

April 2020

Edge computing for industrial affairs

Atos

Contents

3

Introduction

4

Why Edge computing

5

Edge computing enhance
“Cloud first” strategy

8

Artificial intelligence at the edge

10

Edge Data Center

12

Edge computing to move
towards a low-carbon
future

13

Conclusion



➤ Explore all of the edge computing
offer on bullsequaedge/atos.net

Introduction

By 2022, Gartner predicts that 80% of enterprise-generated data will be created and processed outside of the data center and cloud infrastructures compared with 10% today.* Edge computing will pick up this job by extending cloud computing capabilities and allowing us to deal with the data tsunami being created by the Internet of Things (IoT) revolution.

Edge will allow us to move away from a reliance on centralized data centers and remove the issue of networks having to transmit ever-increasing volumes of data. It will help us make optimal use of the growing reservoir of unstructured data produced by IoT devices and of rapidly maturing analytics technologies and will enable IoT devices to make full use of artificial intelligence and offer ever more opportunities for innovation.

It means real-time information processing at the source enabling faster reaction times. It will change the factories of the future, create smart cities and has the potential to completely change the way we travel through autonomous vehicles.

In Manufacturing, 1TeraBytes of production data is created daily by the average factory, but less than 1% of production data is analyzed and acted upon in real time.**We see that Artificial intelligence and Machine learning's ability to collect and handle big data and its applications in various manufacturing applications such as Industrial IoTs, predictive analytics and machinery inspection, quality control, and cybersecurity are growing***.

Edge computing is today a business reality in Manufacturing. Artificial Intelligence Computer vision, Lower latency, Autonomy, Security and reduced network costs are enabled by moving computing closer to the end-user and the Operational Technologies (OT).

This paper focuses on Edge computing for Computer vision, Data Analytics and Edge Micro Data Center (EDC) infrastructures for industrial affairs.

*Source: The Edge Completes the Cloud-A Gartner Trend Insight Report (09/18)

**Source: IBM Institute for Business Value. "The A.I. effect on industrial products." February 2018

***Source: Artificial Intelligence in Manufacturing Market Published Date: Jan 2019 | Report Code: SE 5470

Why edge computing?

Edge computing is an important element of the post-cloud era, extending rather than replacing the cloud.

It allows data to be processed rapidly at the edge, close to where devices are generating it.

The decentralized and distributed nature of edge computing avoids unnecessary network transmission to the cloud and enables the near real-time actuation of connected things.

Simply put, edge computing makes large-scale AI at the edge not only possible but also cost-effective. In doing so, it opens the door to unprecedented innovation.

Edge Computing is data driven with a significant growth potential that is boosting High Performance compute at the Edge.

A huge data explosion

The IoT is fast becoming an essential source of data, its volumes limitless. Unlimited AI, social networks, applications, sensors and captors, among other things, are only adding to the soaring data and content. By 2030, the total volume produced could reach 1 yottabyte - that's a trillion terabytes or a million trillion megabytes.

Data shifts to the edge

Around 75% of data is expected to be produced at the edge by 2020, with only 25% still produced within the data center.

A growing complexity of the data

More and more complex and unstructured data is produced at the edge. Data is produced from many various sources in different formats including text, voices, images, video streams, sounds and sensors.

Lessons learned from a large shipping company to equip vessels with edge computing

Our customer needed edge computing to collect and process data relating to the vessels technical installations:

Connected machines

All vessels assets (engine, engine cooling system and ventilation system) are connected and generate operational data which is consolidated in one common data lake (Tera bytes of data created every day).

Data-driven predictive maintenance

Atos combined the vessel-related data with external data (such as routes, weather, sea currents) and used Artificial Intelligence powered analytics to turn this into relevant insights on technical status, fuel consumption and turnaround time of vessels. Each vessel of the fleet is connected to a central system, which allows those high-value insights to be sent from land to sea in real time.

We used both edge containers & computing server, to run extremely advanced algorithms directly on the vessel - which gives to our customer the flexibility to limit the use of satellite communication and to optimize costs.

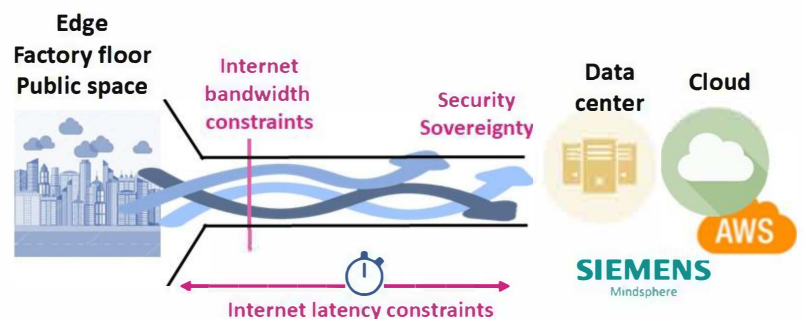
Last but not least, we optimized Waste heat recovery system - The solution enables re-cycling of the emitted exhaust gases and kinetic energy from engines as the power source for the vessel - which has a direct impact on the fuel consumption.

Edge computing enhance “Cloud first” strategy

Cloud computing is now the predominant mechanism for IT service delivery. Enterprises appreciate the benefits it brings – the agility, scalability, cost efficiency and more. This current approach, however, needs to evolve in the long run. The rapid advance in technologies such as the Industrial Internet of Things (IIoT), big data analytics, machine learning, artificial intelligence (AI) and Computer Vision requires an alternative...

Why?

Because ingesting massive data sets from geographically distributed edge devices and processing it in the cloud generates critical issues. Today's networks are not yet ready for the incredible growth expected in data to be transferred. Potential risks include low latency, bandwidth congestion, poor scalability, not to mention privacy, sovereignty and security issues.



Lessons learned from our experience with Mindsphere On-premise/ Edge Data Centers

Atos Provides **MindSphere On-premise/Edge Data Centers** solutions to many manufacturing customers so they can be able to securely deploy and manage their own MindSphere business-driven applications on shop floor. This Solution leads to a number of essential **benefits like data confidentiality, low-latency and data processing in close proximity to technical equipment at shop floor**. As a strategic partner to Siemens, Atos has been one of the first partners to enter the MindSphere ecosystem by developing an approach to Manufacturing IoT exploitation which balances innovation and compliance in equal measure and which delivers quantifiable benefit rapidly and reliably. In particular Atos offers MindSphere customers a structured approach; from quick-start services with first real results to production readiness of MindSphere applications, in order to accelerate the prototyping, validation and deployment of MindSphere applications through automated and predefined use cases. By offering Mindsphere On-premise/Edge Data Centers services, Atos manages as well all aspects of cloud infrastructure and delivery, scaling up as business needs change and ensuring performance, security and clarity of reporting at all times.

With crucial production environments needing real-time (or near real-time) reactions in closed-loop environments, there is a pressing need to transfer real-time data processing and analysis nearer to the source of data. Compute capability needs to be provided inside an environment where connectivity and response times can be tightly controlled. Edge computing provides a perfect response to these high-stakes challenges.



Lessons learned from a large manufacturer

Our customer is an innovator in health and creator of the world's first apple cider vinegar (ACV) gummy. Our customer is experiencing exponential growth. They are very limited with manufacturing at the moment because they only have two manufacturing lines and cannot keep up with the demand of their product nor scale their business. Because of this, down time on the manufacturing lines is extremely critical. Our customer is in the process of scaling their manufacturing lines from 2 to 7 lines and are getting ahead on the predictive maintenance of the lines. They want to **anticipate machine failure, monitor asset behavior and analyze trends** with robust reporting tools in their factories. Atos is providing Predictive Maintenance-as-a-Service by implementing data processing and data science software that leverages **compute at the edge** to produce predictive insights to maintenance and operations centers. **"Bull Sequana edge servers" and "Atos Codex Datalake Engine" are providing real-time data science predictions and data ingestion to anticipate machine failure, monitor asset behavior and analyze trends with robust reporting tools.**

The exponential growth of intelligent sensors and devices is generating an unprecedented amount of data. This is reshaping IT architectures, as increasingly powerful processing and machine learning inference capabilities are required at the edge of the networks to enable next generation, transformative AI and IoT applications.

5 criteria to move towards the Edge computing

Atos considers criteria essential when considering the evolution towards edge



Real-time

Addresses IoT latency issues, enabling a near real-time response by bringing computation close to data sources.



Video analytics

Capacity to analyze massive and complex video data in real time.



Security and privacy

Data in motion and data at rest as well as the physical server are protected by an advanced chain of security measures.



Cost reduction

Massive and complex data induct high network costs and cloud provider costs. BullSequana Edge can be completely used independently.



Local autonomy

It can communicate through radio, private LTE or Wi-Fi networks. Can be mounted in a 2U form factor rack. Reduced dependence on cloud and data center availability and connectivity



Lessons Learned about Inference* costs at the Edge vs Cloud

Use case 1: Apply inference model for 1 image per second
1 Frame per second @30 days = 2.6 million images inferred against ONE model.
3Y TCO for 1 inference model - Edge Computing vs cloud API

Inference using AWS REKOGNITION	Inference at the EDGE
<ul style="list-style-type: none"> Provides object detection within video images for \$1 per 1000 images up to 1M and \$0.80 per 1000 after that. 2.6 million inferences per month = \$2280 per model Three-year total cost= \$82,080 per model 	<ul style="list-style-type: none"> 1 BullSequana Edge Server Three-year total cost= \$20,000 per model

To illustrate benefits of Edge computing for AI inference, the following cases depicts the pricing comparison between "Inference on the Edge" and "inference on AWS cloud using Recognition API" - as of April 2020.







Use case 2: Apply 10 inference models for 1 image per second
 3Y TCO for 10 inference models - Edge Computing vs cloud API

Inference using AWS REKOGNITION	Inference at the EDGE
<ul style="list-style-type: none"> Three-year total cost= \$800,000 	<ul style="list-style-type: none"> 3 BullSequana Edge Server Three-year total cost= \$60,000

* Inference: Inference refers to the process of using a trained machine learning algorithm to make a prediction.

3 solutions to move towards the Edge computing

At Atos we believe that there are 3 solutions for infinity applications to solve your industrial challenges.

Edge Computer Vision	Edge Data Analytics	Edge Datacenter
		
<ul style="list-style-type: none"> Realtime Video Analytics Complex AI Inference at the Edge Video Analytics features 	<ul style="list-style-type: none"> Data collection and management Real time Data Analytics Scalable to absorb Data growth 	<ul style="list-style-type: none"> Complete Infrastructure at the Edge Secure and Rationalize IT outside Datacenters Large data storage anywhere
Key Drivers <ul style="list-style-type: none"> Video AI Computer Vision 	Key Drivers <ul style="list-style-type: none"> Industrial lot OT Sensors 	Key Drivers <ul style="list-style-type: none"> Micro Data Centers Edge Containers Real time Business applications
Atos Edge Server 	Atos Edge Server 	DC in a rack or Safe rack 
Industry protected wall mount stainless steel solution 	DC In a Rack embedding Datalake & Data Analytics 	EDC in a Container 

Usage & considerations for those 3 solutions is described in the next chapters:





- Artificial Intelligence at the Edge” for Computer vision & Data Analytics solutions
- Edge Data Centers” for Edge Data centers solutions





Artificial intelligence at the edge

Artificial Intelligence (AI) is increasingly finding its place in the manufacturing sector. As the technology matures and costs drop, Atos identified use cases and applications for which AI algorithms can make complex decisions showcasing better capabilities predicting what will be needed and when. Computer vision is one of these applications.

This technology deals with how computers can gain high-level understanding from digital images or video. Quality control, safety of production sites, automation of product lines, or management of personnel get the full benefits of this technology. It requires high level of computing power provided by CPU and GPUs and low latency for real-time analysis. Data Analysis is another key application. This becomes increasingly important in predictive maintenance for equipment and production lines, with sensors tracking operating conditions and performance, learning to predict issues and malfunctions, and taking (or recommending) preemptive actions.

The table below summarizes many of the key use cases for manufacturing, best fit in an Edge context:

Category	Use case	Business Benefits & costs	 Real-Time	 Video-analytics	 Local Autonomy	 Security & Privacy
Quality Control	<ul style="list-style-type: none"> Quality Inspection 3D Vision Inspection & 3D in-line metrology: Machin e vision inspection systems where defects get classified according to their type and are assigned an accompanying grade or default Package Inspection Count items before placing them into boxes/containers, check for broken or partially formed packages (right color, length, width, and whole) so defective items/box/containers are then rejected 	<ul style="list-style-type: none"> Increase First Pass Yield Decrease Rework, scrap, recall Reduce cost of control Improve On-Time Delivery 	High	High	High	Medium
Safety of production sites	<ul style="list-style-type: none"> Detection of an operator in poor posture Detection of operators that are in a hazardous and life-threatening situation Detection of the absence of Personal Protective Equipment (PPE) or incorrectly worn PPE Detection of abnormal situation (People lying on the ground, unauthorized areas) Detection of dangerous driving situations with forklifts, trucks... 	<ul style="list-style-type: none"> Improve workers Safety Decrease number of reported accidents & incidents Improve productivity and decrease monthly health and safety prevention costs 	High	High	High	Medium
Automation of product lines	<ul style="list-style-type: none"> Monitoring manual operations with smart cameras Forklifts and autonomous factory equipment (collision detection, pallet management, vision system for autonomous factory equipment) Track and Trace Reading, identifying and processing barcodes at scale 	<ul style="list-style-type: none"> Improve security & Safety of personnel Increase throughput of the production line / Productivity Improve Overall Operating Efficiency, Overall Equipment Effectiveness 	High	High	High	Medium

Category	Use case	Business Benefits & costs	 Real-Time	 Video-analytics	 Local Autonomy	 Security & Privacy
Predictive maintenance	<ul style="list-style-type: none"> Collect & Analyze data flows & images from cameras attached to robots/industrial machines to identify potential problems before they arise Digital Twins Execute a digital copy of the physical system to perform real-time optimization and run Machine Learning algorithms 	<ul style="list-style-type: none"> Increase throughput of the production line / Productivity Improve Overall Operating Efficiency, Overall Equipment Effectiveness Decrease Machine Downtime, Unscheduled Down Time 	High (lot)	Medium	High (lot)	Medium
Parking Management	<ul style="list-style-type: none"> Parking Capacity & utilization Optimization Enforce regulations rules and detect violations (SEVESO sites) 	<ul style="list-style-type: none"> Detection of number of available places in real time Detection of violations (unauthorized vehicles, defensive driving rules, wrong ways) Cost control by reusing existing video systems in place 	High	High	Medium	High
Outdoor smoke & fire detection	<ul style="list-style-type: none"> Video detection of smoke and fire outdoor 	<ul style="list-style-type: none"> Reduce intervention time Cost control by reusing existing video systems in place 	High	High	High	High

Edge Data Centers

Edge Data Centers (EDC) refer to data centers which are positioned closer to end-users and source of the data, delivering faster services with minimal latency.

They help in enhancing the efficiency of content delivery and bringing cloud computing resources and cached streaming content to local end-users. These data centers are normally smaller facilities that extend edge of a network and deliver cached streaming content to local end-users. They control data processing at the edge of a network compared to keeping in cloud or a centralized warehouse. These facilities are mainly used to support applications that demand a high amount of bandwidth, are latency-sensitive, and require rapid response times. Edge Data Centers range from small clusters of "Edge Cloud" resources located near a cobot for example, to a few racks located in a shelter at the end of a production line, inside buildings or in a warehouse, to a self-contained ready-to-drop and plug-in container in a temporary facility. The following needs to be taken into consideration when evaluating the opportunity of EDC in a plant are the following :

Telecom and Computation

For applications demanding high computing and/or storage, costs related to telecommunication and cloud computation can significantly increase. Those costs can be reduced by pre-processing data at the edge and transferring only identified valued data to the cloud (Incident detected, Deviation from set-value, pre-processed reports..).

*Example: Video Analytics for quality control in syrup bottling or access control to labs
Predictive Maintenance on complex machines that generate high volume of IIoT data.*

Latency & real time

Due to the distance between AWS Datacenters (located in / out of the country) and the factory in a pure Cloud architecture, latency may become an issue when real time decision is necessary.

Example of a sensor monitoring the temperature curve for the creation of medicine, EDCs can be the answer to have realtime analysis and make decision when deviations to standard model are identified.

OT security

Production plant are becoming more and more intelligent. With the generalization of IIoT, the number and variety of devices connected to the network generating data is increasing complexity for devices/data management. Then EDC can be an answer to manage this diversity/complexity, as well as security challenges with the potential creation of thousands of risk exposure points on the network.

Local Facilities

Depending on the use case to be addressed and type of EDC the topic « local facilities » needs to be taken into consideration. Uses cases like Video Analytics, or predictive maintenance can be addressed by standalone GPU accelerated edge servers that can be installed in any room, on a wall or a ceiling. More sophisticated services like local implementation of SAP based on dedicated appliance, can request « containerized EDC », specific shelve with dedicated cooling and USP.

Operational Risk

Another point that needs to be taken into consideration is the risk related to intermittent connectivity or communication breakdown between the AWS cloud and the end user. If this risk is limited in the context of a well-established plan, it may become important for a temporary facility.

Cloud providers recognize the interest of Edge Computing and offer Edge devices to "bring cloud capabilities for machine learning, data analytics, processing, and storage to your edge, for migrations, short-term data collection, or even long-term deployments. Edge devices work with or without the internet, do not require a dedicated IT operator, and are designed to be used in remote environments.** (*source: AWS-Snowball)



Lessons learned from a large Entertainment Park on Predictive maintenance

Lessons learned from a large Entertainment Park on Predictive maintenance

The world's largest amusement parks attract an average of more than 40,000 visitors per day. Today, about 11,000 sensors on each ride provide information on the condition of the engine, brakes and structure. When the roller-coaster breaks down, it must be immediately evacuated. The admission price is maintained regardless of the number of opened attractions, which impacts customer satisfaction.

Atos developed predictive maintenance solution hosted in Edge Micro data Centers combining:

- Data Analytics on IoT information, as signs of wear and tear, deformation or other damage to check the integrity of the equipment (mechanical parts, lock nuts, lock washers, safety pins, restraint and locking systems, controls, brakes, emergency stop devices and communication systems).
- Analysis of high-volume video data from cameras, scanning equipment 360 degrees and information from tunnel monitoring, producing 1 GB of data per minute.

Hosting Data capture and data processing at the Edge, made real time analysis possible, avoiding the transfer of huge volume of information on the Cloud.

Powerful deep learning algorithms identify the source and nature of an issue, triggers an alert and recommends a mitigating action to fix or tune the rollercoaster, avoiding ride shutdown. Atos Real-time decision solution speeds up the production process. Installed closed to the ride, BullSequana Edge allows to obtain information on a defective equipment 22x faster. Here, critical lead times on quality control chains are optimized 24/7.

Edge computing to move towards a low-carbon future

Edge as an opportunity

In the coming years, factories are likely to be subject to strict rules on CO2 reduction. In this context, they will have to work to reduce their energy footprint. This constraint is in fact an opportunity, since it has many advantages such as better communication to the general public, less energy expenditure, and a more sustainable business. To address this challenge, the use of Edge Computing can be effective in reducing the energy footprint of digital.

Save space in the factory & optimize energy consumption

A significant part of the energy of companies is due to the consumption of their digital ecosystem. Machine rooms often take up a large amount of space, in which IT-related electricity consumption does not account for the entire energy balance of the room. For example, data centers use additional power sources such as an external cooling system. Several studies have been carried out on the PUE of company-run data centers. One of the most complete is conducted every year by the "Uptime Institute", which surveys thousands of companies worldwide. In 2019 the average PUE was around 1.67 (Uptime Institute's Data Center Survey, 2019) which means that for every 100w of IT, companies used 67w of auxiliary power such as air conditioning.

Save energy costs

Less energy consumption means less energy-related expenses. The electricity bill, depending on the country in which the energy is consumed, can be quite varied. Nevertheless, over several years, the amount can turn out to be high. In this context, the use of Edge Computing seems to be relevant and can save a significant part of the consumption and thus reduce the bill.

Thanks to its independence from these other external systems, the Edge Servers can solve this problem: Edge Servers are coming with relaxed operating temperature specifications (+5°C to 45°C), which means that the operator does not need to deploy an external air-conditioning solution.

BullSequana Edge Servers allows maximum inference capability in a minimum power consumption

Reduce the pressure on network bandwidth

In a digital ecosystem, it is relatively difficult to accurately measure the amount of energy used by the network. Nevertheless, a few serious studies have been conducted to determine an order of magnitude*. This explains that the transfer of 1GB of data through the mobile network would consume about 0.8KWH of energy. This figure is not an absolute truth, can vary depending on many parameters, but can give an idea of what the network consumes. Unlike a remote private cloud system, the Edge can be installed on-site, directly in the same network. This means that the pressure on bandwidth will be lower. It will be possible to conserve bandwidth for other uses, or even not use it at all and reduce the energy consumption of the network infrastructure

The Edge combines a large IT capacity in a minimum volume, with a power usage that will not exceed 380w. In this way, the BullSequana Edge offers high computing power with minimum energy consumption. The inference capacity of the machine, thanks to two AI acceleration capabilities provided by Nvidia T4s and its low power supply, offers a power density between 0.4 to 0.6 mwh/Top (Top is TeraOperations), in contrast to DELL's PowerEdge XR2 range of 0.8 to 0.9 mwh/Top. In addition, the power density of the Edge allows the operator to deploy multiple applications in the same machine. Parallelization solutions such as containerization can thus save a large number of machines that were previously used for each application at low capacity.

Conclusion

Edge computing and Artificial Intelligence are completely transforming manufacturing to drive better efficiency and productivity while also controlling costs.

Atos is uniquely positioned to support Sanofi in the post-cloud era, helping Sanofi understand, deploy and fully leverage edge computing.

Atos delivers a complete edge computing solution, including hardware, software, professional services and ready-to-go use cases. This set of expertise is a prerequisite to consolidate data in a data lake engine while training new analytical models, developing inference models and analyzing real-time data streams - using an edge server in a micro data center at the edge.

About Atos

Atos is a global leader in digital transformation with 120,000 employees in 73 countries and annual revenue of € 13 billion.

European number one in Cloud, Cybersecurity and High-Performance Computing, the Group provides end-to-end Orchestrated Hybrid Cloud, Big Data, Business Applications and Digital Workplace solutions through its Digital Transformation Factory, as well as transactional services through Worldline, the European leader in the payment industry. With its cutting-edge technologies and industry knowledge, Atos supports the digital transformation of its clients across all business sectors. The Group is the Worldwide Information Technology Partner for the Olympic & Paralympic Games and operates under the brands Atos, Atos Syntel, Unify and Worldline. Atos is listed on the CAC40 Paris stock index.

Find out more about us
atos.net/bullsequanaedge

Let's start a discussion together



Atos, the Atos logo, Atos Syntel, Unify, and Worldline are registered trademarks of the Atos group. January 2019. © 2019 Atos. Confidential information owned by Atos, to be used by the recipient only. This document, or any part of it, may not be reproduced, copied, circulated and/or distributed nor quoted without prior written approval from Atos.