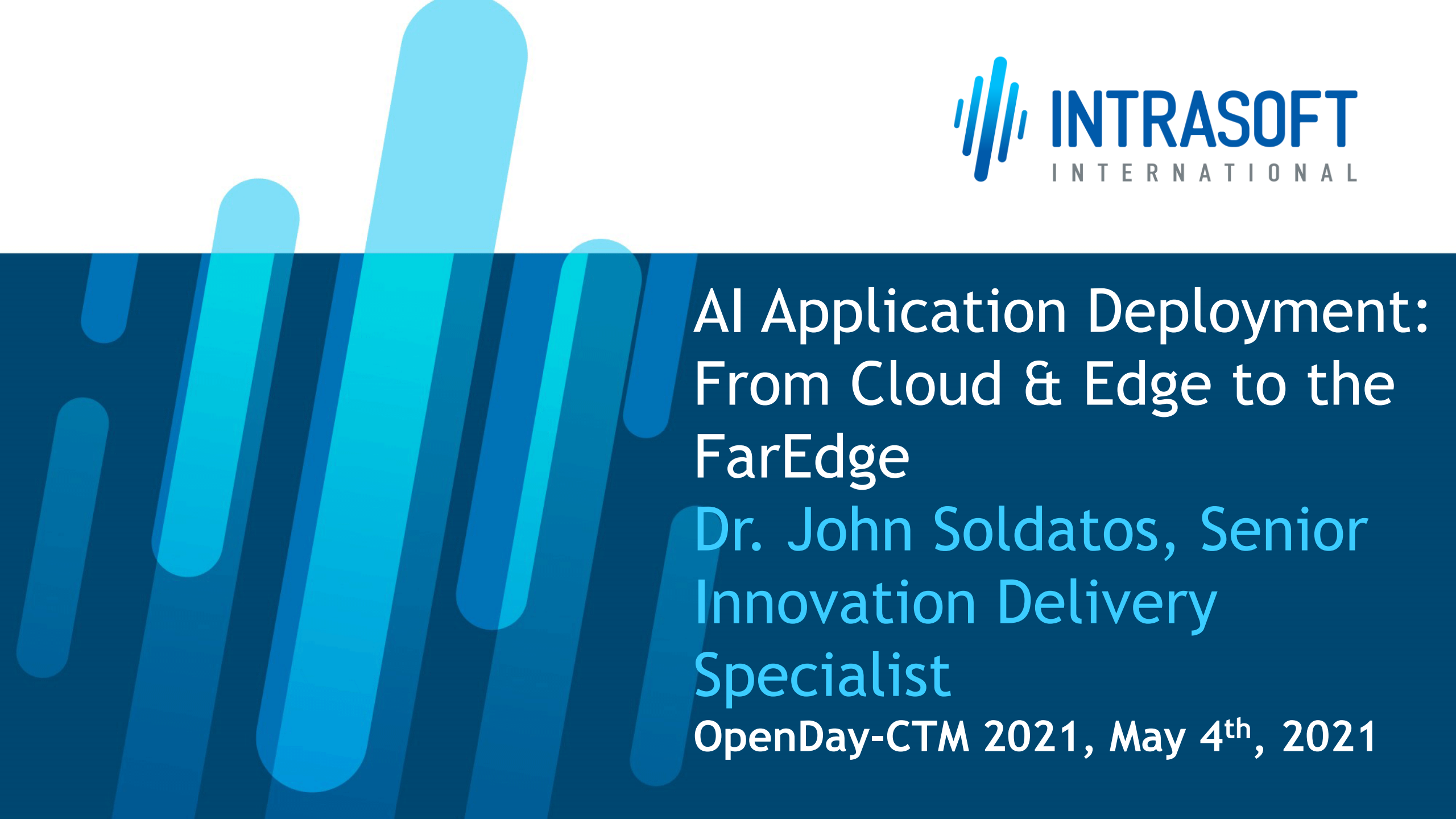


AI Application Deployment: From Cloud & Edge to the FarEdge

Dr. John Soldatos, Senior Innovation Delivery Specialist

OpenDay-CTM 2021, May 4th, 2021

INTRASOFT in a nutshell

INTRASOFT INTERNATIONAL

A leading European IT Solutions & Services company

Established in 1996 in Luxembourg, operations in 13 countries

A key player in E.U. Institutions and Agencies for 20 years

Outstanding record in over 70 countries around the globe

Providing Configurable & bespoke applications - Managed services

Subject Matter Expertise

- Customs
- Taxation
- Social security
- Telco
- Finance

Innovation

- Analytics
- Machine learning
- Blockchain
- Cloud

Talent

- Attract the best
- Continuous development
- Culture of leadership & ownership

Methodology

- Tailored Agile & DEVOPS
- ITIL
- Security

Research & Innovation Development Unit

Over 20 years of successful R&D projects implementation *and*…

- System Integration & Architecture Specification
- Technical Coordination
- Application **Development**
- **Communication**

Efficient/effective **coordination** and team-leading EC co-funded R&D practices:

- FP5/FP6/FP7/H2020
- CIP/ECSEL/CEF
- Studies/Impact assessment/Policy support

Pre-commercialisation of Information and Communication Technologies **research results**

…innovation

Innovation-related activities:

- Internal Innovation Competition
- **Innovation radar**/Technology transfer
- Joint ventures

Strategy positioning:

- Feasibility study of cutting-edge technologies (Big Data, Data Analytics, Blockchain, …)
- Opening up new innovative fields (agrifood, AI, maritime security, …)

Product Development and cross – collaboration:

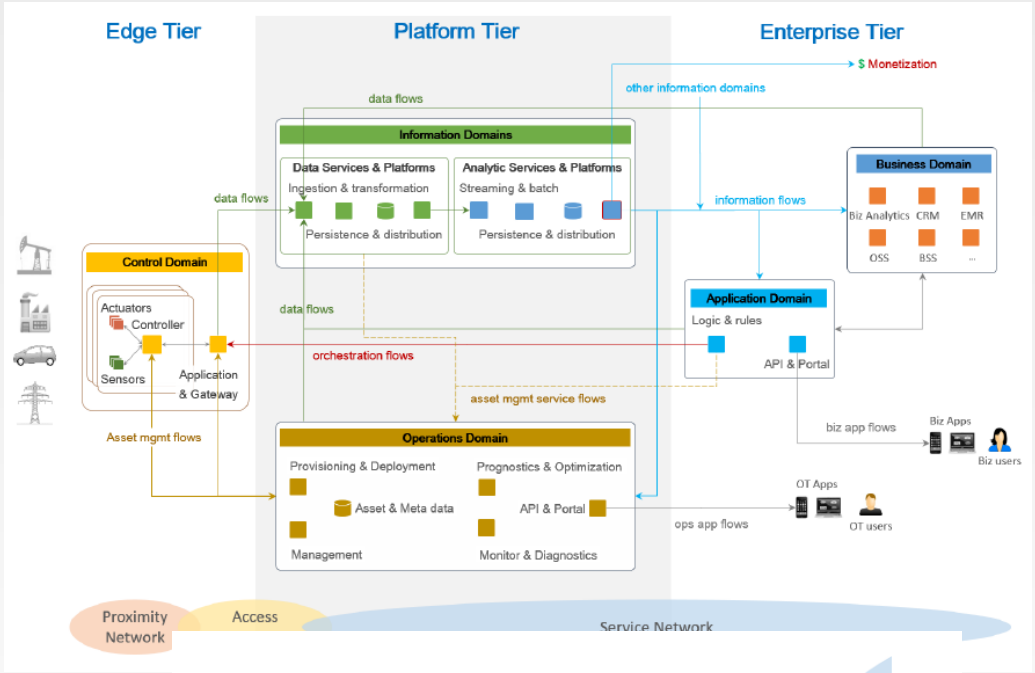
- Extend/New **product offerings**
- Cross-departmental business and commercial collaboration
- **Spin-off** (SmartPACT – blockchain in insurance)

Edge Computing for Industrial Use Cases: When Cloud is Not Enough

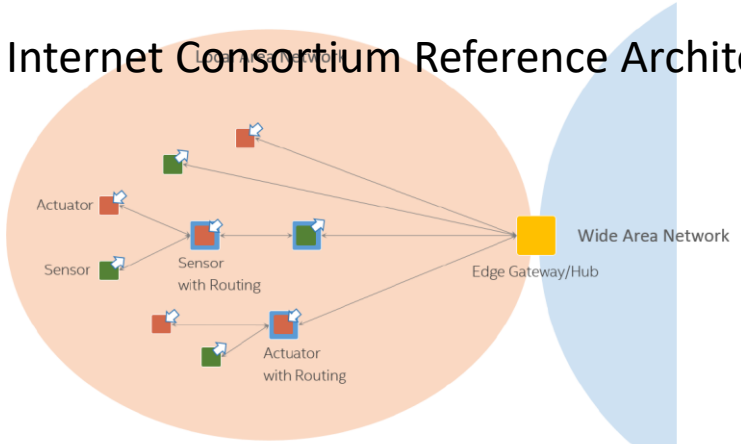
INTRASOFT INTERNATIONAL

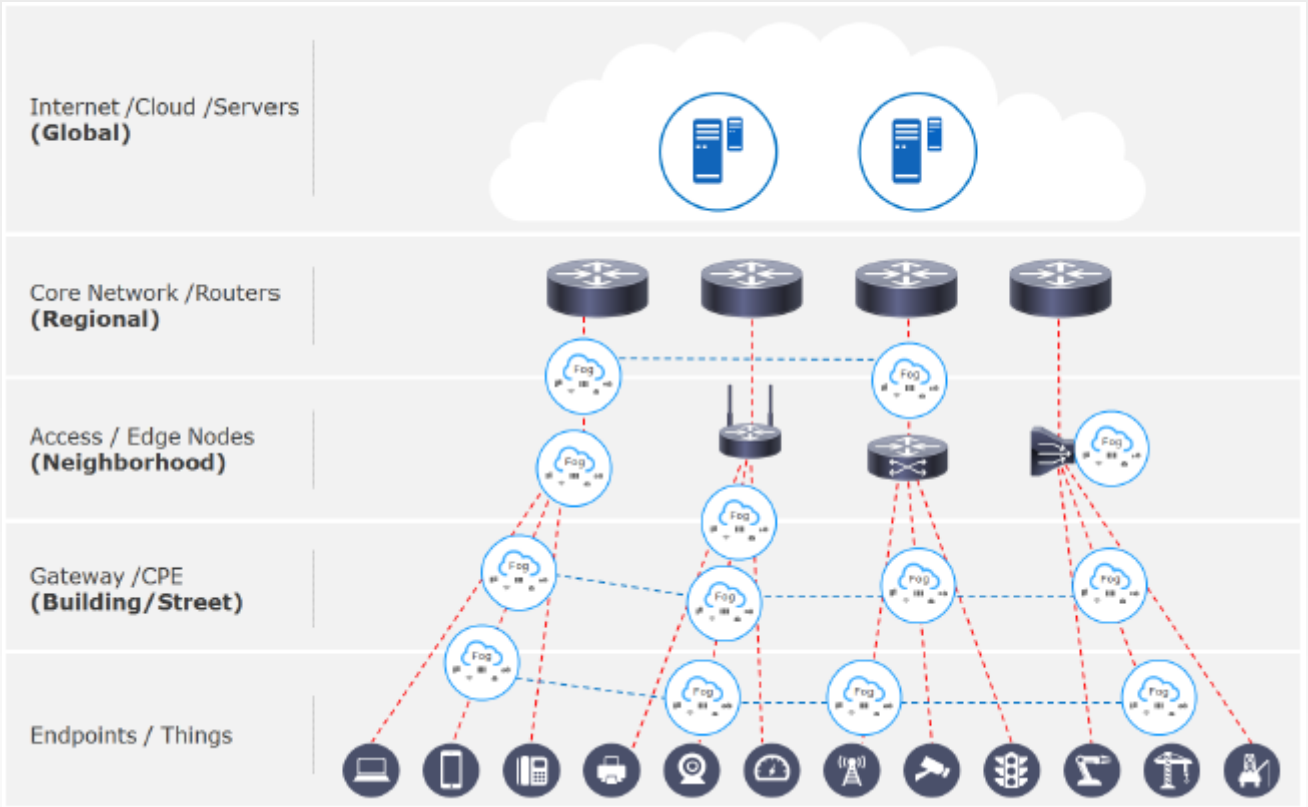
- Rationale for Edge Computing in Industrial Use Cases:
  - Latency & Real-Time Field Operations (e.g., Real-Time Actuation and Control)
  - Energy Efficiency (e.g., Sustainable Manufacturing)
  - Data Privacy and Protection (e.g., Industrial Data)
- Different Edge Computing Approaches:
  - Edge Computing with Local Clouds
  - Fog Computing
  - FarEdge Approaches (e.g., Embedded Machine Learning)
  - Federated Machine Learning
  - Decentralization based on Distributed Ledger Technologies

Edge Computing VS. Standard Industrial Architectures (source: https://www.iiconsortium.org/)



Industrial Internet Consortium Reference Architecture





OpenFog Reference Architecture

Cloud vs. Edge vs. FarEdge

FEATURES	CLOUD (e.g., IOT/Cloud Integration (WAN))	EDGE (e.g., Local Clouds, Fog Computing)	FarEdge (e.g.,AIoT, Embedded Devices, Microcontollers)
Data Points Availability	High	Medium	Low
Energy Efficiency	Low	Medium-to-High	Very High
Privacy	Low-Medium	Medium-to-High	High
Real-Time Opportunities	Low	Medium-to-High	Very High

Cloud vs. Edge vs. FarEdge

INTRASOFT INTERNATIONAL

Transfer Learning

INDUSTRY USE CASE	CLOUD	EDGE	FarEdge
Asset Management & Predictive Maintenance	RUL Calculation at Cloud	RUL Calculation at Edge Cluster or Edge Devices	Fault Detection Inside the Machinery
Quality Management & Zero Defect Manufacturing (ZDM)	Quality Inspection – Defect Prediction	Defect Prediction & Near Real-Time Control	Defect Detection & Real-Time Control
Value Proposition	ML Model Accuracy	Speed vs. Accuracy Balance	Speed and Power Efficiency

Predictive, Preventive, and Reactive Strategies to Asset Management & Quality Management

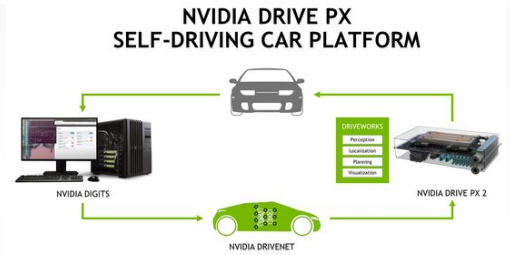
TinyML Background & Motivation

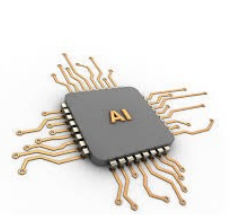
- Is AI/ML in the Cloud Always the Best Option?
  - Siri, Alexa, OK Google etc.: benefit from an instant response
  - Industrial maintenance: More timely detection of failures and abnormalies
  - Agriculture: Instant disease detection using a plant's image
  - In several cases ML/AI «at the edge» dramatically reduces costs & complexity, and limits potential data privacy leaks.
- TinyML: An Alternative Form of Machine Learning and AI at the Far Edge

Cloud AI (32 GB, TFLOPS/s)

Smart Phone AI (4GB, GFLOPs/s)

TinyML (<500KB, MFLOPs/s)

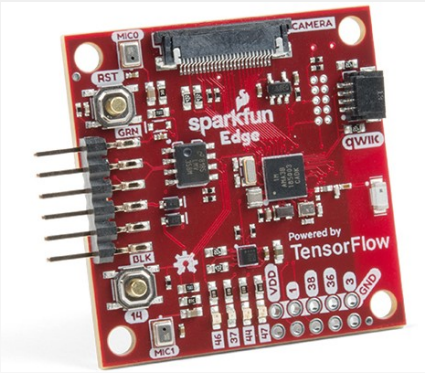




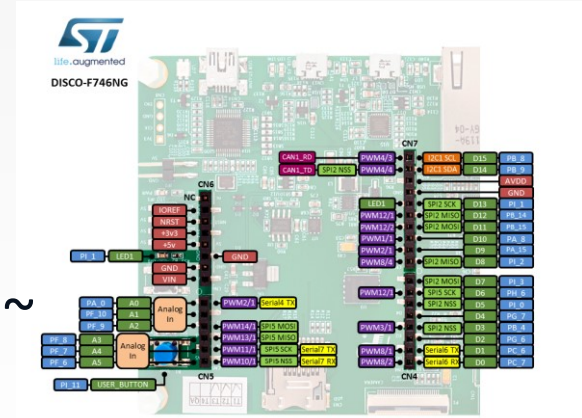
TinyML Hardware Examples

- ARDUINO NANO 33 BLE SENSE WITH HEADERS, Cost ~ 30€

- SparkFun Edge Development Board - Apollo3 Blue DEV-15170, Cost ~ 15€



- STM32F746G-DISCO discovery board (32F746GDISCOVERY) - complete platform for STMicroelectronics ARM® Cortex®-M7 core-based STM32F746NGH6 microcontroller, Cost ~ 70€

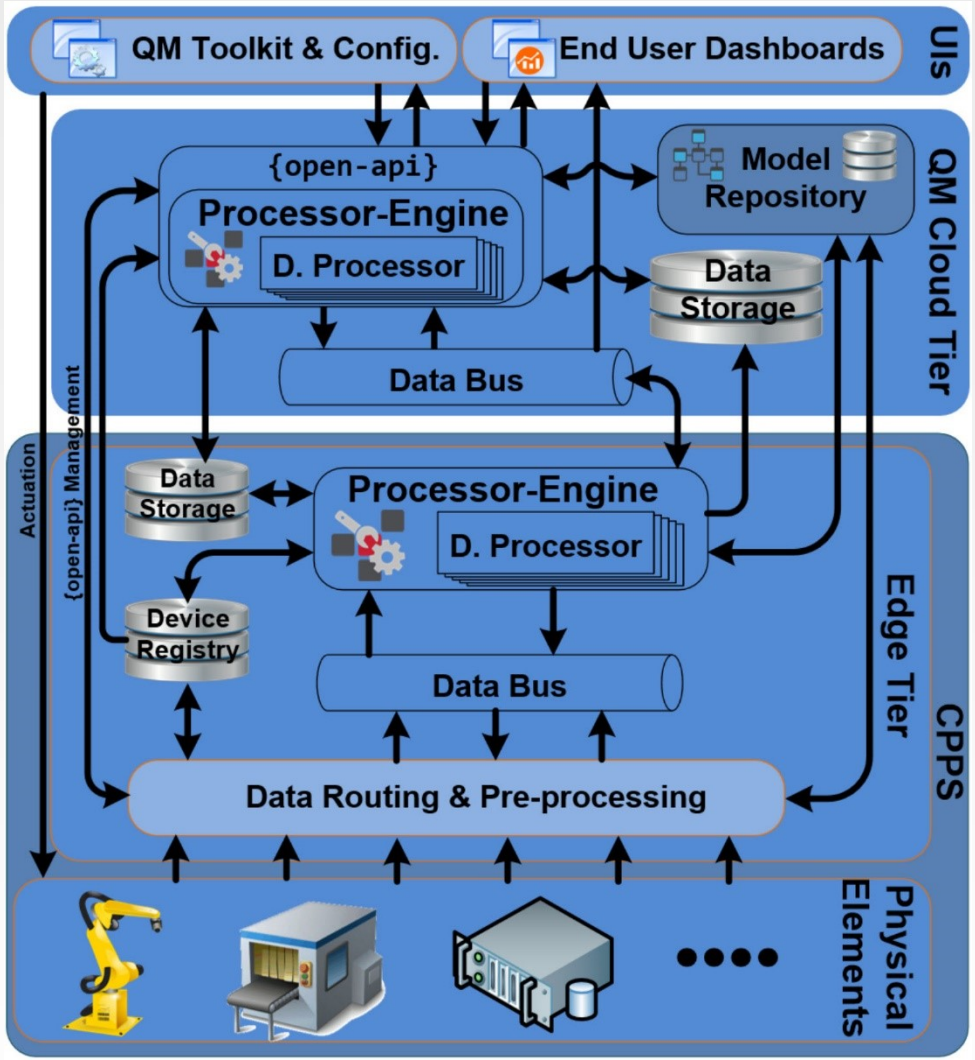


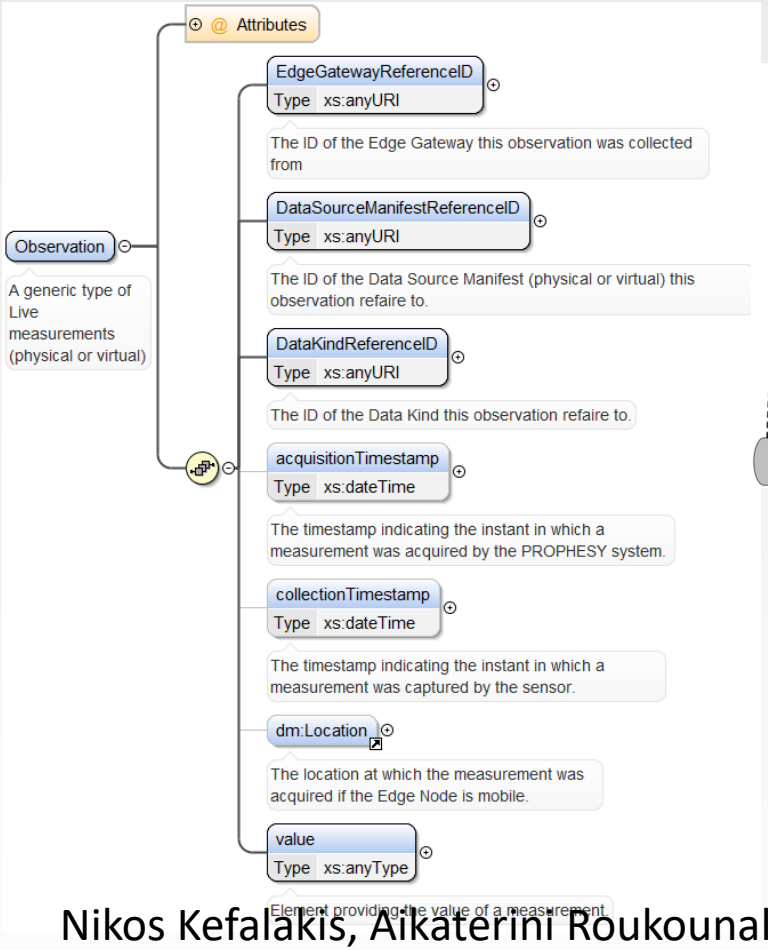
Building TinyML Models and Applications

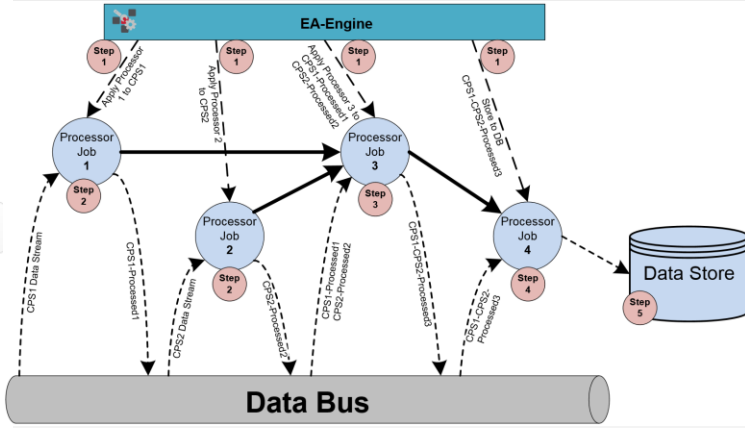
INTRASOFT INTERNATIONAL

- Standard Tools Work for Data Science & Model Development:
  - Python, Jupyter Notebooks, Arduino IDE,…
- Machine Learning Framework for TinyML:
  - **TensorFlow**: Suite of tools for building and running ML models
  - **Keras**: TensorFlow's high-level API focused on building and training Deep Learning applications
  - **TensorFlow Lite**: Specifically designed for inference on devices with limited computing capacity (e.g., phones, tablets, embedded devices).
  - **TensorFlow Lite Micro**: Deploy models on microcontrollers and other devices with only few kilobytes of memory e.g.,
    - Core runtime just fits in 16 KB on an Arm Cortex M3
    - Doesn't require operating system support, any standard C or C++ libraries, or dynamic memory allocation

INTRASOFT's DataCrop IIOT Platform







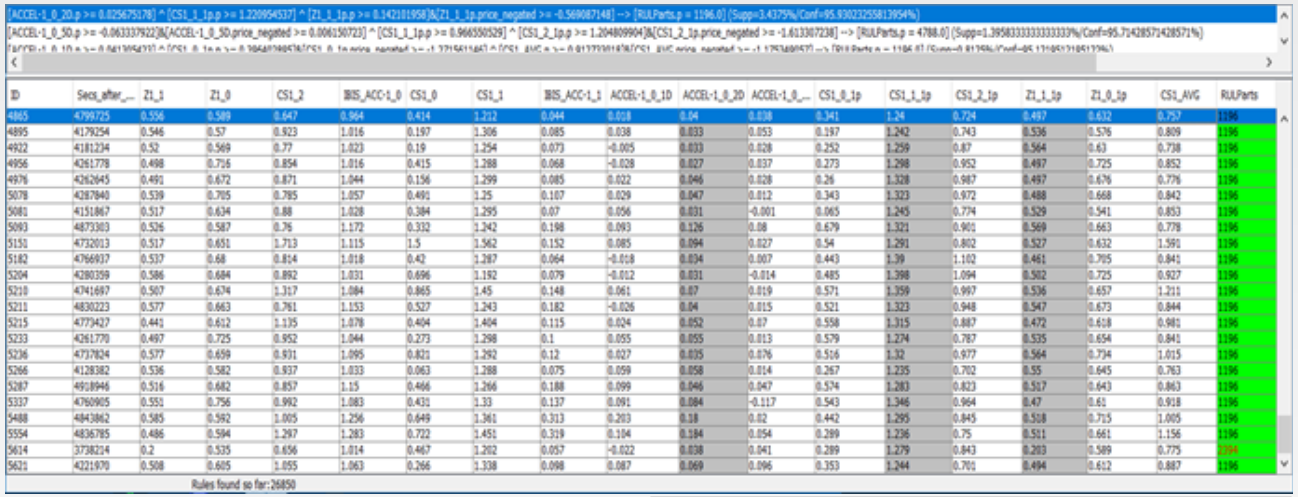
Processing Jobs are of three types:

- Pre-processing
- Storage
- Analytics (e.g., ML, calculation)

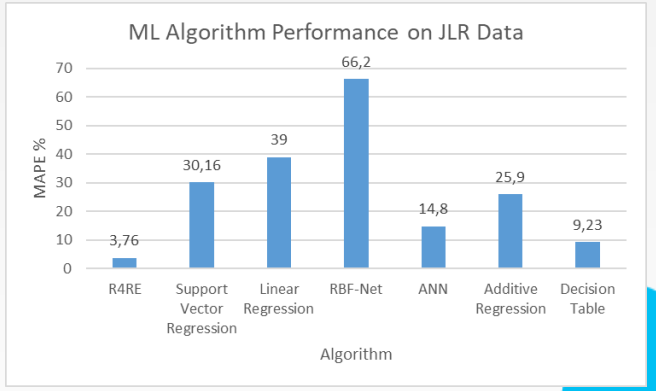
Nikos Kefalakis, Aikaterini Roukounaki, John Soldatos: Configurable Distributed Data Management for the Internet of the Things. Inf. 10(12): 360 (2019)

ML Framework Integration over DataCrop - QARMA

- Different ML Algorithms run over DataCrop
  - Recurrent Neural Networks (RNN)
  - Long Short-Term Memory Networks (LSTM)
  - Attention-based Networks
- QARMA4Industry:
  - Rules Mining Approach
  - Identify Rules that hold on data and quantify them
  - Used in H2020 QU4LITY Project



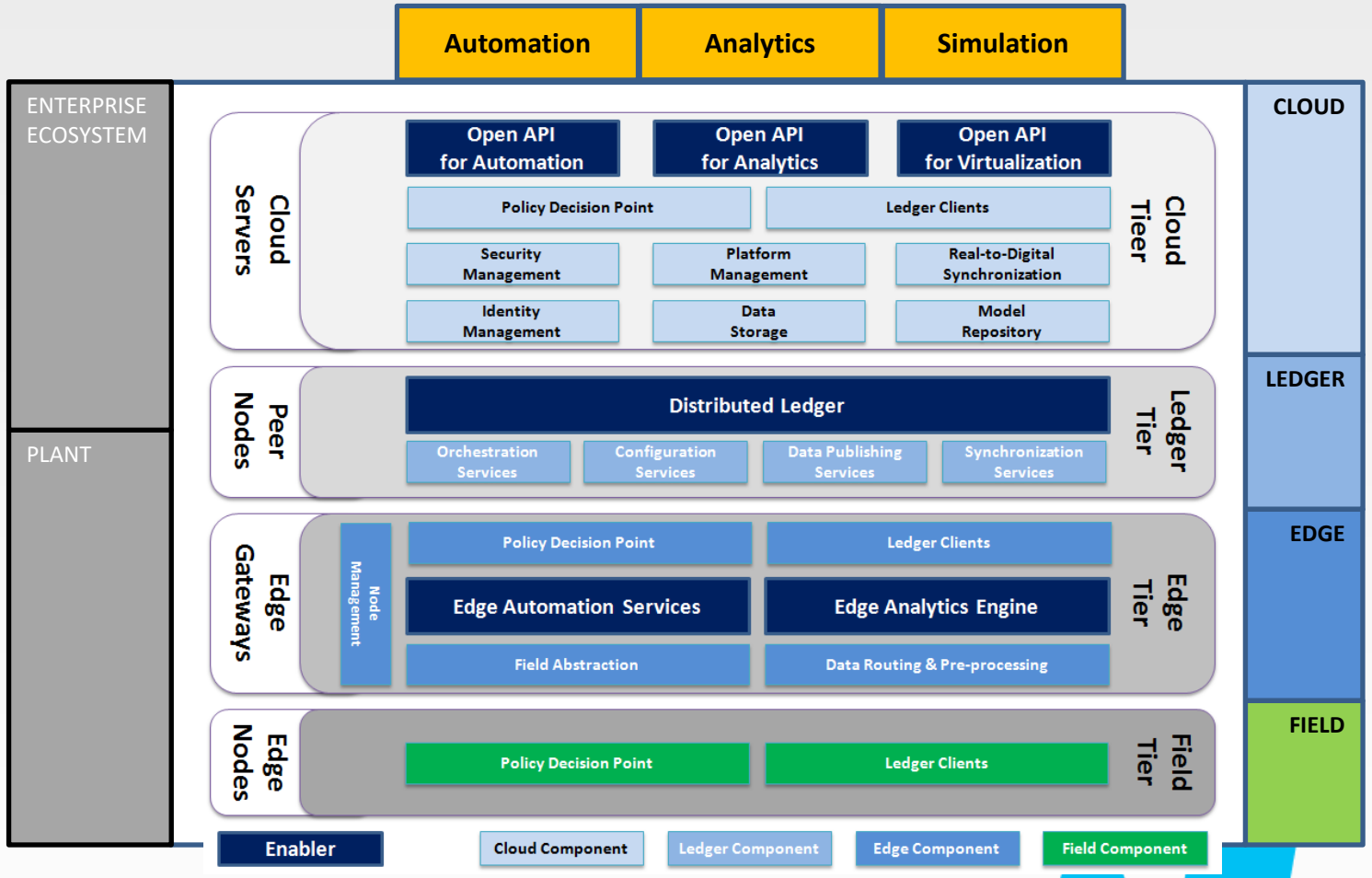
QARMA4Industry outperforms popular algorithms



Ioannis T. Christou, Nikos Kefalakis, Andreas Zalonis, John Soldatos, Raimund Bröchler, End-to-End Industrial IoT Platform for Actionable Predictive Maintenance, IFAC-PapersOnLine,
Volume 53, Issue 3, 2020, Pages 173-178, ISSN 2405-8963,

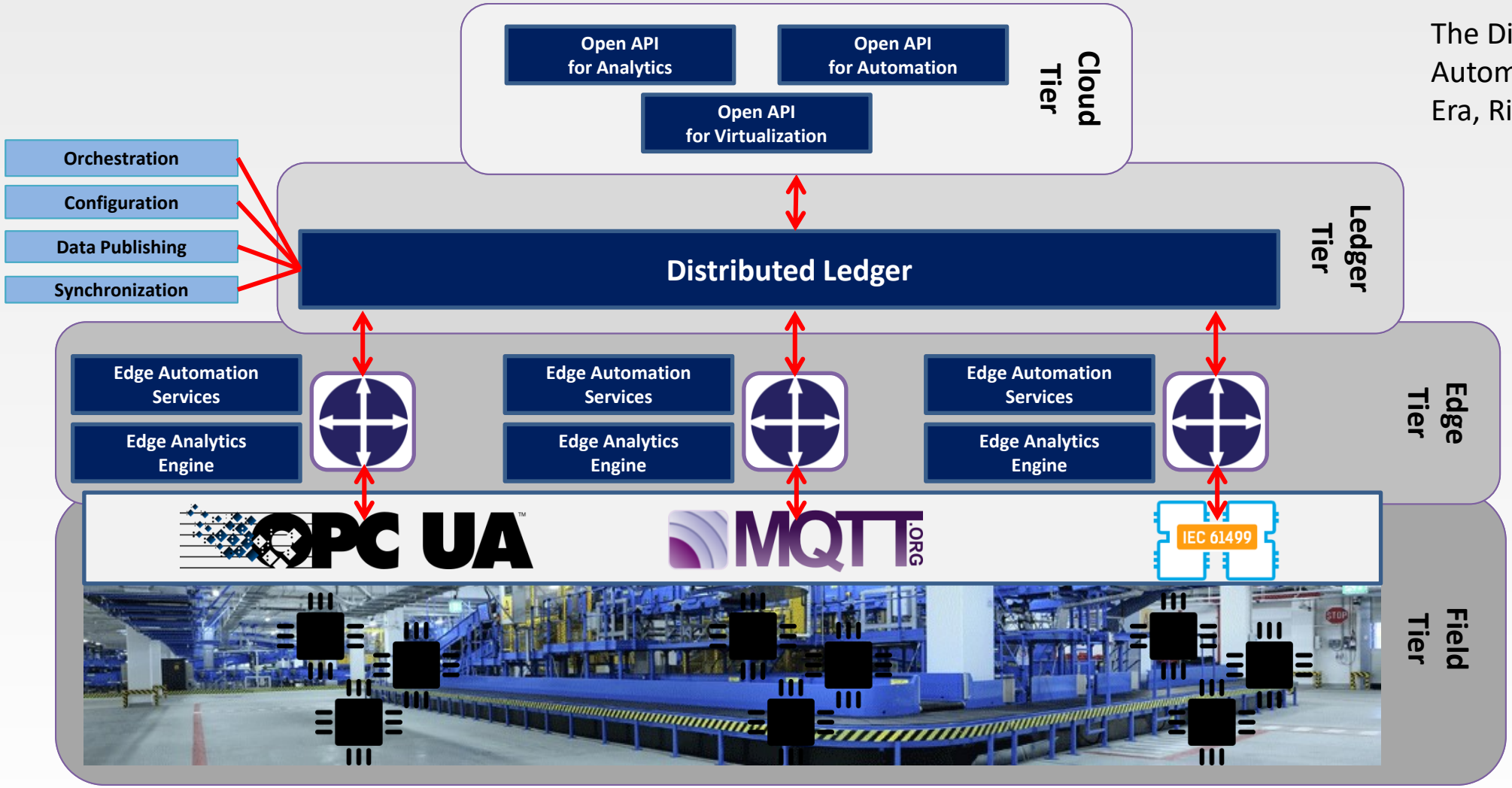
H2020 FAR-EDGE : Distributed Ledger Technologies for Orchestrating & Configuring Processes at the Edge

- Edge Computing Functions:
  - Edge Automation & Edge Analytics
- Edge Automation:
  - Low-overhead, low-latency connectivity
- Two way interactions
  - Edge Analytics:
  - Real-time analytics
  - One-way data savvy operations

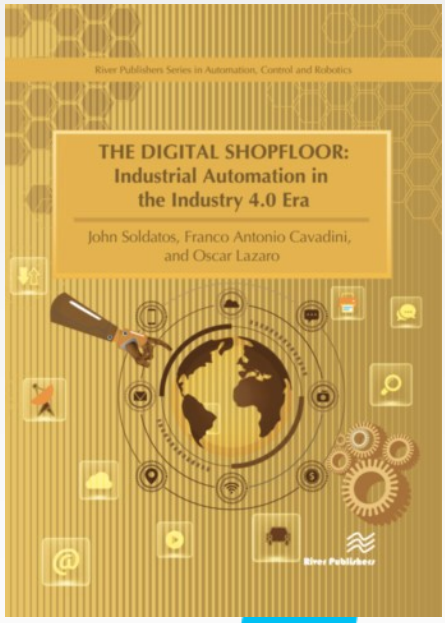


Mauro Isaja, John Soldatos:
Distributed ledger technology for decentralization of manufacturing processes. ICPS 2018: 696-701

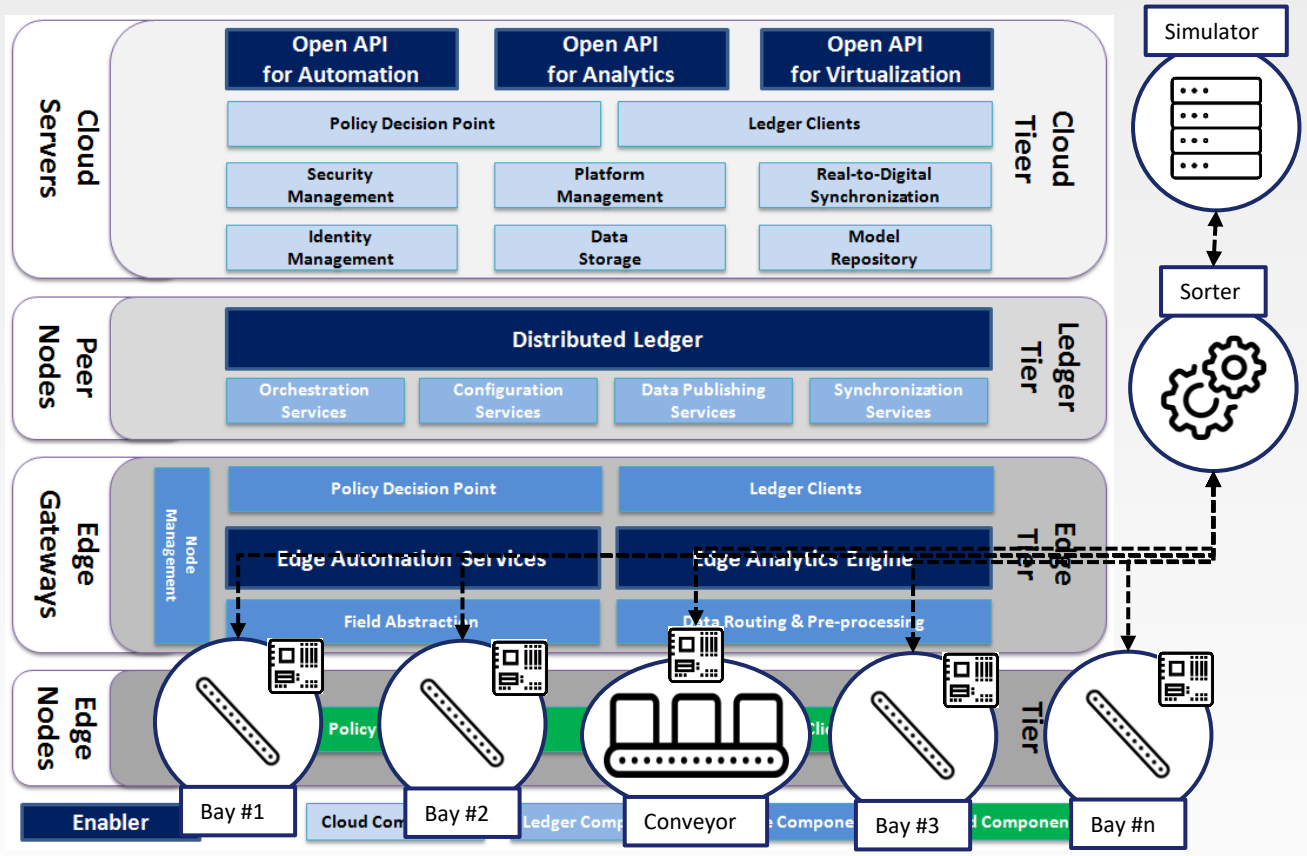
DLT Technologies for Edge Computing Functions (see: www.Edge4Industry.eu)

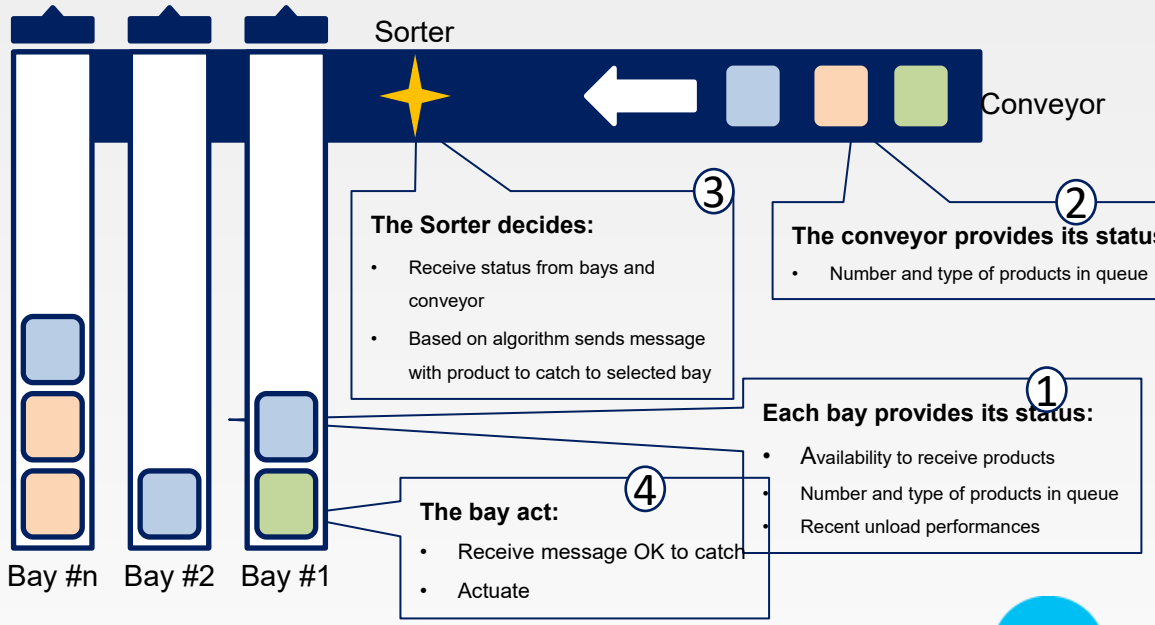


The Digital Shopfloor: Industrial Automation in the Industry 4.0 Era, RiverPublishers 2019

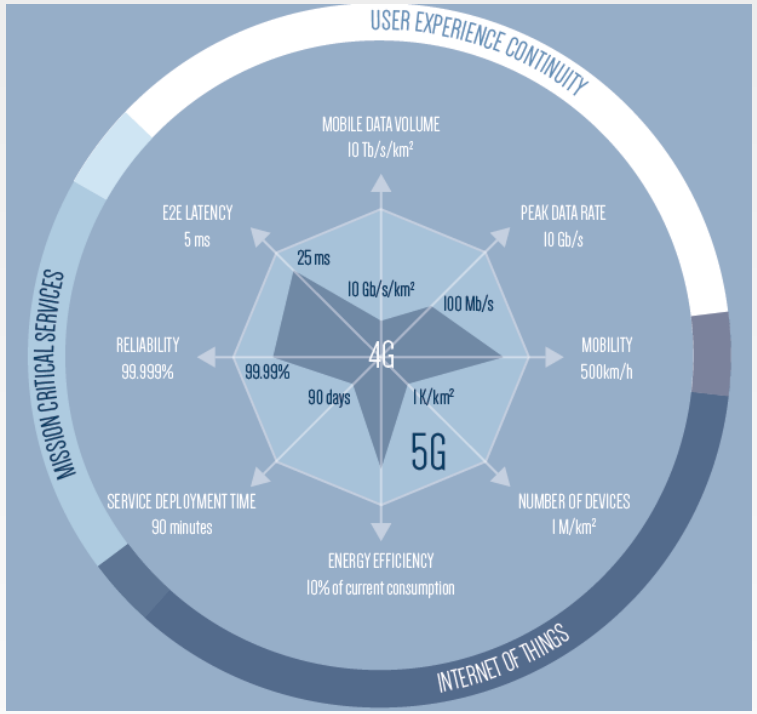


Edge Simulation Use Case (WhiteAppliances Factory) (source: FAR-EDGE Project)

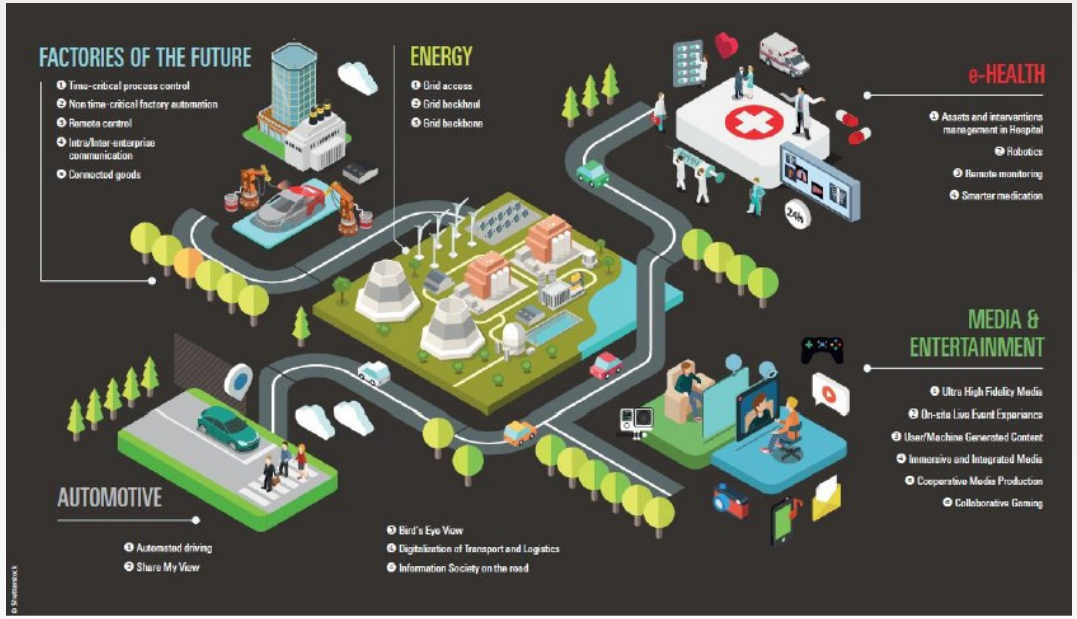




5G: New Capabilities & Services – Impact on Vertical Sectors



Source: *5G Infrastructure Association: Vision White Paper, February 2015*



Source: *5G Infrastructure Association: 5G Empowering vertical industries. White Paper, 2016*

Why 5G/6G for Edge Intelligence ?

INTRASOFT INTERNATIONAL

5G

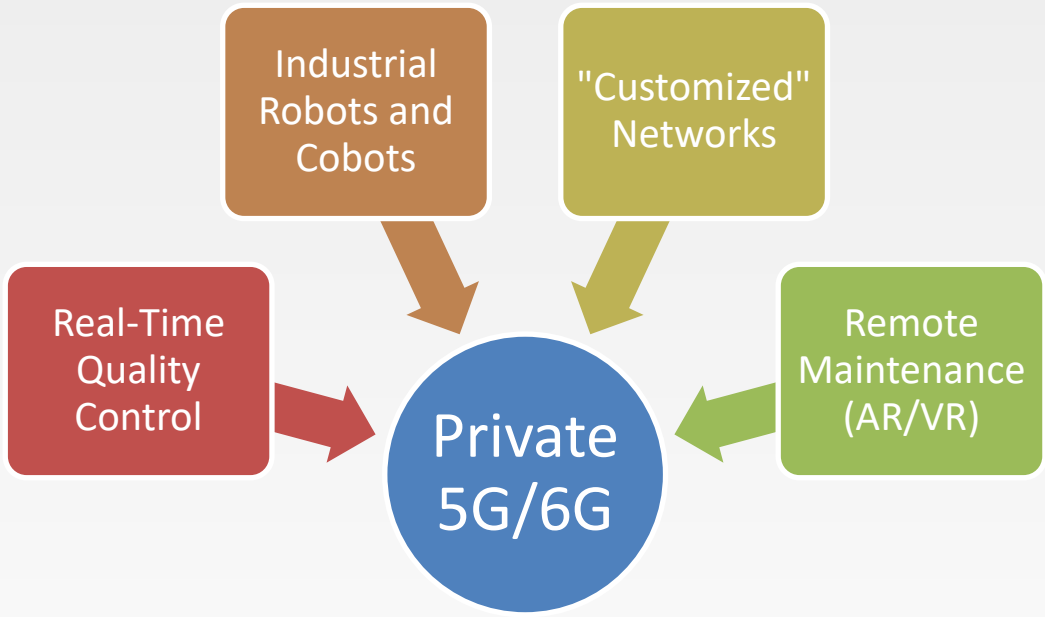
- Handle 1000X times more mobile data than today's cellular systems
- Increases data rate
- Reduce the end-to-end latency
- Improved coverage
- Flexibility through Slicing and NFVs

6G

- Extra speed and latency
- In-Network Analytics
- Meeting the stringent network demands of the next decade (2020-2030)

Enabling

- Seamless connectivity across multiple IoT technologies (e.g., RFID, short-range communications, UWB, Blue Tooth, etc.)
- Tactile Internet
- Edge AI



Source: ***Why 5G is the way to go for IoT: https://www.linkedin.com/pulse/why-5g-way-go-internet-things-john-soldatos/***

Conclusions

- Alternative Edge Computing Approaches in Industrial Environments
  - With 5G/6G as key Enabler
- No Silver Bullet: Selection Depends on Requirements:
  - Latency / Real-Time Concerns
  - Power Efficiency
  - Privacy
  - Decentralization
  - Security & Robustness
  - Federated Learning (not addressed in this presentation)
- An “Edge Operating System” should take into account relevant trade-offs
  - Zero Defects Manufacturing Approach Combines Reactive, Preventive and Predictive Strategies
- Infrastructures like GAIA-X can boost a modular approach

INTRASOFT INTERNATIONAL

Thank you for your attention

Follow us