT. First, data quantity at the edge is too large, which will lead to huge unnecessary bandwidth and computing resource usage. Second, the privacy protection requirement will pose an obstacle for cloud computing in IoT. Lastly, most of the end nodes in IoT are energy constrained things, and the wireless communication module is usually very energy hungry, so ofﬂoading some computing tasks to the edge could be more energy efﬁcient.

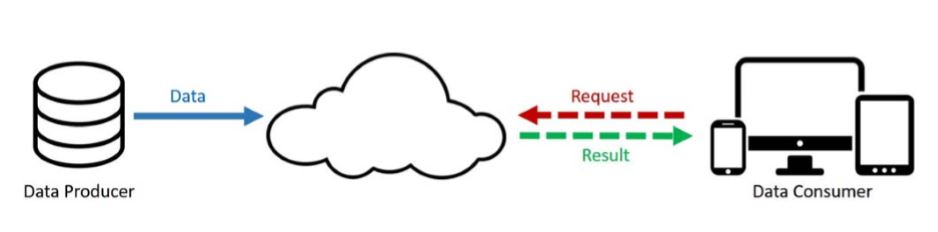


Fig 1: Cloud computing paradigm

1. **Change from Data Consumer to Producer**: In the cloud computing paradigm, the end devices at the edge usually play as data consumer, for example, watching a YouTube video on your smart phone. However, people are also producing data nowadays from their mobile devices. The change from data consumer to data producer/consumer requires more function placement at the edge. For example, it is very normal that people today take photos or do video recording then share the data through a cloud service such as YouTube, Facebook, Twitter, or Instagram. Moreover, every single minute, YouTube users upload 72 h of new video content; Facebook users share nearly 2.5 million pieces of content; Twitter users tweet nearly 300000 times; Instagram users post nearly 220000 new photos. However, the image or video clip could be fairly large and it would occupy a lot of bandwidth for uploading. In this case, the video clip should be demised and adjusted to suitable resolution at the edge before uploading to cloud. Another example would be wearable health devices. Since the physical data collected by the things at the edge of the network is usually private, processing the data at the edge could protect user privacy better than uploading raw data to cloud.

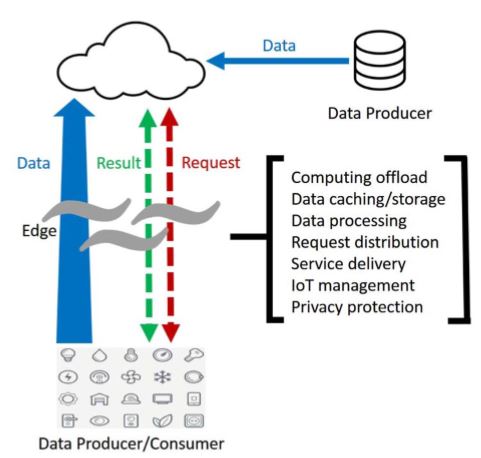


Fig: Edge computing paradigm

**Edge computing vs. cloud computing**

First of all, we need to say that **edge computing** is a special computing infrastructure existing at the edges of data sources, e.g. devices (industrial machines like turbines, magnetic resonance systems, self-driving cars, smart homes, and other smart devices envisaging incorporating many sensors and operating with their data). In other words, it’s pushing the computing applications frontier away from centralized nodes to the network extremes. That means edge computing requires leveraging device resources so that they don’t need to be connected to the network (or data center) continuously.

The opposing method, **cloud computing**, requires that all things be connected to the central data storage, where huge volumes of information are processed to find optimization solutions or make business decisions. As a rule, cloud computing is associated with complex data processing operations requiring significant computational power. At the same time, data accumulation and processing are not quick enough to be applied in some special spheres where the computational results need to be applied instantly.

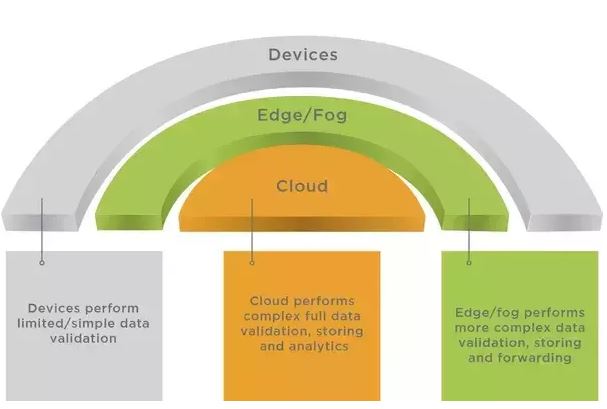
The problem with cloud computing is widely known. It’s resulted in the appearance of a middle tier in the data circulation model – fog computing. Fog computing is an attempt to push computing powers closer to the data sources, eliminating response times without affecting efficiency. In fog computing, the computing is distributed in the most logical, efficient place between the data source and the cloud – in a “fog”. Fog is territorially closer to the devices as compared to the cloud, however, it’s still just a middle chain pushing the information further even if it’s able to make some decisions on the fly.

For industrial companies to fully realize the value of the massive amounts of data being generated by machines, edge computing and cloud computing must work together.

**In IIoT context**

When you consider these two technologies, think about the way you use your two hands. You will use one or both depending on action required. Apply that to an IIoT example, where one hand is edge and the other hand is cloud, and you can quickly see how in certain workloads your “edge hand” will play a more prominent role while in other situations your “cloud hand” will take a lead position. And there will be times when both hands are needed in equal measure.

Scenarios in which edge will dominate include a need for low latency (speed is of the essence) or where there are bandwidth constraints (locations such as a mine or an offshore oil platform that make it neither practical nor affordable, and in some cases impossible, to send all data from machines to the cloud). It will also be important when Internet or cellular connections are spotty. Cloud computing will take a more dominant position when actions require significant computing power, managing data volumes from across plants, asset health monitoring and machine learning, and so on.



The bottom line is this: cloud and edge are both necessary to industrial operations to gain the most value from today’s sophisticated, varied, and volume of data applied across cloud and edge, wherever it makes the most sense to achieve the desired outcomes.

**Edge Computing vs. Fog computing**

As the edge computing market takes shape, there’s an important term related to edge that is catching on: fog computing. Both edge computing and fog computing are strongly on the rise for the same exact reasons: an IoT data deluge.

This IoT data deluge, among others, takes place in the converging worlds of IT and OT (again predominantly Industrial IoT) and occurs in general as we keep adding more IoT devices in the scope of mainly large-scale IoT projects, the industrial markets of fourth industrial revolution and IoT use cases and applications where a lot of data needs to be analyzed and leveraged, often also in an IT and OT environment as we, for instance, find them in IoT in manufacturing.

Fog refers to the network connections between edge devices and the cloud. Edge, on the other hand, refers more specifically to the computational processes being done close to the edge devices. So, fog includes edge computing, but fog would also incorporate the network needed to get processed data to its final destination.

Some have predicted that edge computing could displace the cloud. But Mung Chaing, dean of Purdue University’s School of Engineering and co-chair of the OpenFog Consortium, believes that no single computing domain will dominate; rather there will be a continuum. Edge and fog computing are useful when real-time analysis of field data is required.

**Edge computing examples**

Here’s a couple of examples to help bring this computing concept to life.

**Autonomous vehicles**

With autonomous automobiles—essentially a datacenter on wheels— this type of computing plays a dominant role. Intel, estimates that autonomous cars, with hundreds of on-vehicle sensors, will generate 40TB of data for every eight hours of driving. That’s a lot of data. It is unsafe, unnecessary, and impractical to send all that data to the cloud.

It’s unsafe because the sensing, thinking, and acting attributes of edge computing in this use case must be done in real-time with ultra-low latency to ensure safe operation for passengers and the public. An autonomous car sending data to the cloud for analysis and decision-making as it traverses city streets and highways would prove catastrophic. For example, consider a child chasing a ball into the street in front of an oncoming autonomous car. In this scenario, low latency is required for decision and subsequent actuation (the car needs to brake NOW!).

It’s unnecessary to send all that data to the cloud because this particular set of data has only short-term value (a particular ball, a particular child on a collision with a particular car). Speed of actuation on that data is paramount. It’s simply impractical (not to mention cost-prohibitive) to transport vast volumes of data generated from machines to the cloud.

However, the cloud is still an important part of IIoT equation. The simple fact that the car had to respond to such an immediate and specific event might be valuable data when aggregated into a digital twin, and compared with the performance of other cars of its class.

**Fleet management**

In a scenario where a company has a fleet (think trucking company, for example), the main goal could be to ingest, aggregate, and send data from multiple operational data points (think wheels, brakes, battery, electrical) to the cloud. The cloud performs analytics to monitor the health of key operational components. A fleet manager utilizes a fleet management solution to proactively service the vehicle to maximize uptime and lower cost. The operator can track KPIs such as cost over time by part, and/or the average cost of a given truck model over time. This in turn helps maintain optimal performance at a lower cost and higher safety.

Coca Cola free style dispensers are using edge computing to quickly understand the consumer behavior and help to be more responsive to needs.

GE locomotives take advantage of edge computing by gathering and processing real-time data about railway conditions, train maintenance, and even crew morale to help railroad companies move trains through crowded railway corridors in as safe and efficient a manner as possible.

With Digital Transformation and emerging technologies that will enable “smart” everything – cities, agriculture, cars, health, etc – in the future require the massive deployment of Internet of Things (IoT) sensors while edge computing will drive the implementations.

**Edge computing terms and definitions**

**Edge:** What the edge is depends on the use case. In a telecommunications field, perhaps the edge is a cell phone or maybe it’s a cell tower. In an automotive scenario, the edge of the network could be a car. In manufacturing, it could be a machine on a shop floor; in enterprise IT, the edge could be a laptop.

**Edge server:** Edge servers can be defined as “a computer for running middleware or applications that sits close to the edge of the network, where the digital world meets the real world. Edge servers are put in warehouses, distribution centers and factories, as opposed to corporate headquarters.”

**Edge devices**: These can be any device that produces data. These could be sensors, industrial machines or other devices that produce or collect data.

**Mobile edge computing:** This refers to the buildout of edge computing systems in telecommunications systems, particularly 5G scenarios.Mobile edge computing is a network architecture concept that enables cloud computing capabilities and an IT service environment at the edge of the cellular network. The basic idea behind MEC is that by running applications and performing related processing tasks closer to the cellular customer, network congestion is reduced and applications perform better.

GE Digital, Qualcomm Technologies and Nokia announced a successful demonstration of a private LTE network for Industrial IoT. As the news describes, “Industrial companies often have local connectivity needs and operate in remote locations or temporary sites, such as mines, power plants, offshore oil platforms, factories, warehouses or ports—connectivity for these environments can be challenging. A standalone LTE network to serve devices and users within a localized area can help improve performance and reliability for these industrial settings.”

**Edge gateway**: A gateway is the buffer between where edge computing processing is done and the broader fog network. The gateway is the window into the larger environment beyond the edge of the network.

**Fat client**: Software that can do some data processing in edge devices. This is opposed to a thin client, which would merely transfer data.

**Examples of edge devices**

An edge device can be defined in several ways. You could think of it as an entry point into enterprise or service provider core networks. Examples include routers, switches, integrated access devices (IADs), multiplexers, and a variety of metropolitan area network (MAN) and wide area network (WAN) access devices. These devices also provide connections into carrier and service provider networks. Edge computing uses a range of existing and new equipment. Many devices, sensors and machines can be outfitted to work in an edge computing environment by simply making them Internet-accessible. Cisco and other hardware vendors have a line of ruggedized network equipment that has hardened exteriors meant to be used in field environments. A range of computer servers, converged systems and even storage-based hardware systems like Amazon Web Service’s Snowball can be used in edge computing deployments.

**Edge computing and IoT in 2018 and beyond**

With real-time information even being a proven competitive differentiator it is clear the in the increasing unstructured data deluge of which the IoT and sensor data deluge is part, traditional approaches don’t fit anymore as we’ll see.

There are even applications and industries where, just on the level of sending data, traditional networks don’t suffice, let alone can be used, for instance because of their remoteness and the costs it takes to send all this data through, for instance, satellite communications.

So, for a mix of reasons (bandwidth, costs, speed, automation, maintenance, predictive analytics, remoteness, you name it) we need a faster, cheaper and smarter approach than the traditional one which typically goes like: gather the data, send them through networks to the cloud or other environments where they can get processed and leveraged and so forth.

That’s where both edge computing and fog computing really come in. If your data is generated at the edge in IoT, then why not bring all your intelligence and analysis as close to the edge, the source, as possible, with all the obvious benefits. And it’s also where those promised forecasts on edge computing and IoT come in.

**Edge computing and IoT predictions**

According to IDC (data announced in its November 1, 2017, worldwide IoT forecasts webcast) by 2020, the IT spend on edge infrastructure will reach up to 18% of the total spend on IoT infrastructure. That spend is driven by the deployment of converged IT and OT systems which reduces the time to value of data collected from their connected devices IDC adds. It’s what we explained and illustrated in a nutshell.

According to a November 1, 2017, announcement regarding research of the edge computing market across hardware, platforms, solutions and applications (smart city, augmented reality, analytics etc.) the global edge computing market is expected to reach USD 6.72 billion by 2022 at a compound annual growth rate of a whopping 35.4 percent.

The major trends responsible for the growth of the market in North America are all too familiar: a growing number of devices and dependency on IoT devices, the need for faster processing, the increase in cloud adoption, and the increase in pressure on networks.

In an October 2018 blog post, Gartner’s Rob van der Meulen said that currently, around 10% of enterprise-generated data is created and processed outside a traditional centralized data center or cloud. By 2022, Gartner predicts this figure will reach 50 percent.

Gartner’s definition of edge computing: “Gartner defines edge computing as solutions that facilitate data processing at or near the source of data generation. For example, in the context of the Internet of Things (IoT), the sources of data generation are usually things with sensors or embedded devices. Edge computing serves as the decentralized extension of the campus networks, cellular networks, data center networks or the cloud.”