How to Right-Size and Future-Proof a Container-First Edge AI Infrastructure

Carl Moberg
CTO at Avassa

Zoie Rittling
Business Development Manager at OnLogic

avassa ONLOGIC

Edge AI: There's more to it than the model and the application

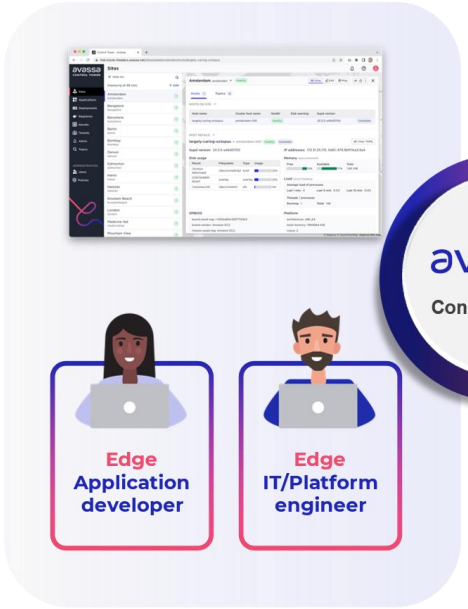
Edge AI starts with a model and an application…

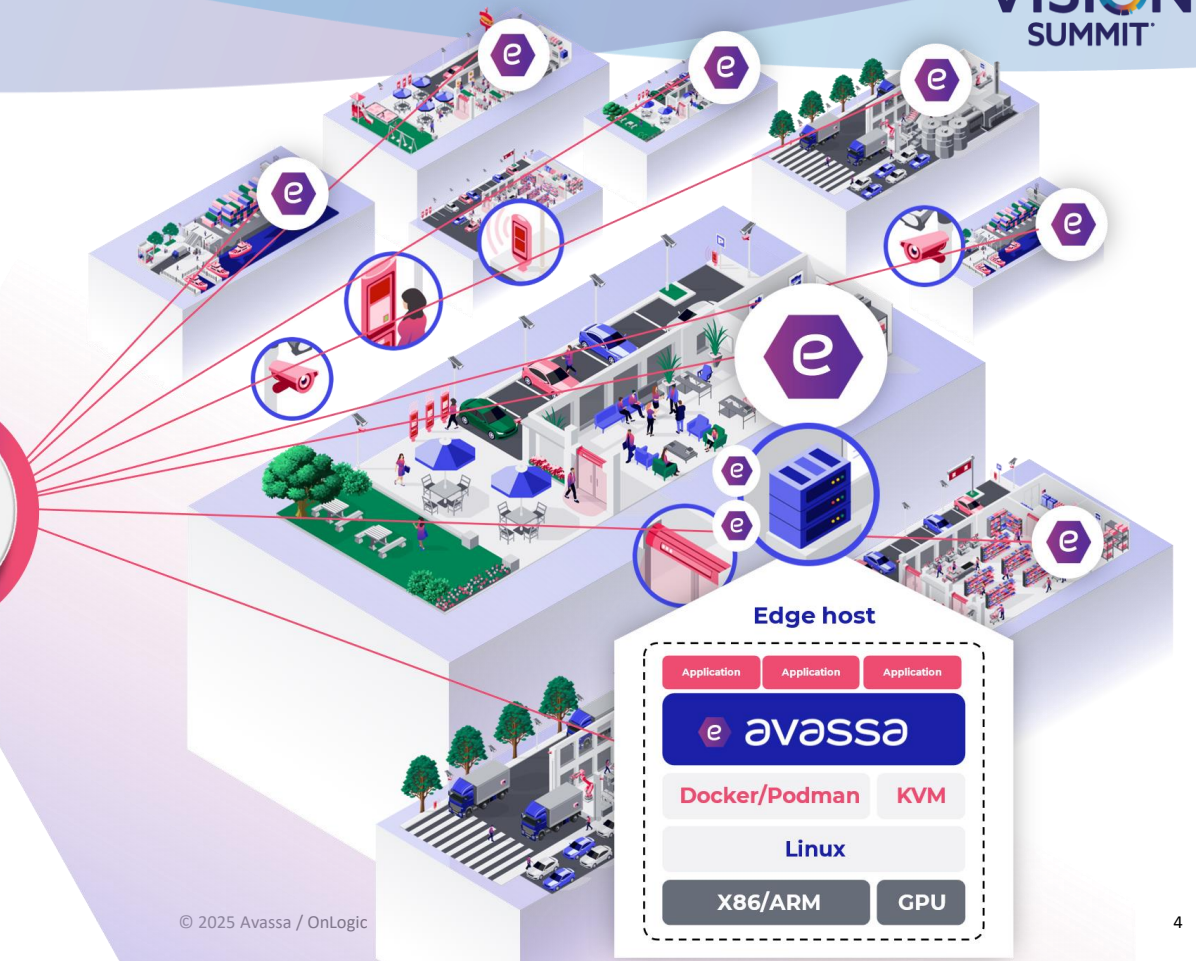
…but operating AI at the edge in resource-constrained environments also requires **scalable, purpose-built infrastructure**.

In this presentation we will…

- Learn what to consider when:
  - Selecting edge AI devices
  - Managing the lifecycle of edge AI applications
- Lay out the components of a best-of-breed approach to an edge AI stack

Introducing Avassa





avassa ONLOGIC

Introducing OnLogic

Millions of Configurations

Systems tailored for your exact needs

Collaborative Sales Approach

No middleman when it comes to solving your challenges

Engineered for the Edge

Rugged and reliable performance where it matters most

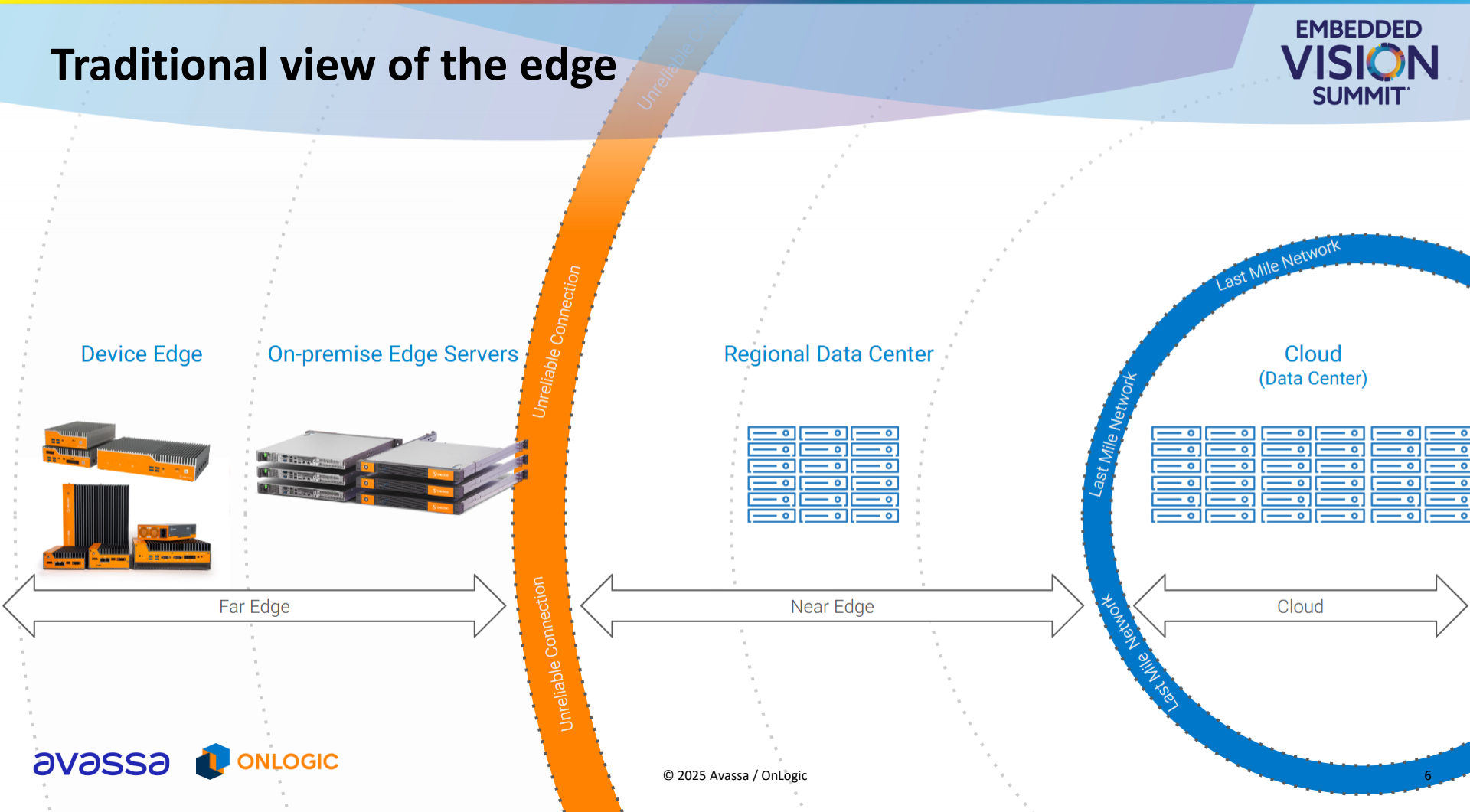
Prototype to Production

Our engineers can design, deploy, and scale any project

Built for scale

Expansive global production capacity to meet growing demand

Traditional view of the edge



What do I need to consider when selecting edge devices?

Edge hardware considerations

Temperature

Dust

Constrained space

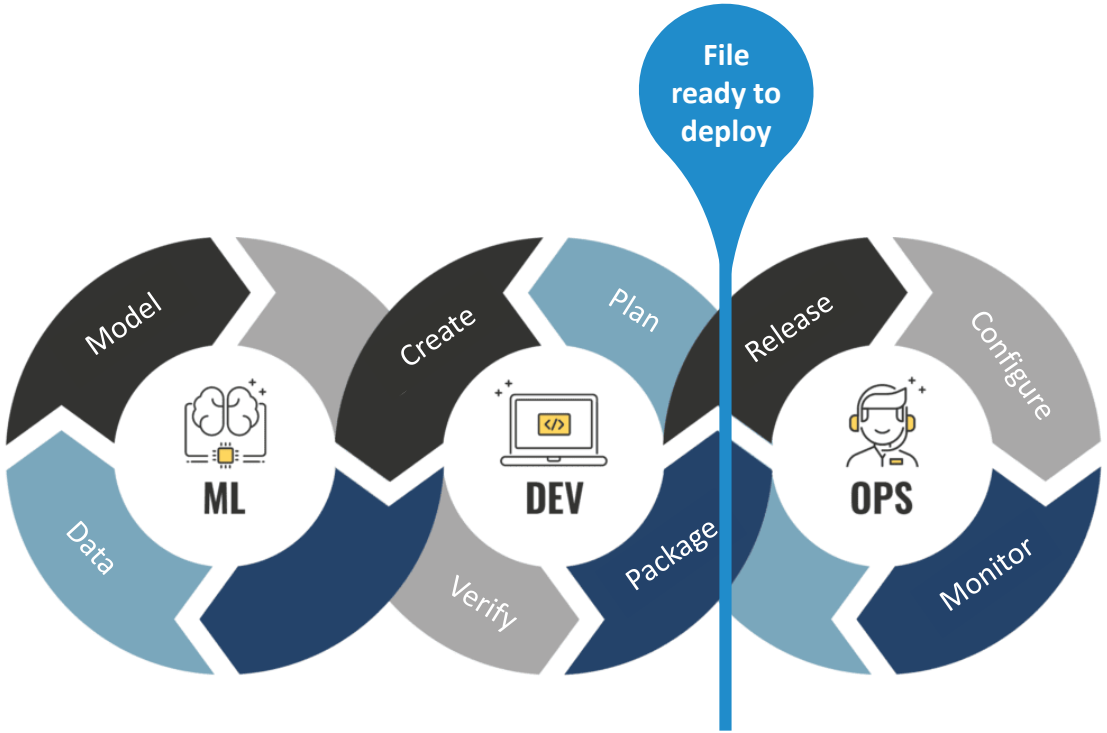
Vibration

… The edge is **a complex place with different and diverse challenges.**

What do I need to consider when managing edge AI applications?

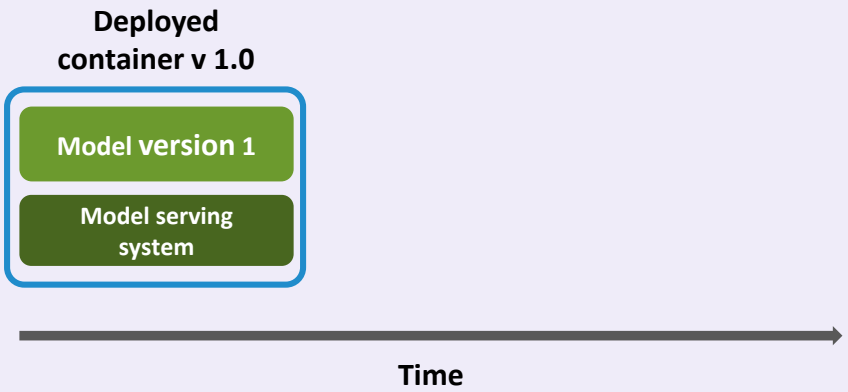
MLOps meets lifecycle management

MLOps



Lifecycle Management

Purpose-driven model lifecycle management

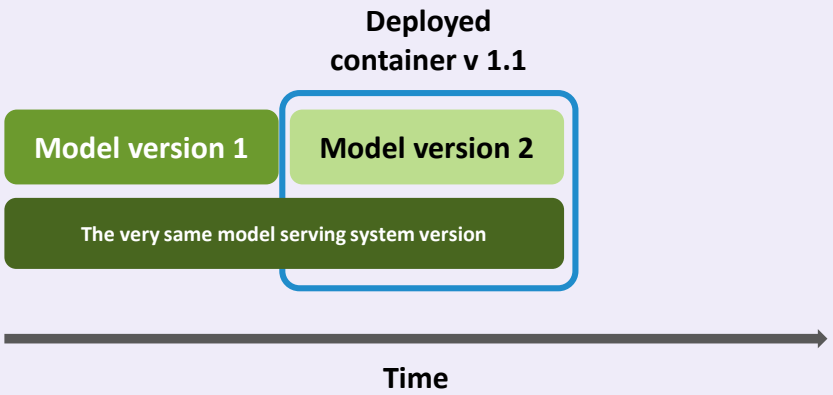


Container Layers

Model Container Layer

Model Serving Container Layer

Purpose-driven model lifecycle management

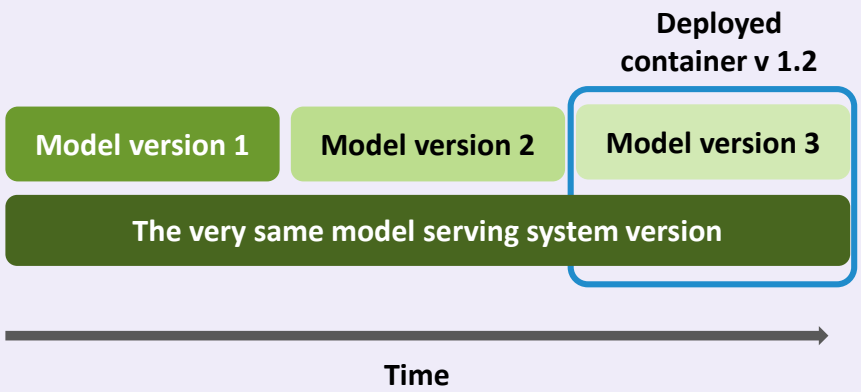


Container Layers

Model Container Layer

Model Serving Container Layer

Purpose-driven model lifecycle management



Container Layers

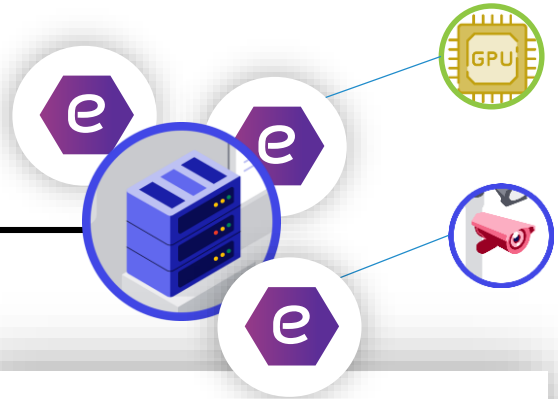
Model Container Layer

Model Serving Container Layer

Targeted deployments of containers, only where they are needed

Discovery and rule-based labels

for application placement



Targeted deployments

- Declarative requirements in your application
  - Define what and where
- Automatic placement of application at the edge

Automatic discovery

- Leaf device discovery using Linux kernel features
- Local rule-based label management
- Automatic identification and mounting of local devices
- Local placement based on GPU availability

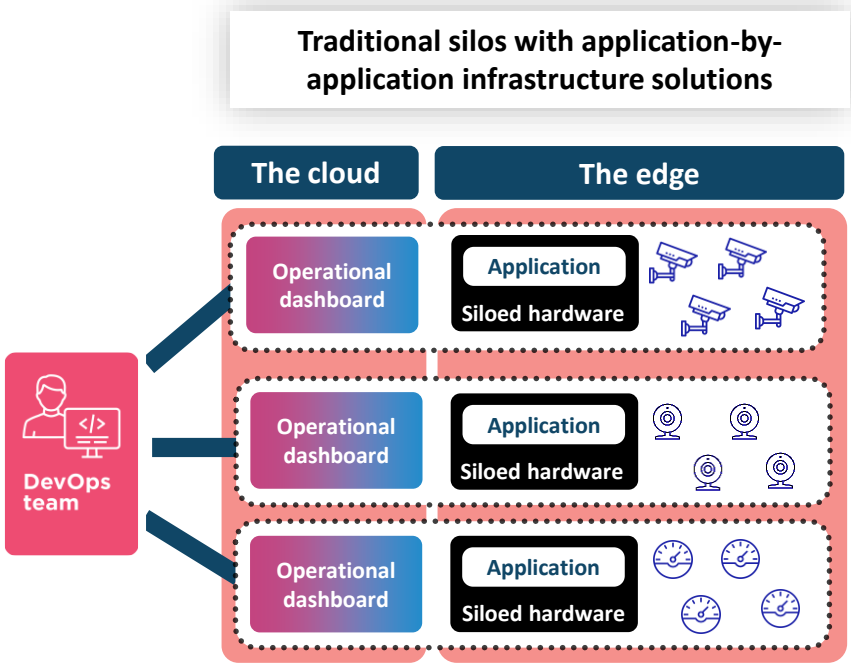
Lifecycle managing AI models at the edge

Key considerations:

- The model changes faster **than the application**
- Hardware and GPU requirements **drive application placements**
- Allow for offline scenarios by making sure your **key application services are kept local**
- Re-utilize existing investment and **integrate with current systems**

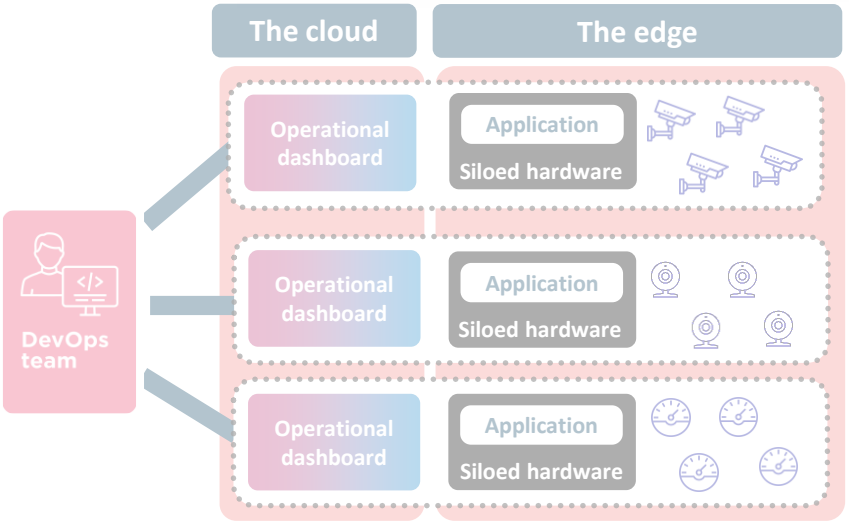
Why best of breed?

Avoid silos and redundant stacks

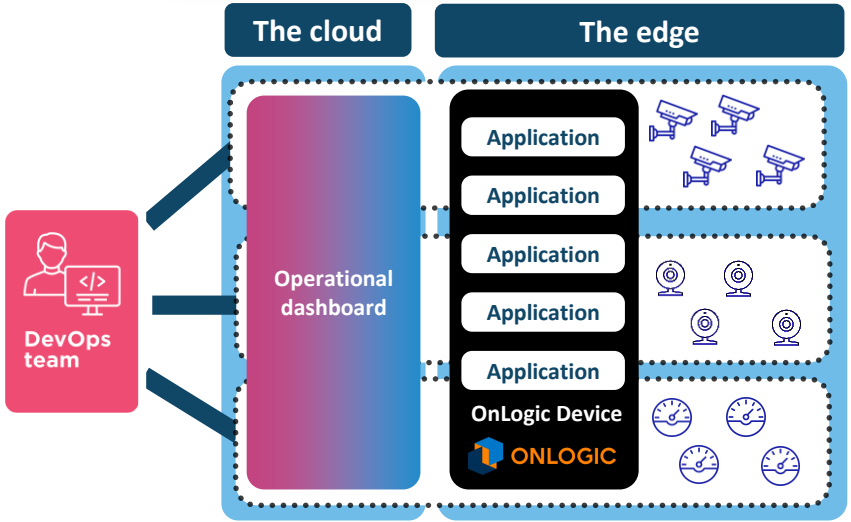


Avoid silos and redundant stacks

Traditional silos with application-by-application infrastructure solutions



A modern unified platform with streamlined infrastructure and single-pane-of-glass overview



Hardware that withstands its environment

Industrial Computers

- Fanless or active cooling
- x86 and ARM architecture
- Industrial operating temp. (0-50°C)

Rugged Computers

- Resistant to shock & vibration
- Wide input (12-48VDC)
- Wide operating temp. (-40-70°C)

Panel PCs / HMIs

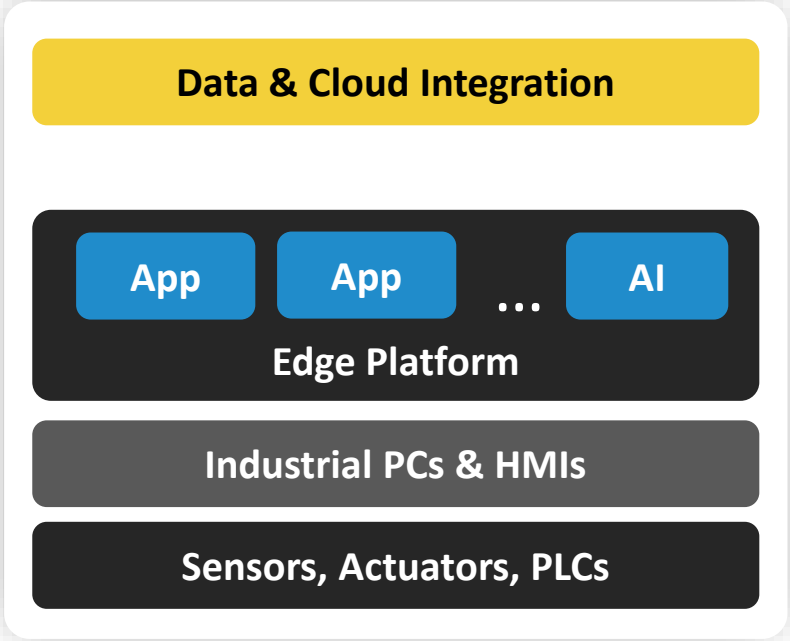
- 8.4” to 24” screen sizes
- Resistive or capacitive touch
- Up to IP69K ingress protection

Edge Servers

- 1U to 4U sizes
- Intel and AMD options
- Highly customizable

avassa ONLOGIC

A modern, right-sized infrastructure stack for edge AI



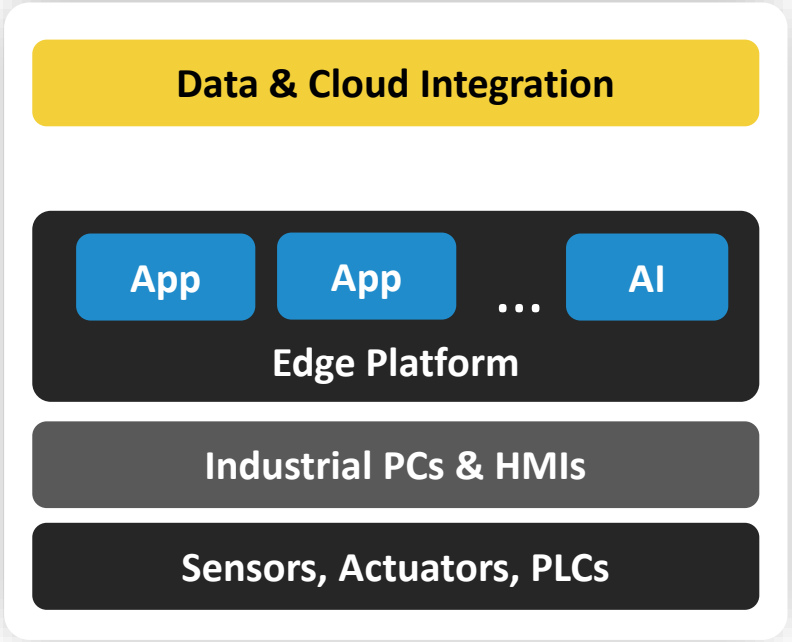
MLOps integrations to continuously deploy and monitor AI workloads through APIs

AI workloads comprise several parts (server, code, model, configuration) with separate lifecycle needs

CPUs with integrated GPUs and NPUs for efficiency, dedicated accelerators when needed.

Control logic going virtual – towards distributed, software-defined control architectures.

A modern, right-sized infrastructure stack for edge AI



avassa

ONLOGIC

A unified stack, allowing multiple workloads to share common infrastructure

Conclusions

- **Edge AI success** demands rugged hardware, scalable orchestration, and seamless cloud integration — not just smarter models.
- **A best-of-breed stack** right-sized infrastructure prevents silos and manual overhead, unlocking a unified, agile, and scalable edge environment ready for the challenges of tomorrow.
- **Pairing industrial-grade devices with smart application management** builds Edge AI systems that adapt and thrive over time.

Resources

[Making the edge lovable for your development and application operations team]

[Optimizing Edge AI: Combining MLOps and Edge Orchestration for Success]

[Solution description: Avassa for Edge AI]

[Solution description: OnLogic for Edge AI]

2025 Embedded Vision Summit

Come see us at booth 521!

Carl Moberg

calle@avassa.io

Zoie Rittling

zoie.rittling@onlogic.com