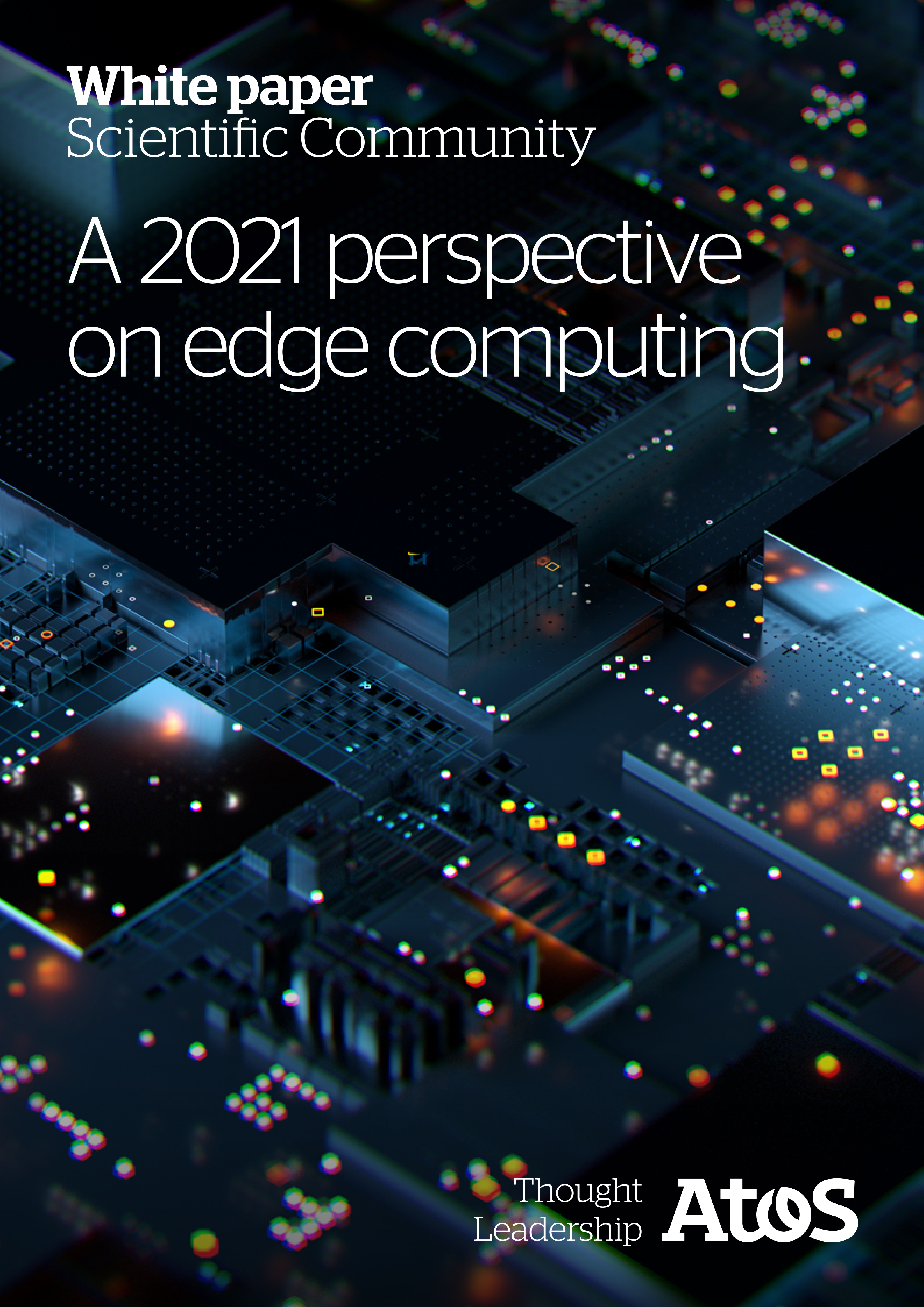


White paper
Scientific Community

A 2021 perspective on edge computing

Thought
Leadership

Executive summary

Edge computing continues to be a major IT trend in 2021.

This paper provides a review on the state-of-the-art and today's edge computing practices, specifically focusing on the software tools that enable execution of workloads at the edge. Our intention is to introduce the different scenarios for edge computing and analyze the software available.

We have provided a review of the existing development frameworks for each type of IoT edge device, as well as the open source frameworks and toolsets that enable the development of IoT edge services and applications.

Additionally, we explore key issues for the future development of edge services, such as the operation of services at the edge and the status of edge standardization efforts.

Finally, we provide our view on the complexity that edge computing brings to the IT landscape and outline our perspective on how edge computing may evolve.

Introduction

The emergence of edge computing is linked to the need to process data close to data generation sources in IoT devices. According to the latest estimates, we could see more than 75.44 billion IoT devices by 2025[1] — a 5X increase since 2015. In this context, the challenge is not simply managing IoT devices but more importantly, **how to cope with the volume of data and content that connected devices will produce**.

In today's prevailing IoT architecture, sensor data is transmitted over a wide area network to be centralized, processed and analyzed — which creates an additional supply of enriched data. These data and analytical models are intended to trigger actions either on the thing itself, in upstream business systems, or in other platforms that can use the data outside of the original context.

Unfortunately, this IoT approach is unsustainable in the long term, due to the number of device connections, volume of data, latency across different locations and networks, and the asynchronous nature of many connections between data flow and analytical cloud services. In addition, today's heterogeneous networks are unable to manage the massive growth anticipated in the number of endpoints and the data volume. Finally, smart ways to distribute the data are not yet available.

These challenges have created a pressing need to move information and analytical models closer to the source of data, in order to provide compute capability inside an environment where connectivity and response times can be controlled. To address these needs, a new class of edge computing and connectivity has emerged. IDC predicts that:

> "By 2023, over 50% of new enterprise IT infrastructure deployed will be at the Edge rather than corporate datacenters, up from less than 10% today; by 2024, the number of apps at the Edge will increase 800%." [2]

At present, IoT devices are not only proliferating, but simultaneously increasing in sophistication. The compute demands of artificial intelligence (AI) processes demand that future IoT environments are composed of more than simple sensors with 8-bit microprocessors.

Gradually, they will be populated by devices capable of porting diverse sensors and actuators, combined with heterogeneous (general purpose and GPU) microprocessors. In fact, IDC expects this trend to be the main source of growth for the microprocessor industry, with edge devices representing 40.5% of the market by 2023.[3]

The proliferation of these AI-powered rich IoT edge devices (the so-called "empowered edge" [4]) will enable the presence of an increasingly growing computing continuum — ranging from the cloud to numerous devices at the edge referred to as "autonomous things" by business analysts like Gartner[34].

This paper will provide deeper insight into state-of-the-art of edge and edge computing, with a specific emphasis on the software tools which enable execution of workloads at the edge. Our goal is to identify the technologies that enable edge implementations, understand the market adoption status, and highlight the technologies and vendors that — in our opinion — will lead the way.

With this in mind, we analyzed existing edge computing software frameworks from two perspectives: (1) a device perspective, providing a study of existent development frameworks by the type of IoT edge device; and (2) open source frameworks and toolsets which can facilitate the development of IoT edge solutions. Finally, we will highlight the fundamental challenges of managing the growing complexity that edge computing brings to the IT landscape.

Before we dive deeper, let's first define a few terms:

Edge computing, according to the Open Glossary[5], refers to the delivery of computing capabilities to the logical extremes of a network in order to improve the performance, operating cost and reliability of applications and services. By shortening the distance between devices and the cloud resources that serve them and reducing network hops, edge computing mitigates the latency and bandwidth constraints of today's Internet, ushering in new classes of applications.

Technology vendors (data center OEMs and software manufacturers) take the edge opportunity in order to push to the market IoT gateways, edge servers, edge computing platforms and solutions. Every endpoint and its associated edge infrastructure become a de facto extension of the cloud. As a consequence, overall management complexity increases substantially. By the same token, the cloud is also transformed because the stream of data is included in a more complex workflow to orchestrate, adding real-time constraints to decision making to loop back to the edge.

For the purposes of clarity, when we refer to "the edge" in this document, we define it as a heterogeneous set of edge components such as edge devices, edge gateways and edge servers, interlinked by edge connectors.

Edge devices are able to collect and process data from physical things in the field. Edge devices incorporate transducers to connect with the physical world (sensors and/or actuators); some processing capability (memory, microcontroller, AI accelerators) and some sort of communications. An edge device has also a uniquely identifiable address for the given network.

More and more edge devices are gaining processing capabilities by incorporating not only microcontrollers, but full-fledged microprocessors. The execution of AI workloads is the main driver for these developments. Existing developments are expected to be surpassed soon, thanks to existing efforts in defining neuromorphic edge processors specifically designed for edge execution of neural networks and other AI workloads.

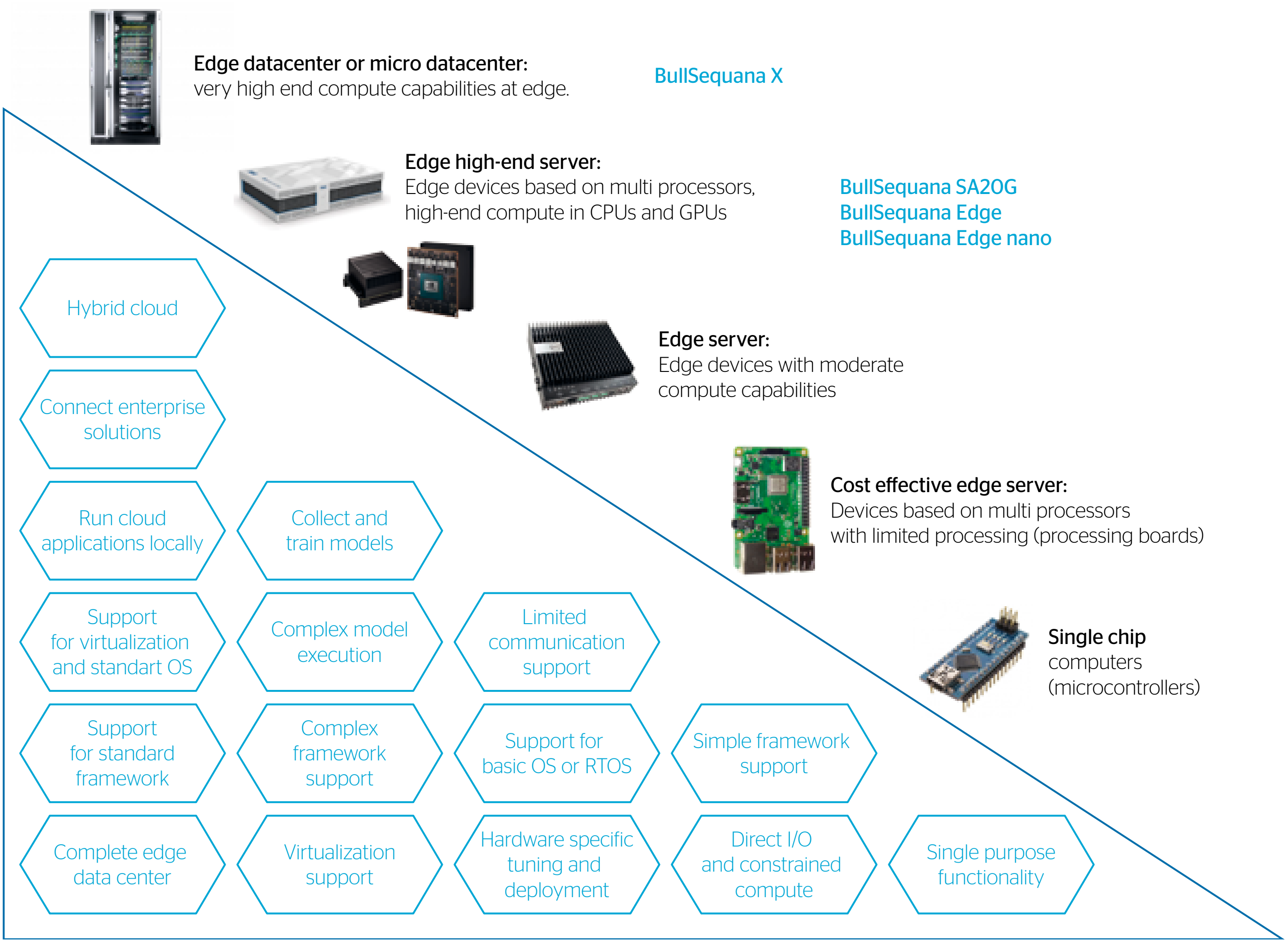
The increasing computing demand, driven by the need to provide more intelligent edge features, is prompting the emergence of innovative sets of compute boards designed to be embedded into real life objects.

Edge servers, according to Gartner, "collect data, deliver content, and perform analytics and decision making close to data producers (e.g., sensors and cameras) and data consumers (e.g., people and actuators). They can range from standard servers configured for edge processing (deployed in a data center/closet), to micro data centers that have self-contained power and cooling (deployed anywhere)" [6]. **Figure 1** represents the diversity of current edge servers and highlights their different capabilities.

Edge gateways (or IoT gateways) are intermediate devices between edge devices and the applications that create value from their data and access. Edge gateways allow the collection and secure transport of data from devices, remote users and applications. Edge servers with adequate software can act as edge gateways, and certain edge gateway devices include enough capacity to host edge services.

Edge connectors represent the different connectivity options (usually other than TCP/IP) between edge devices and gateways, in a wide range of specifications like bandwidth, power consumption and range.

Figure 1: Classification of edge servers


Edge datacenter or micro datacenter:
very high end compute capabilities at edge.
BullSequana X
Edge high-end server:
Edge devices based on multi processors,
high-end compute in CPUs and GPUs
BullSequana SA20G
BullSequana Edge
BullSequana Edge nano
Edge server:
Edge devices with moderate
compute capabilities
Cost effective edge server:
Devices based on multi processors
with limited processing (processing boards)
Single chip
computers
(microcontrollers)
Hybrid cloud
Connect enterprise solutions
Run cloud applications locally
Collect and train models
Support for virtualization and standart OS
Complex model execution
Limited communication support
Support for standard framework
Complex framework support
Support for basic OS or RTOS
Simple framework support
Complete edge data center
Virtualization support
Hardware specific tuning and deployment
Direct I/O and constrained compute
Single purpose functionality

Edge data handling and security

Every class of edge device has different processing capabilities in terms of compute and storage. This ranges from running embedded programs for device controls, performing business rule execution, to executing models that support complete business solution platforms. We classify data handling into the following broad categories:

- **Distribution only:** In this scenario, edge devices provide consumers with a closer access point for large amounts of content. Data is static or changes infrequently from a central control point. Data handling is focused primarily on fast distribution over a wide geography.
- **Simple processing:** Provides connectivity to high velocity data in order to aggregate data, qualify value, monitor certain patterns to reduce bandwidth, and to bring the right data to cloud for additional processing. This category of edge solutions is often involved in pre-processing data.
- **Complex processing:** Edge solutions in this category pre-process data before sending it to cloud. It also brings additional cloud-based business logic to the edge, which can take the form of applications for controlling assets or analytical models — including complex business rules. Some examples include handling video data and operating in disconnected environments or industrial environments where a near real-time response is required based on the data.
- **Multi-application processing:** This category not only supports complex processing, but also supports the deployment of entire data platforms including multiple applications. Processing and consumption of data is generally handled from the edge. These are high-end edge data centers that provide localized, completely independent solutions that can be part of a hybrid cloud setup. Edge devices in this category are more complex in nature, generally multi-node hardware with a virtualization layer.

One of the consequences of edge adoption is the need for new security paradigms and controls, in order to protect data and applications at the edge. Existing security mechanisms do not work well with distributed architectures, computing spread among multiple locations, and when most data is created, processed and consumed at the edge without entering the central data center perimeter.

Edge forces security to be more efficient than ever, and **edge security controls must incorporate intelligence**. Classic architectures typically benefit from "defense-in-depth" approaches, where multi-layered security controls protect the data hidden at the back-end. Such architectures can withstand some controls being defeated or having mispatched/misconfigured systems while still providing a high degree of confidence and resiliency, because other security layers provide assurance.

With the advent of edge, businesses can no longer afford such an approach. There is a need for more dynamic security controls that are able to adapt to heterogeneous environments without centralized monitoring and administration.

One clue to how the cybersecurity market is rapidly adapting to edge is the way analysts talk about the topic. Gartner recently coined the term "SASE" — which stands for Secure Access Service Edge. SASE describes a new network security model that combines multiple security controls such as Zero Trust Network Access (ZTNA), cloud access security broker (CASB), firewall as a service (FWaaS), data loss protection (DLP) and more, to provide assurance to the ongoing transformation of networking and security in the cloud.

Security at the edge can be examined at various levels, such as physical, internal or network related. The security risk areas can be classified as follows, even if specific edge devices may demand different types of security measures:

- **Physical or perimeter security:** The nature of edge computing deployment opens up new frontiers in which physical security and virtual security are equally critical. Physical threats could include tampering with devices to introduce malware through physical access, or unintentional actions that damage the device and data. Both threats must be addressed and are valid for every class of edge device. Countermeasures range from secure data center processes like access controls and audio/video monitoring, to monitoring temperature changes and movement or protecting against natural disasters like fires and floods.

  Edge devices deployed in the "wild" outside a controlled environment (e.g. a smart city solution) pose a greater risk. Countermeasures must be physically installed on the edge device itself, to act as a lock. For example, the Atos BullSequana Edge includes physical intrusion sensors which disable the system if an intrusion is detected.

- **Firmware and operating systems:** The next level of protection (most importantly from software threats) is the device's firmware, operating system and identity. In a world of billions of devices, protecting the identity of a device that sends data to your cloud services is imperative. False or corrupted data transmitted from an impostor device can harm dependent services. Hardware-based root of trust such as that provided by Azure Sphere[5] ensures that no device can be separated from its identity.

  Secure, boot-signed firmware and disk-level encryption are some additional measures that can be taken to secure the base of edge compute solutions. Remote update capabilities are the key to regularly updating operating systems and drivers from OEMs and other vendors.

- **Edge applications:** Security must be considered right from the start of application design, and specially formulated for the data and processes that run on the edge device. Critical data like certificates, data configurations and parameters must be protected.

  Selecting the right secure coding practices for each device class is essential. These include ensuring logged data does not contain sensitive information, that local data stores are encrypted, that credentials and other critical information are not stored in plain text on the device or written to a console, configuring users with minimum privileges and securing internal communication between edge components.

  As new devices and edge solutions emerge, it will be an ongoing challenge to stay up-to-date with new measures to secure the end-to-end chain and data handling of edge to cloud communication.

- **Edge networks:** Edge devices are connected today via various network types like Wi-Fi, LORAWAN, 4G, Sig-fox, NB-IOT and soon, 5G. They provide both inter-edge device connectivity as well as edge-to-cloud connectivity. Communication must be secured using encryption techniques for message transport as well as for the message itself.

  Consider a scenario where the cloud platform sends an unsecured command to an edge device to execute some operation. Typical "man in the middle" attacks allow an attacker to gain unauthorized control of devices.

  Depending on the edge device class and capabilities, it may be difficult to deploy additional network inspection tools. For example, a small footprint device like Raspberry Pi will not allow advanced network monitoring tools due to limited compute availability.

  Edge devices deployed in critical areas should not be directly connected to the Internet but via gateways, to avoid directly exposing the edge devices to external attack. For instance, edge devices deployed in a factory should connect to the shop floor network and use edge and network gateways to transmit data to / from the Internet.

Edge software frameworks

The cloud computing model relies on economies of scale, automation and high degrees of resource homogeneity. By employing standardized, homogeneous hardware platforms, cloud services are usually deployed in huge server farms that offer cost effective compute, storage and networking resources. In contrast, edge computing environments tend to be characterized by heterogeneity. The expected massive growth in connected IoT devices — together with a wide variety of use cases — brings diversity at multiple levels. Servers at the edge can range from simple microcontrollers embedded into products, to micro-data center installations with powerful compute and storage resources. Edge devices can be stationary and hard-wired, or battery-powered mobile devices. The need for battery power has its own strict requirements in terms of optimizing device energy usage and autonomy. This heterogeneity is illustrated in the edge device classification shown in **Figure 1.**

Fundamentally, edge device software development does not differ from any other type of software development. It is all about functions, events and triggers. However, there are several best practices to consider when developing edge device software:

- Use a **standard language**, including standard security supported by a broad variety of devices. This allows easier device upgrades, replacements and operations (including lifecycle management).
- Pay close attention to **lifecycle management** for device software. Over the long term, ensuring compatibility between different devices and with the cloud may become very challenging. Try to limit the number of combinations of supported software/releases for different devices.
- Take into account the **nature of the infrastructure**. Processing capabilities may vary, so decisions about what can be processed or analyzed where (on the device vs. in the cloud) may differ from device to device. In edge computing, analysis and intelligence should occur as close to the data source as possible to reduce the data transfer volume and processing latency. In turn, this reduces transmission costs and increases throughput, speed and quality of service (QoS). Real-time control functions must be executed at the edge.
- Create a robust multi-cloud **device test environment** to support the onboarding of new devices. Code can be tested locally and distributed to other devices after ensuring everything works correctly.
- Consider the **throughput and cloud connectivity** when making decisions about software distribution, lifecycle management and — in the case of development — where each type of data should be processed.
- Employ the **Agile development** methodology, which is the accepted norm to deliver business value as quickly as possible.

Edge frameworks support two main aspects of edge devices. First, they provide a platform to provision and run workloads near to the source of data with full lifecycle support. Second, they provide the means to orchestrate such workloads.
State-of-the-art edge frameworks have the following characteristics, and provide features that enable critical device functions:

Modularity	Modular architecture enables you to customize the solutions that need to run on edge. Depending on the use case, framework modules can be combined to support the modularization required.
Deployment flexibility	Containerization is not merely a cloud phenomenon, but is well suited to edge. Edge frameworks need to support flexible deployment models, and the resilient and portable nature of Docker containers makes them ideal for edge scenarios.
Manageability	By their nature, edge devices are generally deployed into inaccessible and widespread areas. Managing edge devices remotely with over-the-air (OTA) updates is essential to smoothly upgrade frameworks and deployed solution payloads.
Security	Edge devices are the direct and closest control units for “things,” and high control and access security is required at all levels — including access to the physical device, access to the OS, local data and communications.
Orchestration	Edge devices must perform multiple operations in parallel as well as bring a workflow flavor to solutions for ease of integration and adaptation. Support for orchestrating message flows and service calls is key.
Communication	Edge frameworks should be able to send data from edge to cloud, cloud to edge and even edge to edge, so you must support communication flows towards “things” as well as towards the platform. Abstraction, rather than low-level communication protocols, is needed to develop solutions which are focused only on business logic.
Device optimized	Edge devices do not have the infrastructure scalability of cloud, so be aware of the hardware limitations of your devices. Edge devices range from single chip MCUs to high-end edge data centers (and every combination in-between) but even in this extreme range, physical compute power is a limitation.
Support processing	Provide operational capabilities to process data flows at the edge. Edge use cases generally require solutions to process (filter, compute, analyze) data generated at the edge before it is sent to cloud. Some frameworks support machine learning models that are optimized for specific hardware.

The following sections provide an overview of software frameworks that support the development of solutions based on the classification of edge devices. For our analysis, we studied the characteristics of software frameworks for microcontroller edge devices, as well as software frameworks for mid-size edge servers and edge data center or micro data center devices.
In addition, we will analyze existing open source edge management tools for both edge data centers and mid-size edge computers.

Analysis of edge software frameworks from a device perspective

Microcontroller frameworks

A microcontroller unit (MCU) is a very low capacity device that is generally embedded as a part of a larger "thing." Microcontrollers embedded with use-case specific software solutions fall within this category. Billions of "things" — from toys to large industrial equipment — use microcontrollers today, connecting the devices to open up a wide range of business opportunities.

Various chip manufacturers and cloud providers are working to incorporate elements like running small computation and analytical models, local decision making or support for completely disconnected scenarios. A number of available IoT boards provide extremely battery efficient operation, along with multiple built-in connectivity options and development solutions. Some of these include:

- NXP i.MX RT[7] with AWS Alexa integration
- Azure Sphere MCU[8]
- Google Edge TPU[9]
- Texas Instruments SimpleLink[10]
- ARM Cortex Microcontroller Software Interface Standard (CMSIS) [11]with MBed OS
- Microchip Advanced Software Framework (ASF)[12]
- ATMEL Software Framework[13]

MCUs are based on customized real-time operating systems (RTOS) like AWS FreeRTOS[14], MBed OS[15] or TI-RTOS[16], and support IoT development using their own software development kits (SDKs). Developers can also use common programming languages like C/Python to ease the development.

MCUs like Azure Sphere or ARM MBed OS specifically address security aspects that are critical, considering the billions of devices connected to the Internet. Azure Sphere and ARM MBed OS provide hardware-printed root certificates which secure the device and cloud service connectivity. In addition, they provide a hardened OS and continuous security patching using cloud service.

In other cases, like the NXP i.MX RT solution, the MCU boards are pre-integrated with AWS Alexa, enabling them to bring voice-controlled features to any device. Voice processing like noise cancelling and connectivity to the AWS IoT core is also supported. Google offers Tensorflow lite on microcontrollers like Edge TPU to run analytical workloads on smart devices.

Integrated MCU kits, SDKs and cloud development support are increasingly making IoT connectivity and edge computation more accessible and easier to integrate.

Mid-size edge computer frameworks

While MCUs can be compared with single chip computers providing simple integration options, midsize edge devices are analogous to compute devices supporting multi-core CPUs and operating systems like Linux or Windows.

These are essentially miniaturized general-purpose computers supporting displays, USB ports, cameras and local device storage — thus providing support for higher compute solutions.

Devices are supported by a wide variety of edge frameworks and programming languages. A number of existing open source initiatives are beginning to offer handling data and applications at the edge, which are presented in detail in section 3.2. These open source offerings complement existing solutions from cloud services providers.

Public cloud providers like AWS, Azure and Google offer their own solutions, which provide integration with cloud-based solutions and bring common cloud solution elements to the edge. AWS Greengrass supports the execution of AWS Lambda functions on edge and brings AWS ML capabilities to edge devices. Similarly, Azure provides support for Azure Functions and Azure ML on edge.

In addition to specific tools, these frameworks enable remote management of solutions that run on top, which simplifies operations and updates. Azure IoT Edge is open source and provides support for running non-Azure specific modules on top.

With solutions like Azure Stack[17] and the very latest Azure Arc, cloud services on-premises can run their own workloads. Custom solutions to run applications or analytical models are still required, as some cloud features and services are missing in the on-premises implementation. With Azure IoT, developments can be simplified in so-called IoT edge modules. These can be easily tested and deployed to edge devices in a similar approach. IoT edge modules run as Docker containers and can execute in both Windows and Linux devices. This is the same model used for AWS Greengrass, which uses AWS Lambda functions for development[18] in order to easily test and transfer code locally or on the edge device. More details about these solutions can be found in a previous Atos Edge Computing Whitepaper[19].

In addition to edge framework offerings from public cloud providers, edge hardware vendors are beginning to offer software frameworks to help customers to exploit compute capacities in their platforms. An example of this approach is the NVIDIA EGX Edge Computing Platform[20] for the NVIDIA Jetson hardware family. The platform is compatible with Kubernetes management framework and AI execution, with the goal of offering real-time processing.

Atos offers its own edge management framework by means of Codex Smart Edge. This solution is presented in detail in section 7 the end of this paper.

Edge data center frameworks

In this category, we present a number of solutions from cloud hyperscalers which enable the local reproduction of public cloud environments at the Edge. These solutions are in fact, novel forms of hybrid cloud. Hybrid cloud compute is very important in use cases that have a strong requirement for data sovereignty or connectivity limitations, where IoT data generated at source needs to be processed efficiently at the edge.

Google Anthos[21] provides an on-premises application runtime environment which can run workloads near the data source. Anthos provides an optimized stack to execute workloads on top of bare metal servers at the edge. Anthos relies on GKE on-prem[22] and facilitates the creation, management and update of Kubernetes clusters in local environments, while providing a unified experience for application management across edge, internal and public cloud offerings.

Microsoft Azure Stack[23] offers on-premises Azure services. It consists of three main solutions: the previously introduced Azure Stack Edge, for execution of AI and general purpose applications at the edge; Azure Stack HCI, for the management of virtual machines in local infrastructures; and Azure Stack Hub, that brings the experiences of Azure cloud on-premises, permitting disconnected solutions. Like Google Anthos, it enables customers to develop unified DevOps experiences across diverse edge, local and cloud environments.

Along these lines, AWS has announced the general availability for AWS Outposts[24]. This solution develops the AWS experience in local infrastructures. It offers services for virtual machines and container management, databases, networking and data analytics.

In addition to offerings from major cloud providers, much experimentation has demonstrated how open source software management frameworks such as OpenStack can be used at the edge[25], typically in 5G network functions virtualization (NFV) scenarios.

More complex on-premises architectures based on traditional virtualization of container management are often considered to be "edge data centers," but are excluded from the scope of this paper.

Open source software frameworks for edge devices

This section analyzes some existing open source initiatives that currently revolve around developing solutions for edge computing management. It is important to note that this list is not intended to be exhaustive, but instead to presents some existing work that we consider of greatest interest.

Two different approaches can be found today in these solutions. First, solutions that enable low-cost compute devices to act as an IoT gateway (EdgeX Foundry, Eclipse Kura) and second, solutions that bring workload distribution features to the edge (KubeEdge, Starlingx).

Our main interest when presenting these solutions is to provide insights about existing developments and baseline solutions that can potentially be used to enrich more complex, business-driven use cases.

EdgeX Foundry

At the time of publication, EdgeX Foundry[26] offers a microservices software platform which enables users to build IoT gateway functionality out of an edge device. EdgeX Foundry software components allow you to obtain data from the so-called "South side" (IoT objects within the physical realm) with cloud services ("North side") in which data can be stored, aggregated, analyzed and converted into actionable information. It is important to note that the framework also contemplates actuation processes from North side to South side. EdgeX Foundry's architecture is structured in five main building blocks:

1

Core Services include the set of services which communicates and orchestrates the North and South sides. These services include persistent data repository for data obtained from South side objects, enablement of actuation requests from cloud (North side) to IoT objects (South side), metadata storage of the IoT objects connected to EdgeX Foundry instance and configurability of the instance.

2

Supporting Services, including monitoring, alert and notification.

3

System Management, which enables EdgeX Foundry services lifecycle management (installation, upgrade, start, stop).

4

Export Services Layer to develop the integration and data distribution to Northbound services.

5

Device Services Layer, which provides a set of services to allow interaction with IoT devices. Interestingly, it offers an extensible mechanism to build EdgeX Foundry Device Service microservice for new devices.

Eclipse Kura

Eclipse Kura has developed an approach similar to EdgeX Foundry,[27] with the goal of offering a platform for building IoT gateways. Eclipse Kura is a Java-based platform which offers diverse device abstractions and constructs basic gateway services using several drivers to encapsulate communication protocols and configuration. For data services, it employs a policy-driven data publishing system, and it offers an API mechanism for communications with cloud services.

KubeEdge

KubeEdge[28] develops an "open source system for extending native containerized application orchestration capabilities to hosts at Edge." It offers infrastructure services that combine networking, application deployment and metadata synchronization among cloud and edge environments. It is important to note that overall KubeEdge architecture considers a centralized edge management which is performed from the cloud level. The consideration of a cluster of devices at the edge currently requires integration among KubeEdge and standard Kubernetes. Its architecture is structured into edge and cloud parts, namely EdgeCore and CloudCore.

At the edge level, EdgeCore offers functionality to manage the lifecycle of applications (pods) at a single node level. In addition, it allows the configuration of monitoring probes, secrets, container runtimes and volumes in the device, as well as managing interactions among the different edge components and the cloud.

CloudCore comprises components for management of the different edge environments and permits the handling of both lifecycle management operations and operation from/to the edge environment. In addition, it offers device management functionalities making use of Kubernetes Custom Resource Definitions for device description. Device instances belong to a model and enable users to get static device data (specifications) and dynamic device data (status).

Open Stack Starlingx

Starlingx[29] describes itself as "a complete cloud infrastructure software stack for the edge used by the most demanding applications in industrial IOT, telecom, video delivery and other ultra-low latency use cases." Starlingx considers geographically distributed edge sites of diverse sizes governed under a central data center which offers orchestration and synchronization services. Its edge virtualization platform includes deployment options for bare metal, VM and container-based workloads. In addition, version 3.0 includes integration of OpenStack and Kubernetes on dedicated physical services.

Starlingx Infrastructure Management includes components for configuration management, providing node discovery and nodes configuration abilities; host management, which defines host interfaces and monitoring; service management, that takes into account high availability, messaging and service monitoring; fault management, allowing user definition of alarms and events, In addition, the forthcoming software management tool aims to support node software update and patches handling.

In addition, Starlingx offers a container platform based on Kubernetes Cluster that includes application management based in Helm as well as integration with private cloud management tools based on OpenStack.

Operating the edge

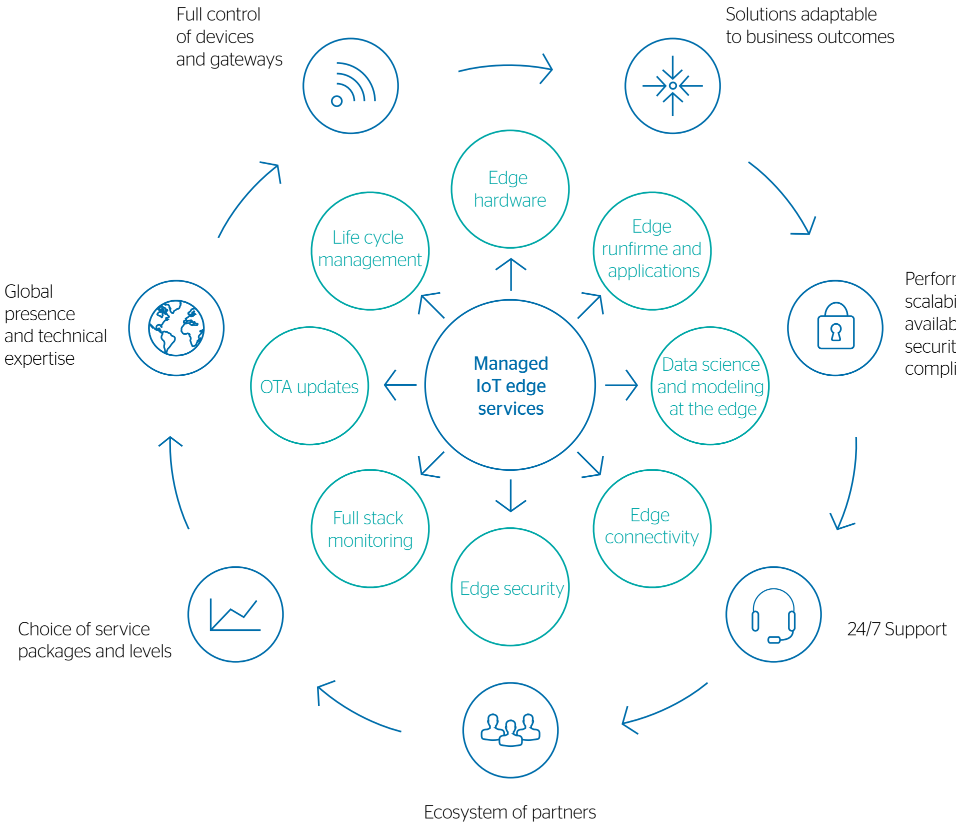
Edge solutions are becoming an essential part of the OT/IT landscape, delivering critical business outcomes. From a service management perspective, it is important to receive timely, accurate business alerts, preventive maintenance alerts, and the speed at which new devices become operational.

It is not just about keeping the "lights on" for edge, but the growing demand for data quality. Receiving a real-time sensor reading is fantastic, but how do you know if it's the right reading? What if other sensors show a different interpretation?

Keeping edge environments operational and ensuring data is processed appropriately creates some specific challenges:

- With hundreds or thousands of edge devices, gateways and servers, plus a potentially heterogeneous technical landscape, lifecycle management can quickly become complex.
- Some devices can be difficult to reach or in hostile environments, such as roadway infrastructure, nuclear power plants or offshore windfarms.
- Ownership, integrity and quality of data must be ensured, in environments that are typically exposed. In addition, potentially huge data volumes will require effective data retention and storage.

Figure 2: Service ecosystem for managed IoT services


Full control of devices and gateways
Solutions adaptable to business outcomes
Performance, scalability, availability, security and compliance
24/7 Support
Ecosystem of partners
Choice of service packages and levels
Global presence and technical expertise
Managed IoT edge services
Edge hardware
Edge runfirme and applications
Data science and modeling at the edge
Edge connectivity
Edge security
Full stack monitoring
OTA updates
Life cycle management

All aspects of edge — including hardware, applications, data and connectivity — need to be managed. Deep and diverse technical expertise is required to implement, connect and manage secure, resilient edge solutions, and to address the different challenges during operations. In addition, a solid ecosystem of trusted partners is essential to provide best-of-breed solutions. The ability to provide full control over a complex variety of edge components in a device-agnostic way is challenging, but crucial.

- Edge applications may include edge runtime such as a container management layer, as well as data analytics and other business applications or middleware. Custom application operations, middleware and interface management, container management, third-party application support, operating system patching, cloud integration and application performance monitoring are part of application management services.
- Edge data management may be limited to standard data storage and telemetry data handling, or may involve more advanced topics such as data analytics, video processing or AI capabilities at the edge. Business rules and alert management can be part of the required service management.
- Compliance with local, regional or global regulations may need to be handled as part of data management services, and end-to-end security needs to be tackled by considering options such as role-based access management, PKI certificate authentication, HSM, vulnerability assessment, end-to-end data encryption and secure industrial control systems. In most cases, both onsite and remote support is required on a 24/7 basis, with faster resolution time requirements than most traditional IT operations. Edge management SLAs may focus on complex concepts such as business continuity and data quality, in addition to or instead of traditional SLA terms such as availability

There is rarely a one-size-fits-all solution for addressing the above challenges. Each facet of the service layer is unique, and multiple software solutions are often required to address each function and problem. Finding the right tooling, skills and processes to solve your most pressing edge computing challenges represents an opportunity for service integrators.

In addition to the toolsets previously identified in this paper as edge frameworks, a complete service ecosystem for edge management must include additional features for managing the edge. Some typical tools required include:

Device managers

Solutions such as such as ARM Pelion[30] or Telit[31] can help manage edge apps on your hardware of choice, when combined with the edge workload management frameworks described in section 3.

Connectivity management platforms

When working with cellular networks, a connectivity management platform provides important services and management tools that enable you to manage the SIM lifecycle, including deployments, SIM activation or deactivation, changing tariff profiles or switching mobile networks.

Zero-touch provisioning

As outlined in section 2, there are now many MCU hardware devices with certificates embedded for major public cloud platform providers. This makes it possible to onboard devices to a platform using a zero-touch provisioning process.

Monitoring tools

Combine traditional infrastructure monitoring (Nagios) with containers (Prometheus), workloads and data (Dynatrace, ELK)

Standardization for edge

At the edge, highly standardized internet services connect with highly variable hardware and software installations on physical devices that form the endpoints of the IoT. On the Internet side of the edge, a proven set of protocols can be selected, as well as standards and services for computing, management and orchestration. Beyond the edge, a very diverse and specialized ecosystem exists.

As long as the scenarios remained static, it posed few problems — and these could be solved in a single integration effort. With the requirements for agile processes and configuration at the edge, it looks very different. The orchestration of devices, networks and services with the exchange of data must at least be manageable and, at best, highly automated.

In order to achieve this interoperability, it is essential to complete the (potentially semantical) description of each participant in the edge in machine readable form. To counteract this complex fragmentation, standards must be developed to address the different dimensions of edge computing.

Currently, there are several pure standardization efforts underway, as well as work on normalizing edge computing definitions and capabilities. The scope of these initiatives is quite wide, covering everything from a pure networking perspective to workload execution at the edge. It is important to note that some of these initiatives make use of the term "Fog computing." [32]

The following list presents some of the standardization initiatives that are currently ongoing as of the date of publication:

Initiatives related to edge computing concepts, term definitions and models:

- NIST Fog Computing Conceptual Model[33], NIST Formal definition of Edge Computing[34]
- IEEE P1934.1 - Nomenclature and Taxonomy for Distributing Computing, Communications and Networking along the Things-to-Cloud Continuum[35] (former OpenFog works)
- ISO ISO/IEC TR 30164:2020 Internet of things (IoT) — Edge computing[36]
- Linux Foundation's LF Edge Open Glossary for Edge Computing[37]
- Industrial Internet Consortium, The Edge Computing Advantage[38].

Initiatives presenting reference architectures and more technical approaches:

- ETSI Multi-access Edge Computing (MEC); Framework and Reference Architecture[39]
- 1934-2018 - IEEE Standard for Adoption of OpenFog Reference Architecture for Fog Computing[40].

Perspectives in edge computing

The term "**edge**" is all over the place. IoT use cases create massive amounts of data, and that data is often required to be processed at some level near its source. Edge computing addresses that challenge.

So, are we witnessing a new architecture paradigm emerging after the "move to the cloud," or are we simply going back to the proven "do the main compute on premises" model? What role was played by the strong movement towards digitalization incentivized by the COVID-19 pandemic?

Increasingly, sustainability and energy efficiency are critical aspects that warrant analysis in every new technology. Studies such as Lean ICT estimate that watching a 10-minute video online consumes as much electricity as a smartphone uses in ten days. [41]

Optimizing energy consumption in large data centers (such as those providing public cloud services) has been a large area of study and development in the last decade, and has achieved important advances like the definition of the PUE (Power Usage Effectiveness) metric.[42]

However, advances in edge computing bring their own set of challenges:

- **Rightsizing:** How to avoid oversizing the edge and duplicating cloud and edge resources.
- **Pressure for power efficient edge devices:** The compute per watt must be improved by several levels of magnitude.
- **Waste management** must consider the device's design mechanisms and parts traceability, so they can be decommissioned and recycled.
- The **heterogeneity of devices** capable of offering edge services creates a need to reconsider consolidated energy efficiency metrics in data centers, due to the existence of edge devices which do not require cooling systems.
- The **battery limitations** of devices, together with the demonstrated impact of network transmission in energy consumption required of specific edge

With edge computing installations growing and the new dimension of local connectivity between machines and devices, the role of the network will change and have an important impact on performance and structure. In addition, 5G campus networks will increase shop floor flexibility by eliminating the need to re-cable communication links. The feature of software defined networking (SDN) — where routes, bandwidth and other characteristics are managed dynamically — will elevate the edge computing principle in combination with network functions virtualization. The exploitation of computing capability inside the network itself will adapt to this flexibility with virtualizing payload processing to follow network configurations and topologies. In such a distributed network-edge computing ecosystem, the computing demands will reshape and optimize SDN beyond networking characteristics to form a dynamic edge-network computing fabric.

With the compute evolution at the edge, there is an increasing push to perform machine learning (ML) model training on the edge. ML model training is the most resource and time-consuming aspect of the AI execution lifecycle, combined with the complexity of cloud to edge synchronization. For this purpose, approaches such as "learning on the edge" and "on-device learning" are being developed. We anticipate significant developments in this direction both from hardware and edge software frameworks.

In our opinion, the current state of edge computing is an initial step towards an even more decentralized view of computing that fully exploits the emergent computing continuum enabled by the combination of "empowered edge" and "autonomous things".

At Atos, we use the term "swarm computing[43] for this new digital infrastructure, encompassing sophisticated IoT endpoints, edge and multiple cloud platforms working in continuous cooperation to connect entities in the context of large cooperative and self-organizing applications in the emerging compute continuum.

The edge computing paradigm is a driver for technology evolution. Several complementary technologies are also emerging, such as 5G communications. Edge computing is evolving from IoT towards a complement of cloud computing in the form of a distributed cloud and as an intermediate step to the computing continuum and swarm computing[44].

While some aspects and use cases of edge are quite accepted in the industry, the most advanced aspects of edge computing are still on the innovation curve of the hype cycle (see the Gartner Hype Cycle for Edge Computing, 2020). The solution provider market must consolidate, and elements of standardization still need to mature. Against this backdrop, we see several tensions that will drive the evolution of the edge technology landscape:

- **Hyperscalers** dominate the cloud and have the means and weight to push for their de facto standards to dominate massive deployments. On the other hand, an ecosystem of **specialized solution providers** will end up winning some niches or even specialized market sectors (for example NVIDIA in computer vision). **Hardware and software** providers and **telecom operators** will benefit from the shift to the edge.
- While **hyperconnectivity** driven by 5G is generally considered an enabler for edge and swarm (by enabling seamless connectivity within edge and to the cloud), it can also challenge the need for edge computing in latency or bandwidth-driven use cases.
- Due to the very definition of edge itself, increased data sources and more exposed data will condition edge and swarm adoption to specific environments. How do we ensure that data originates from trusted sources? Who owns this data? Who manages it and under what rules? Proper solutions and conventions must be in place to secure **data identity, integrity and sovereignty.**

Despite standardization and convergence, edge brings complexity and ecosystem diversity to the end-to-end value chain, which will represent a large opportunity for service integrators. Edge will drive the need for service offerings that address the specificities of on-premises and embedded solutions, in combination with increased demand for hybrid cloud capabilities.

Acknowledgements

The authors would like to thank **Gert Prieber** for early revisions and comments.

About the Authors

Vincent Couteau
Chief IP Transaction Counsel,
vincent.couteau@atos.net

Jordi Cuartero (Editor-in-Chief)
CTO IoT,
jordi.cuartero@atos.net

Ricky El-Qasem
Automation Domain Leader, Global Head of Application Platforming & Data Center Transformation,
ricky.el-qasem@atos.net

Ana Juan Ferrer (Editor-in-Chief)
Distinguished Expert, Edge Domain,
ana.juanf@atos.net

Kedar Joglekar
Technical Director (Architect),
kedar.joglekar@atos.net

Marc Llanes
Global Cybersecurity Business Development Director,
marc.llanes@atos.net

Martin Pfeil
CTO Global Siemens Account & IDM Germany, Distinguished Expert
martin.pfeil@atos.net

Jesus Ranz
IoT Device and Connectivity Engineer,
jesus.ranz@atos.net

Wolfgang Thronicke
Architect & Consultant for R&D Projects, Enterprise AI, and Mobile Solutions,
wolfgang.thronicke@atos.net

Purshottam Purswani
CTO-APAC, Business & Platform Solutions,
purshottam.purswani@atos.net

Jose Maria Cavanillas
Worldwide IoT Partner Management Director,
jose-maria.cavanillas@atos.net

Said Derradji
Lead Architect, said.derradji@atos.net

Francisco Jose Ruiz Jimenez
Technical Architect Manager, fjose.ruiz@atos.net

Annex 1: Atos edge assets

Atos Codex Smart Edge

Codex Smart Edge is highly scalable industrial IoT and edge computing platform provided by Atos. The platform offers a hardware agnostic software solution for machine-to-machine connectivity and distributed intelligence. It has the following features:

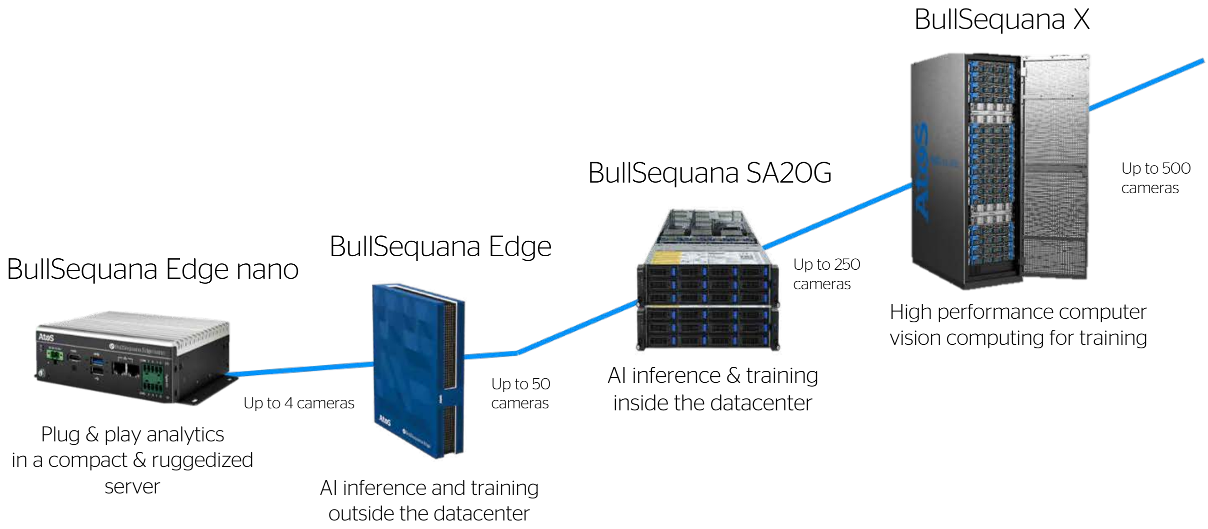
- Edge computing capabilities, from processing simple logic to complex AI/ML jobs at the edge for near real-time decisions
- Distributed mesh, which develops an interconnected smart node mesh to support layered architecture and swarm connectivity services
- Support for disconnected mode operation in scenarios with intermittent or limited connectivity, enabling zero data loss with local data buffering and historization
- Reduced bandwidth needs by using data decimation and multi-layer aggregation
- End-to-end orchestration and management thanks to Kubernetes and Mesh auto surveillance
- "Northbound" services for connectivity to IoT data platforms and clouds
- "Southbound" services to communicate and interact with IoT field devices supporting a diverse set of protocols
- Microservices architecture for extensibility and scalability, including Docker-enabled OS extension support

Atos Codex Smart Edge allows organizations to quickly build and deliver interoperability between devices, applications and services for many industrial use cases. It provides a foundation for industrial IoT and enables interesting development opportunities for the future.

Atos Codex Smart Edge includes ready-to-use native assets to quickly compose project applications aligned to business needs. Components range from industrial data collection and simple calculations to AI-based edge analytics and scoring, data visualization and export to major cloud platforms.

Atos Codex Smart Edge provides the flexibility required to securely and efficiently manage industrial operations in real-time from any location.

Atos edge computing server range

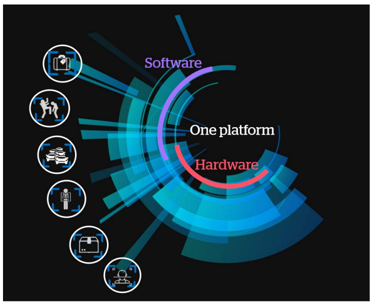

BullSequana X
Up to 500 cameras
High performance computer vision computing for training
BullSequana SA20G
Up to 250 cameras
AI inference & training inside the datacenter
BullSequana Edge
Up to 50 cameras
AI inference and training outside the datacenter
BullSequana Edge nano
Up to 4 cameras
Plug & play analytics in a compact & ruggedized server

Atos has designed a powerful set of edge servers to address the needs of edge computing from edge datacenter/cloud to far edge, which can be combined with Atos Computer Vision Platform, Codex Smart Edge and Edge Data Analytics (predictive analytics use cases).

Atos Computer Vision Platform

Atos Computer Vision Platform is a unique end-to-end computer vision platform providing pre-trained & customizable AI models powered by BullSequana server range and enriched by Atos computer vision experts through worldwide experts labs.

It enables to identify events and behaviours, to reduce error rates, to guarantee people and asset safety, to deliver highest quality, to offer frictionless and personalized customer experiences. Business and organisations keep up the paste of events and demand, by analyzing videos in real time at the edge to drive the best decisions.



- AI expertise through computer vision labs
- Pre-trained AI models
- Market leading software stack based on Ipsotek
- GPU-enabled hardware

Annex 2: edge use cases examples

This annex includes several Edge Use Cases Atos has developed that illustrate how Edge computing is becoming a reality in the context of different industries.

Automatic quality inspection

Quality inspection based on video captured by cameras (linear, 2D or 3D), imaging (RX or confocal) or electrical sensors

Industry: Manufacturing

Customer challenges: reliance on manual visual inspections, product recalls through defects passing undetected, increased production time for rework

Solution: Quality inspection, 3D vision inspection & 3D in-line metrology. Machine vision inspection systems where defects get classified according to their type and are assigned an accompanying grade or default. Edge enables real time detection and analysis of the camera images to raise the alert of a defect in a production line. ML models are executed on the Edge limiting data throughput and limiting latency and response time

Outcomes: Reduction in defects during a production assembly, reducing time to detect defects through manual inspections, increased yield and productivity

Smart control room: Prediction of events and quality

Industry: Manufacturing

Customer challenges: thousands of alarms generated every minute from machines, skilled technician detects key alarms that need attention. Increased time between detection of critical alert to reacting to alert, thereby increasing the chance of production outage.

Solution: Edge analytics applications run on the edge close to the production line and can predict in near real time critical events avoiding interruption to the production cycle and limiting the amounts of stream data uploaded to the cloud.

Outcomes: Reduced number of defects through early detection of critical events, thereby improving production uptime and yield, automation of operators work and reduction in the waste of materials used for production line.

Traffic Analytics

Industry: Transportation, Public Sector and Defense or Retail (People analytics using same solution, detecting people behavior in retail outlets)

Customer challenge: increased traffic through unplanned events, unauthorized access to restricted areas, Automatic Number Plate Recognition for traffic violations, smarter roads for vehicle safety

Solution: Smart camera with 4G/5G capability with built in analytics trained to detect vehicle behavior, transmits traffic violations to an Edge server near the camera. Multiple cameras in a stretch of road can send their violations to a single Edge server.

Outcome: Improved traffic management, preventing road accidents, capturing traffic violations, improved safety for drivers.

People and Asset Tracking

Industry: Multiple industries

Customer challenge: Track and measure equipment usage on the ground to ensure regular maintenance. Monitoring working conditions, lone worker safety.

Solution: LoRa beacons and Smart Edge software on LoRa gateways. Equipment tagged with GPS, LoRa or Sigfox transmitting geolocation data to Smart Edge nodes which convey onto factory/plant Edge Server for local analytics.

Outcomes: Increase the throughput of the maintenance team/productivity, improve overall operating efficiency, overall equipment effectiveness and decrease unscheduled down time.

Process Optimization for Wastewater Treatment plant

Customer challenge: manual inspections and reliance on onsite technicians to identify faults with the processing plant, with issues detected too late thereby causing unplanned outages of the facility.

Solution: Distributed mesh of Smart Edge based on smart nodes to support local micro services for value restitution at field, factory and corporate office level. Dedicated enclosures for field deployment.

Industry: Energy & Utilities

Outcome: Mass data collection and aggregation. Machine learning and advanced analytics embracing large data variety (including external data sources). Improved visibility of treatment plant with better maintenance schedules to reduce unplanned downtimes.

References

[1] IDC FutureScape: Worldwide IT Industry 2020 Predictions, https://www.idc.com/getdoc.jsp?containerId=US45599219

[2] IDC FutureScape: Worldwide IT Industry 2020 Predictions

[3] IDC: Worldwide Edge and Endpoint AI Processing Forecast, 2020–2024, https://emea.idc.com/getdoc.jsp?containerId=US45169219

[4] Gartner Identifies the Top 10 Strategic Technology Trends for 2020, https://www.gartner.com/en/newsroom/press-releases/2019-10-21-gartner-identifies-the-top-10-strategic-technology-trends-for-2020

[5] Open Glossary of Edge Computing [v2.1.0], - https://github.com/State-of-the-Edge/glossary/blob/master/PDFs/OpenGlossaryofEdgeComputing_2019_v2.0.pdf

[6] Bob Gill, Thomas Bittman, Chirag Dekate, Gartner Hype Cycle for Edge Computing, 2019, https://www.gartner.com/en/documents/3956137/hype-cycle-for-edge-computing-2019

[7] i.MX RT1010 Crossover MCU with Arm® Cortex®-M7 core , https://www.nxp.com/products/processors-and-microcontrollers/arm-microcontrollers/i-mx-rt-crossover-mcus/i-mx-rt1010-crossover-mcu-with-arm-cortex-m7-core:i.MX-RT1010?cid=ad_PRG353609_TAC356834_EETECH_PP06&gclid=Cj0KCQjwvb75BRD1ARIsAP6LcquHCuNF8_Ku1wwtpTziBFaipLtx5No3phTy6ORaHfmjPWDrRnjGLsEaAjegEALw_wcB

[8] Azure Sphere , https://azure.microsoft.com/en-us/services/azure-sphere/

[9] Edge TPU, https://cloud.google.com/edge-tpu

[10] Texas Instruments Microcontrollers (MCU), https://www.ti.com/microcontrollers/simplelink-mcus/overview.html

[11] Arm Developer CMSIS, https://developer.arm.com/tools-and-software/embedded/cmsis

[12] Microchip Advanced Software Framework (ASF), https://www.microchip.com/mplab/avr-support/advanced-software-framework

[13] Atmel Software Framework, https://gallery.microchip.com/packages/4CE20911-D794-4550-8B94-6C66A93228B8/

[14] AWS FreeRTOS, https://aws.amazon.com/freertos/

[15] ARM MBed OS, https://os.mbed.com/mbed-os/

[16] Texas Instruments RTOS, https://www.ti.com/tool/TI-RTOS-MCU

[17] https://azure.microsoft.com/

[18] AWS Lambda functions, http://aws.amazon.com/ lambda

[19] Atos Edge Computing Whitepaper, https://atos.net/wp-content/uploads/2019/04/Atos_EdgeComputing_WP-web.pdf

[20] https://www.nvidia.com/en-us/data-center/products/egx-edge-computing

[21] Anthos at the Edge, https://cloud.google.com/solutions/anthos-edge

[22] GKE on-prem overview, https://cloud.google.com/anthos/gke/docs/on-prem/overview

[23] Azure Stack, https://azure.microsoft.com/en-us/overview/azure-stack/#overview

[24] AWS Outposts , https://aws.amazon.com/outposts/

[25] OSF Edge Computing, https://www.openstack.org/edge-computing

[26] EdgeXFoundry: https://www.edgexfoundry.org/

[27] Eclipse Kura: https://www.eclipse.org/kura/

[28] KubeEdge: https://kubeedge.io/en/

[29] StarlingX: https://www.starlingx.io/

[30] ARM Pelion, https://www.pelion.com/

[31] Telit, https://www.telit.com/

[32] Typically, these make use of the differentiation among Edge and Fog terms made in https://www.nebbiolo.tech/wp-content/uploads/whitepaper-fog-vs-edge.pdf

[33] Fog Computing Conceptual Model, https://www.nist.gov/publications/fog-computing-conceptual-model

[34] https://www.nist.gov/publications/formal-definition-edge-computing-emphasis-mobile-cloud-and-iot-composition C. Mahmoudi, F. Mourlin and A. Battou, “Formal definition of edge computing: An emphasis on mobile cloud and IoT composition,” 2018 Third International Conference on Fog and Mobile Edge Computing (FMEC), Barcelona, 2018, pp. 34-42, doi: 10.1109/FMEC.2018.8364042.

[35] P1934.1 - Nomenclature and Taxonomy for Distributing Computing, Communications and Networking along the Things-to-Cloud Continuum , https://standards.ieee.org/project/1934_1.html

[36] ISO/IEC TR 30164:2020 Internet of things (IoT) — Edge computing https://www.iso.org/standard/53284.html

[37] Linux Foundation’s LF Edge, https://www.lfedge.org/projects-old__trashed/openglossary/

[38] IIC The Edge Computing Advantage, https://www.iiconsortium.org/pdf/IIC_Edge_Computing_Advantages_White_Paper_2019-10-24.pdf

[39] ETSI MEC Multi-access Edge Computing (MEC); Framework and Reference Architecture https://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/02.01.01_60/gs_MEC003v020101p.pdf

[40] 1934-2018 - IEEE Standard for Adoption of OpenFog Reference Architecture for Fog Computing, https://standards.ieee.org/standard/1934-2018.html

[41] The Shift project, LEAN ICT: TOWARDS DIGITAL SOBRIETY, https://theshiftproject.org/wp-content/uploads/2019/03/Lean-ICT-Report_The-Shift-Project_2019.pdf

[42] About PUE and DCiE, https://dcimsupport.ecostruxureit.com/hc/en-us/articles/360039292493-About-PUE-and-DCiE

[43] Swarm Computing Concepts, technologies and architecture, https://atos.net/wp-content/uploads/2018/12/atos-swarm-computing-white-paper.pdf

[44] https://atos.net/wp-content/uploads/2018/12/atos-swarm-computing-white-paper.pdf

About Atos

Atos is a global leader in digital transformation with 105,000 employees and annual revenue of over € 11 billion.

European number one in cybersecurity, cloud and high performance computing, the Group provides tailored end-to-end solutions for all industries in 71 countries. A pioneer in decarbonization services and products, Atos is committed to a secure and decarbonized digital for its clients. Atos operates under the brands Atos and Atos|Syntel. Atos is a SE (Societas Europaea), listed on the CAC40 Paris stock index.

The purpose of Atos is to help design the future of the information space. Its expertise and services support the development of knowledge, education and research in a multicultural approach and contribute to the development of scientific and technological excellence. Across the world, the Group enables its customers and employees, and members of societies at large to live, work and develop sustainably, in a safe and secure information space.

Find out more about us
atos.net
atos.net/careers

Let's start a discussion together

For more information: atos.net/scientific-community

CT-210610-RY-J3389-ATOS-WP-PERSPECTIVE