EVOLUTION OF THE INTELLIGENT INDUSTRIAL EDGE

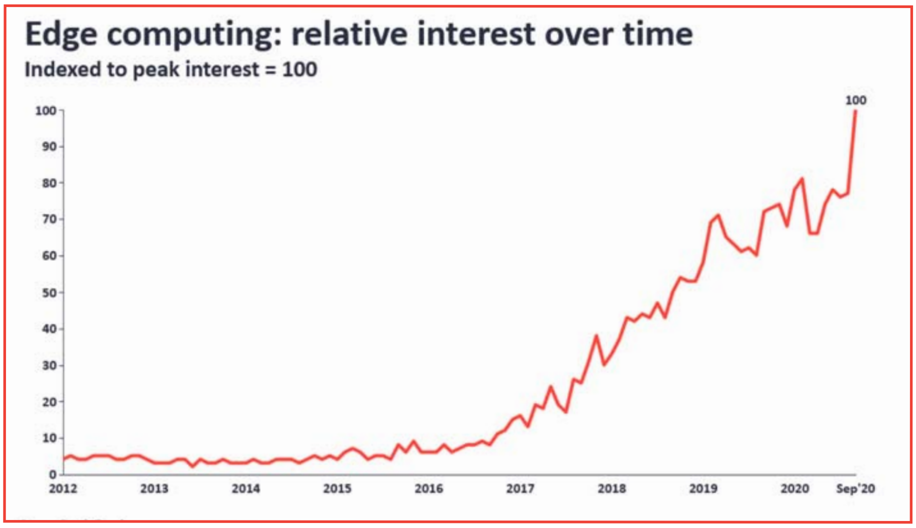
RESEARCHED BY

IOT ANALYTICS
MARKET INSIGHTS FOR THE INTERNET OF THINGS

SPONSORED BY

SIERRA WIRELESS™

1. INTRODUCTION INTO EDGE COMPUTING



Since 2017, public interest in edge computing has increased ten-fold, according to Google Trends.

Figure 1: Edge computing: relative interest over time

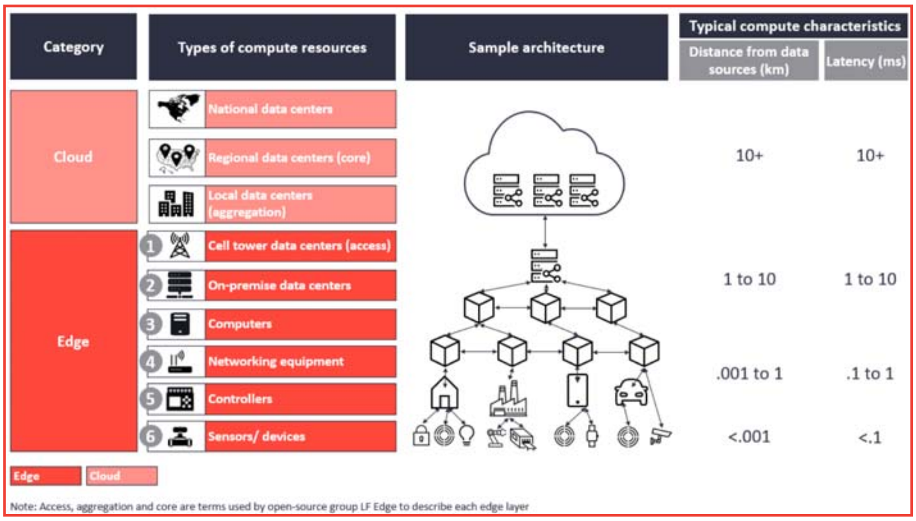
Why has the interest in edge computing become so widespread in recent years?

The primary reason why the edge has become so popular in recent years is because the edge as we know it is becoming increasingly intelligent. This "intelligent edge" opens up a whole new set of opportunities for software applications and disrupts some of today's edge to cloud architectures on all 6 layers of the edge.

Intelligent edge compute resources are replacing "dumb" legacy edge compute resources at an increasing pace. The former makes up a small portion of the market today but is expected to grow much faster than the overall market and thus gain share on the latter. The hype about edge computing is warranted because the replacement of "dumb" edge computing with intelligent edge has major implications for companies in all sectors: from consumer electronics and machinery OEMs to manufacturing facilities and oil and gas wells.

Benefits of switching from "dumb" to "intelligent" edge computing architectures include an increase in system flexibility, functionality, scalability and in many cases a dramatic reduction in costs; in some instances it has been reported that the recent switch from "dumb" to modern intelligent edge computing architectures has resulted in 92% industrial automation cost savings.

Where is the edge?



Category	Types of compute resources	Sample architecture	Typical compute characteristics	
			Distance from data sources (km)	Latency (ms)
Cloud	National data centers			
	Regional data centers (core)		10+	10+
	Local data centers (aggregation)			
Edge	1 Cell tower data centers (access)			
	2 On-premise data centers		1 to 10	1 to 10
	3 Computers			
	4 Networking equipment		.001 to 1	.1 to 1
	5 Controllers			
	6 Sensors/ devices		<.001	<.1

Edge Cloud

Note: Access, aggregation and core are terms used by open-source group LF Edge to describe each edge layer

Source: IoT Analytics Research 2020

Figure 2: Types of cloud and edge compute resources

A lot of great work has been done in recent years to define and explain "the edge". Cisco was an early thought leader in the space, conceptualizing the term "fog computing" and developing IoT solutions designed to run there. LF Edge (an umbrella organization under the Linux Foundation) publishes an annual "State of the Edge" report providing a modern, comprehensive and vendor-neutral definition of the edge. While these broad definitions are certainly helpful, the fact is that the edge is often "in the eye of the beholder".

For instance, a telecommunications provider may view the edge as the micro data center located at the base of a 5G cell tower (often referred to as "Mobile Edge Computing" or MEC). A manufacturing end user, on the other hand, may regard the edge as the vision sensor at the end of the assembly line. The definitions are different because the goals and purposes of hosting workloads at the edge are different: the telco provider is trying to optimize data consumption (i.e. performance issues associated with data consumers), while the manufacturing end user is trying to optimize data generation (i.e. performance issues associated with transmitting and analyzing the data).

IoT Analytics defines edge computing as a term used to describe intelligent computational resources located close to the source of data consumption or generation. "Close" is a relative term and is regarded rather as a continuum than a static place. It is measured by the physical distance of a compute resource from its data source.

Types of edge computing resources

There are six types of edge computing resources, which vary in performance, form factor and distance from the data source. The largest types of compute resources (both in size and compute capacity) are:

1. **Cell tower data centers**, which are rack-based compute resources located at the base of cell towers, and
2. **On-premise data centers**, which are rack-based compute resources located at the same physical location as the sensors generating the data.

Both are typically equipped with components (e.g. high-end CPUs, GPUs, FGPAs) designed to handle compute intensive tasks / workloads, such as data storage and analysis. These resources are typically found between 100 m and 40 km from the data source.

The next three types of compute resources aggregate data from the sensors / devices generating the data. They are typically equipped with middle-tier processors (e.g. Intel i-series, Atom, Arm Cortex-A7A) and sometimes include AI components, such as GPUs or ASICs. These resources are typically found between 1 m and 1 km away from the data source:

3. **Computers** are generic compute resources located outside of the data center (e.g. industrial PCs, panel PCs).
4. **Networking equipment** are intelligent routers, switches, gateways and other communications hardware primarily used for connecting other types of compute resources.
5. **Controllers** are intelligent PLCs, RTUs, DCSs and other related hardware primarily used for controlling processes.

Finally, the smallest types of edge compute resources are

6. **Sensors or devices**, which are physical pieces of hardware that generate data and / or actuate physical objects. They are typically equipped with low-end processors (e.g. Arm Cortex-M3) due to cost and power consumption-related constraints. They are located at the very farthest edge in any architecture, and as the data-generating devices themselves, their distance from the data source is essentially zero.

The industrial edge

Edge computing can be deployed in a variety of environments, but one of the most common is the industrial setting.

Industrial edge computing describes intelligent edge computing resources deployed at industrial sites (e.g. manufacturing facilities, power plants), which run analytics (e.g. rules engines, inference), control (e.g. PLC logic), data ingestion (e.g. OPC servers), storage (e.g. databases) and / or visualization (e.g. HMI) workloads.

2. THE EVOLUTION OF INTELLIGENCE AT THE EDGE

What is intelligence about?

Modern intelligent edge computing architectures are the driving force behind the move to more edge computing and the value-creating use cases associated with the edge. Seven key characteristics distinguish modern intelligent edge computing from legacy systems.

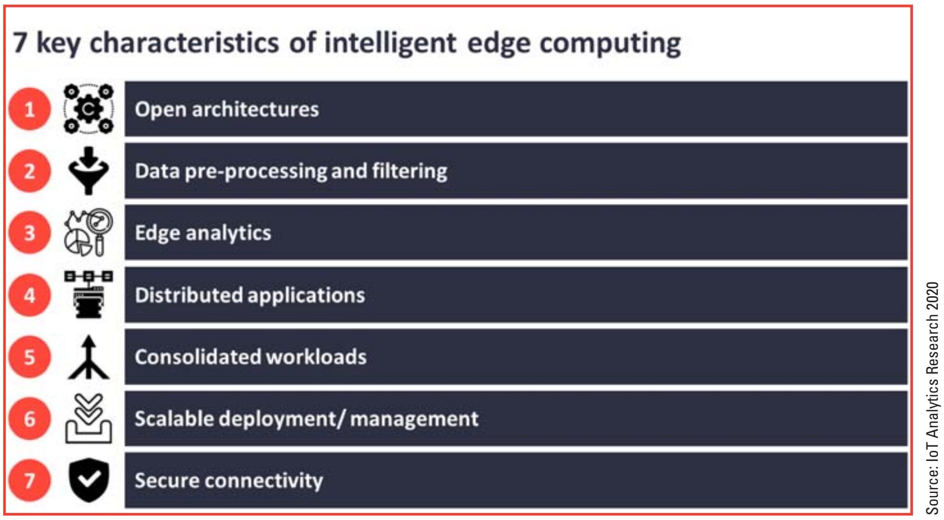


Figure 3: 7 key characteristics of intelligent edge computing

1. Open architectures

Proprietary protocols and closed architectures have been commonplace in edge environments for decades. However, these have often proven to lead to high integration and switching costs as vendors lock their customers in. Modern, intelligent edge computing resources deploy open architectures that, among others, leverage standardized protocols (e.g. CANOpen, OPC UA, MQTT).

2. Data pre-processing and filtering

Transmitting and storing data generated by legacy edge computing resources in the cloud can be expensive and inefficient. Legacy architectures often rely on poll / response setups, in which a remote server requests a value from the "dumb" edge computing resource device on a time interval, regardless of whether or not the value has changed. Intelligent edge computing resources can pre-process data at the edge, while only sending relevant information to the cloud, thus reducing data transmission and storage costs.

3. Edge analytics

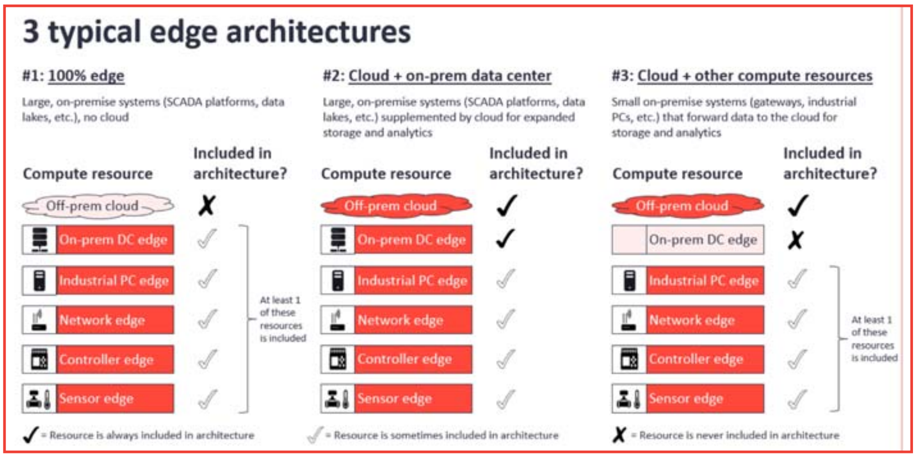
Most legacy edge computing resources have limited processing power and can only perform one specific task or function (e.g. sensors ingest data, controllers control processes). Intelligent edge computing resources typically have more powerful processing capabilities designed to analyze data at the edge. These edge analytics applications enable new use cases that rely on low latency and high data throughput.

4. Distributed applications

The applications that run on legacy edge computing devices are often tightly coupled to the hardware on which they run. Intelligent edge computing resources de-couple applications from the underlying hardware and enable flexible architectures in which applications can move from one intelligent compute resource to another. This de-coupling enables applications to move both vertically (e.g. from the intelligent edge computing resource to the cloud) and horizontally (e.g. from one intelligent edge computing resource to another) as needed. Three types of edge architectures are used to deploy intelligent edge applications:

I. **100% edge architectures**, where all compute resources are on premise. These architectures are often used by organizations that do not send data to the cloud for security / privacy reasons (e.g. defense suppliers, pharmaceutical companies) and / or large organizations that have already invested in on-premise computing infrastructure.
II. **Cloud + on-premise data center architectures** include an on-premise data center with cloud compute resources and optional additional edge compute resources. These architectures are often used by large organizations that have invested in on-premise data centers but leverage the cloud to aggregate and analyze data from multiple facilities.
III. **Cloud + other compute resources architectures** always include cloud compute resources connected to at least one smaller edge compute resource (i.e. not an on-premise data center). These architectures are often used to collect data from remote assets, which are not part of the existing plant network.

The choice of architecture and where an application is distributed in the architecture is largely dependent on the specific edge application and use case.



Source: IoT Analytics Research 2020

Figure 4: Three main types of intelligent edge architectures (Source: IoT Analytics)

Modern edge applications need to be architected so that they can run on any of the three edge architectures. Lightweight edge “agents” and containerized applications in general are two examples of modern edge applications which enable more flexibility when designing edge architectures.

5. Consolidated workloads

Most “dumb” edge computing resources run proprietary applications on top of proprietary real-time operating systems (RTOS) installed directly on the compute resource. Intelligent edge computing resources are often equipped with hypervisors, which abstract the operating system and application from the underlying hardware. This enables an intelligent edge computing resource to run multiple operating systems and applications on a single edge device. This in turn, leads to workload consolidation, which reduces the physical footprint of the compute resources required at the edge. Device and equipment manufacturers that previously relied on multiple physical compute resources can enjoy the benefits of lower cost of goods sold.

6. Scalable deployment/ management

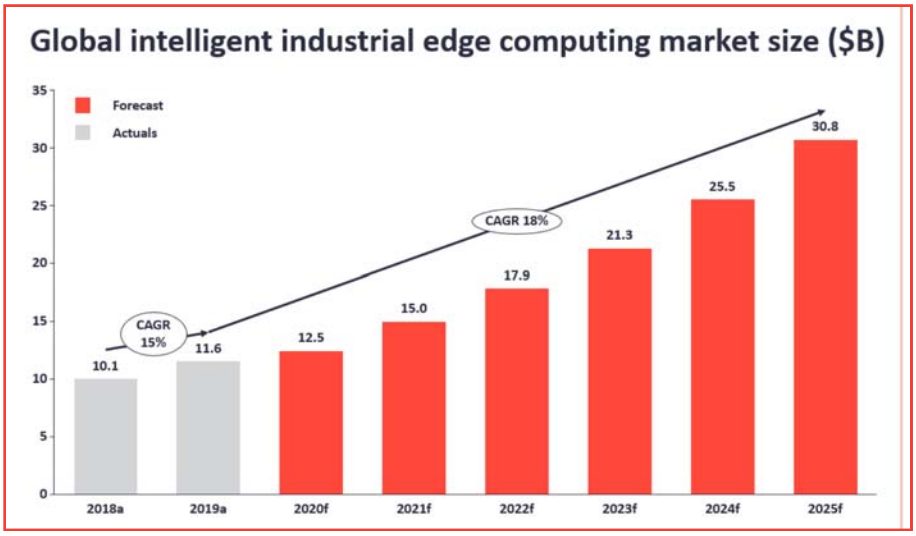
Legacy compute resources often use serial (often proprietary) communication protocols which are difficult to update and manage at scale. Intelligent edge computing resources are securely connected to local or wide area networks (LAN, WAN) and can thus be easily deployed and managed from a central location. Edge management platforms are increasingly being used to handle the administrative tasks associated with large scale deployments.

7. Secure connectivity

“Security by obscurity” is a common practice for securing legacy compute devices. These legacy devices often have proprietary communication protocols and serial networking interfaces, which do add a layer of “security by obscurity”; however, this type of security comes at a cost of much higher management and integration costs. Advancements in cybersecurity technology (e.g. hardware security modules – HSMs) are making it easier and safer than ever to securely connect intelligent devices. Different levels of security can be provided throughout the product lifecycle depending on the specific needs of the application.

The market for intelligent industrial edge computing

Intelligent industrial edge computing will make up an increasingly large share of the overall industrial automation market, growing from ~7% of the overall market in 2019 to ~16% by 2025. The total market for intelligent industrial edge computing (hardware, software and services) reached $11.6B in 2019 and is expected to increase to $30.8B by 2025.



Source: IoT Analytics Research 2020

Figure 5: Global intelligent industrial edge computing market, 2018-2025. Source: IoT Analytics

The intelligent industrial edge computing market is comprised of 3 segments: hardware, software and services. Software makes up the largest portion of the intelligent industrial edge market (41% in 2020) and is expected to grow faster than edge computing hardware and services.

3. CASE STUDIES OF INTELLIGENT INDUSTRIAL EDGE COMPUTING

How Ekatra makes fertilizer tanks smart

"Out of all the 30 vendors we evaluated, Octave was the only solution that took care of the entire edge-to-cloud infrastructure, allowing us to confine our focus to building the business application"
Alexey Klimenko, CTO at Ekatra

Background. Ekatra, a US-based provider of IoT security solutions, worked with Sierra Wireless to develop an intelligent industrial edge computing solution. The company helps organizations utilize large numbers of high-value physical assets and protect them against theft, tampering, and disrepair. Ekatra's proprietary solutions protect assets via smart IoT devices with specialized sensors that learn to predict and prevent disasters in a unique way with Edge AI machine learning.

The situation. Ekatra wanted to build an Industrial IoT (IIoT) application that would detect leakage and prevent overflows on fertilizer tanks, enabling agriculture companies to minimize labor and material costs associated with corroded tank walls, and to avoid hefty regulatory fines related to soil degradation and environmental damage. However, as the company's founder and CTO, Alexey Klimenko, explains, Ekatra is not an IoT gateway company, and did not want to spend time and money on building the underlying IIoT infrastructure for the application. They decided to build the remote monitoring application using Sierra Wireless "Octave", an all-in-one edge-to-cloud solution for connecting industrial assets. Octave offered Ekatra a wide range of capabilities and benefits in one solution.

Short time-to-market by using an all-in-one solution. In order to get to speed up the time-to-market, Ekatra decided to go with an end-to-end solution that integrates natively with edge devices, comes with device management capabilities, provides the necessary cellular connectivity, and comes with certain cloud APIs.

"It took us just three months to bring out IoT application to market. Without Octave, it would have taken six months"
Alexey Klimenko, CTO at Ekatra

"Our strategy is to do one simple product that can go everywhere. As the gateway is part of the Legato common platform, we don't have to re-write 6,000 lines of code to go from 3G to 4G"
Damien Hoyen, Project leader IoT at Atlas Copco

"A traditional 4G product is tailored for a particular region in terms of frequency bands, but low power wide area connectivity is much simpler. With a software-defined radio, you can support all the bands used worldwide in one product. LTE-M also supports roaming out of the box, which is a big plus."
Damien Hoyen, Project leader IoT at Atlas Copco

Edge analytics & pre-processing. Octave's intelligent edge proved critical in ensuring that fertilizer does not overflow during a refill and thus does not cause hundreds of thousands dollar damages. This was done by having Octave's intelligent edge configured to automatically shut down the tank's intake valves. Furthermore, Octave allows Ekatra to make instant changes to its edge processing logic as business requirements change, in many cases within just a couple of hours.

Open architectures. Octave supports various standard protocols such as Modbus, or CANopen. It was therefore ideally positioned to connect to the fertilizer tanks for which Modbus had been chosen as the protocol of choice.

Consumption-based pricing. Unlike standard cellular plans, which are built around megabyte consumption, Ekatra pays for the end-2-end solution by number of messages sent. This makes budgeting and planning of the entire solution more predictable and provides Ekatra with much needed flexibility.

The characteristics of intelligence used in the Ekatra solution:

Characteristic	Used?	How?
Open architectures	✔	Support for standard Modbus and CANopen protocols
Data pre-processing and filtering	✔	Edge pre-processing used to optimize costs associated with consumption-based pricing
Edge analytics	✔	Edge logic automatically shuts down tanks to prevent overflows
Scalable deployment/ management	✔	Ability to instantly update edge applications as equirements evolve
Secure connectivity	✔	Secure, global cellular connectivity

How Atlas Copco connects air compressors remotely

IOT ANALYTICS
MARKET INSIGHTS FOR THE INTERNET OF THINGS

Author of this Report,
Matthew Wopata
Principal Analyst at IoT Analytics
Matthew Wopata is a principal analyst covering Industrial IoT and Industry 4.0. He has 10+ years of IIoT / I4.0 experience including prior roles in product management, systems integration and strategy. Matthew was the lead author of IoT Analytics' Industrial Edge Computing Market Report, published in October 2020

Background. With 34,000 employees and annual revenues of about $9.5 billion, Stockholm-based Atlas Copco is one of the world's leading suppliers of air compressors, vacuum solutions, generators, pumps, power tools, assembly systems and other industrial equipment.

The situation. Atlas Copco has connected its air compressors remotely with the overall objective to enhance its products and services, helping customers avoid downtime. Air compressor downtime can cost millions of euros, for example in the case of a large chemical plant. Atlas Copco chose Sierra Wireless' device-to-cloud solution to connect their assets. The solution utilizes an FX30 programmable IoT gateway that runs the Legato® open-source Linux platform and is tightly integrated with the Sierra Wireless AirVantage® IoT software platform. The new setup enabled Atlas Copco to quickly develop, deploy and scale its SmartBox solution as a worldwide IoT service focused on customer success. SmartBox is able to support 2G, 3G, 4G, LTE-M and NB-IoT connectivity.

Benefits of the overall IoT solution. The new solution provides access to up-to-date information on the usage and condition of a customer's compressor. Local service teams can now deliver the right service visit at the right time for faster fault resolution, thereby saving costs associated with breakdowns and production loss. The data collected from air compressors also enables Atlas Copco's customers to manage their energy usage efficiently and further increases product reliability.

Benefits of the technical setup using an edge gateway. The FX30 IoT gateway supports not only NB-IoT and LTE-M, but also 2G as a fall-back connectivity option, giving Atlas Copco different options regarding the connectivity. The gateway is able to change its frequency to adapt to most cellular networks across the world. Therefore Atlas Copco is able to deploy the same connectivity module in all the markets where it operates. This allows the company to cater to its customer base in more than 180 countries using this new solution.

The characteristics of intelligence used in the Atlas Copco solution:

Characteristic	Used?	How?
Distributed applications	✔	Low-latency applications such as data ingestion and fault detection can occur at the edge, while applications requiring more compute power such as data mining and analysis can occur in the cloud.
Consolidated workloads	✔	Edge hardware running open Linux operating systems can run multiple edge applications
Scalable deployment/ management	✔	Global device management platform streamlines edge device / application updates
Secure connectivity	✔	Secure, global cellular connectivity