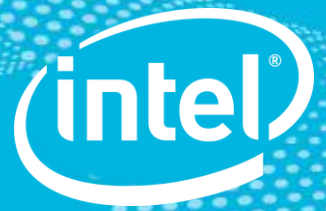


Federated Edge Computing System Architectures

Vaidy Krishnamoorthy, Intel Corporation

September 2020

Notice and Disclaimers

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at www.intel.com.

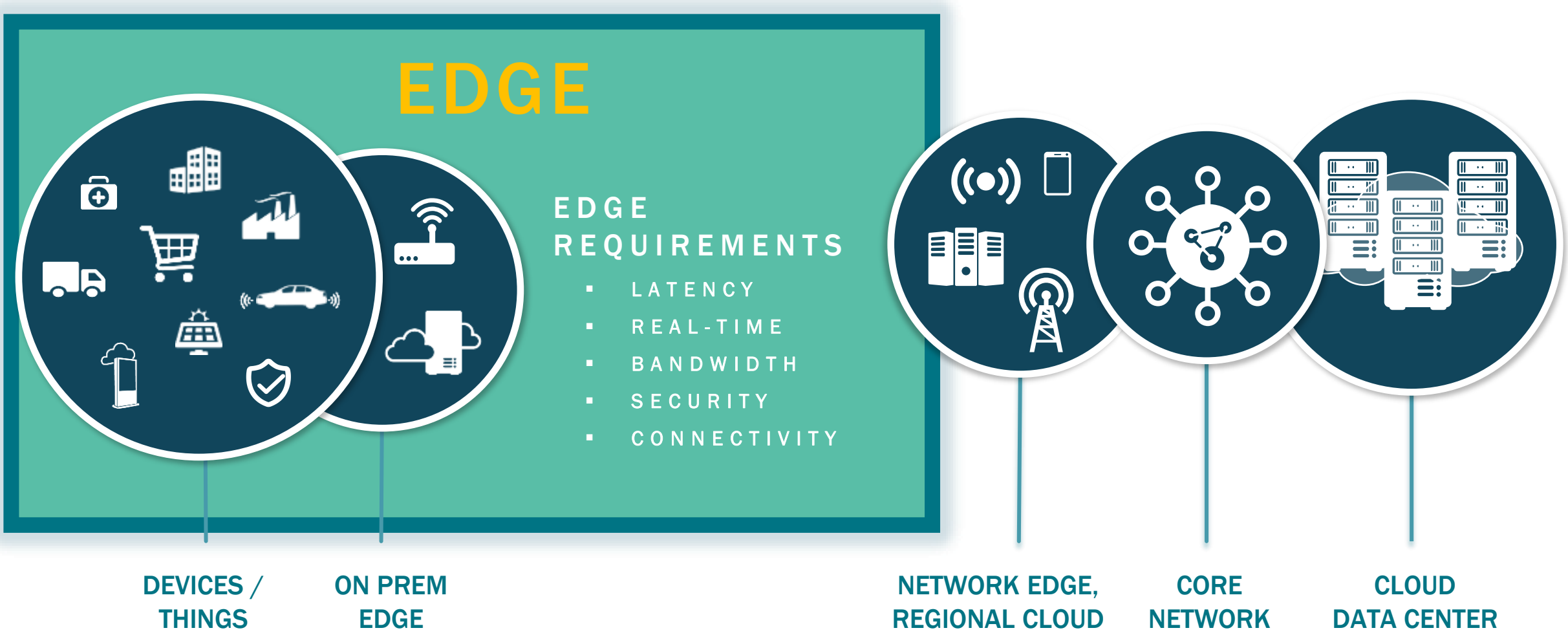
This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

Agenda

- Need for Edge Inferencing
- Federated Edge Inferencing Architectures
  - Home / Small Retail
  - Large Retail
  - Edge Cloud Orchestration with Kubernetes
  - Edge with AI Accelerators
  - Edge with AI Appliances
  - Industrial Edge (Private 5G)
  - Telco Edge (future)
- Intel Products addressing Edge Inferencing

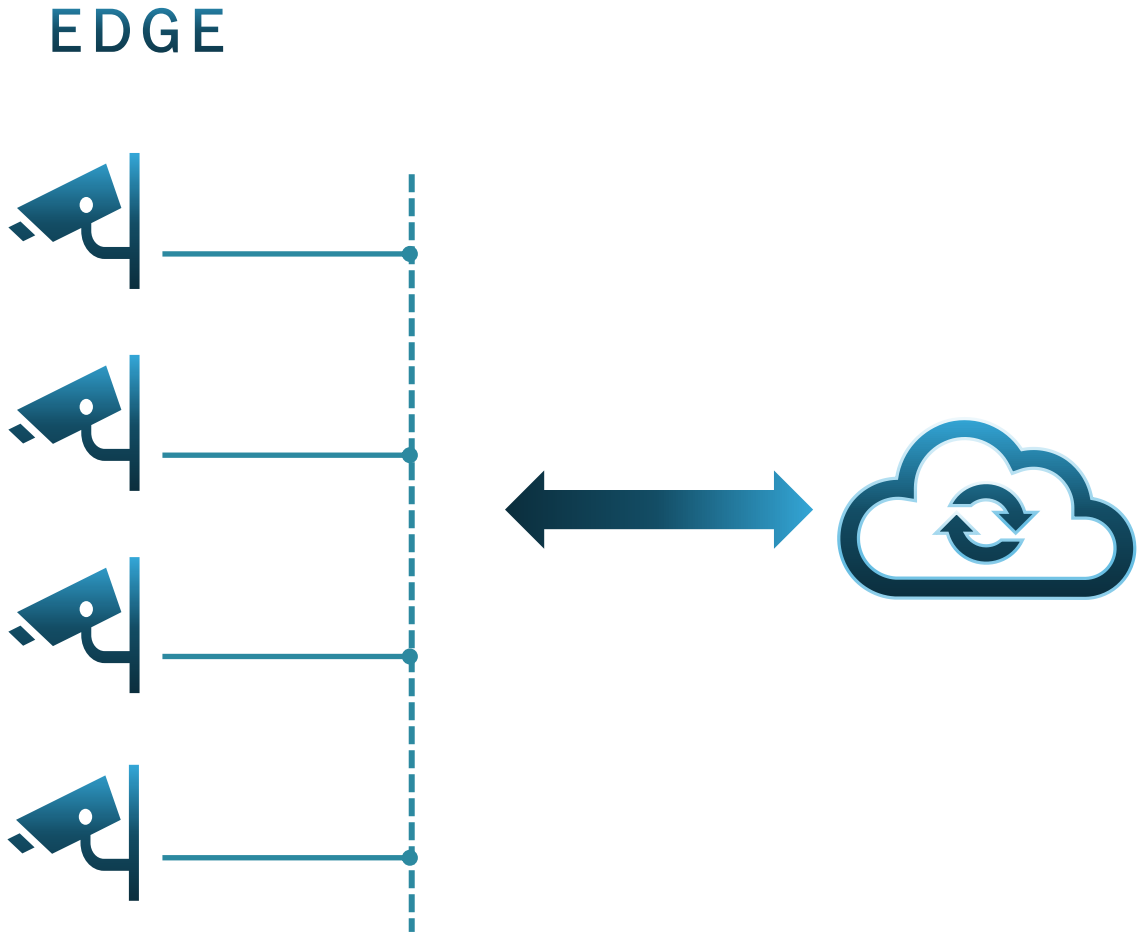
Need for Edge Inferencing

The Edge and its Challenges



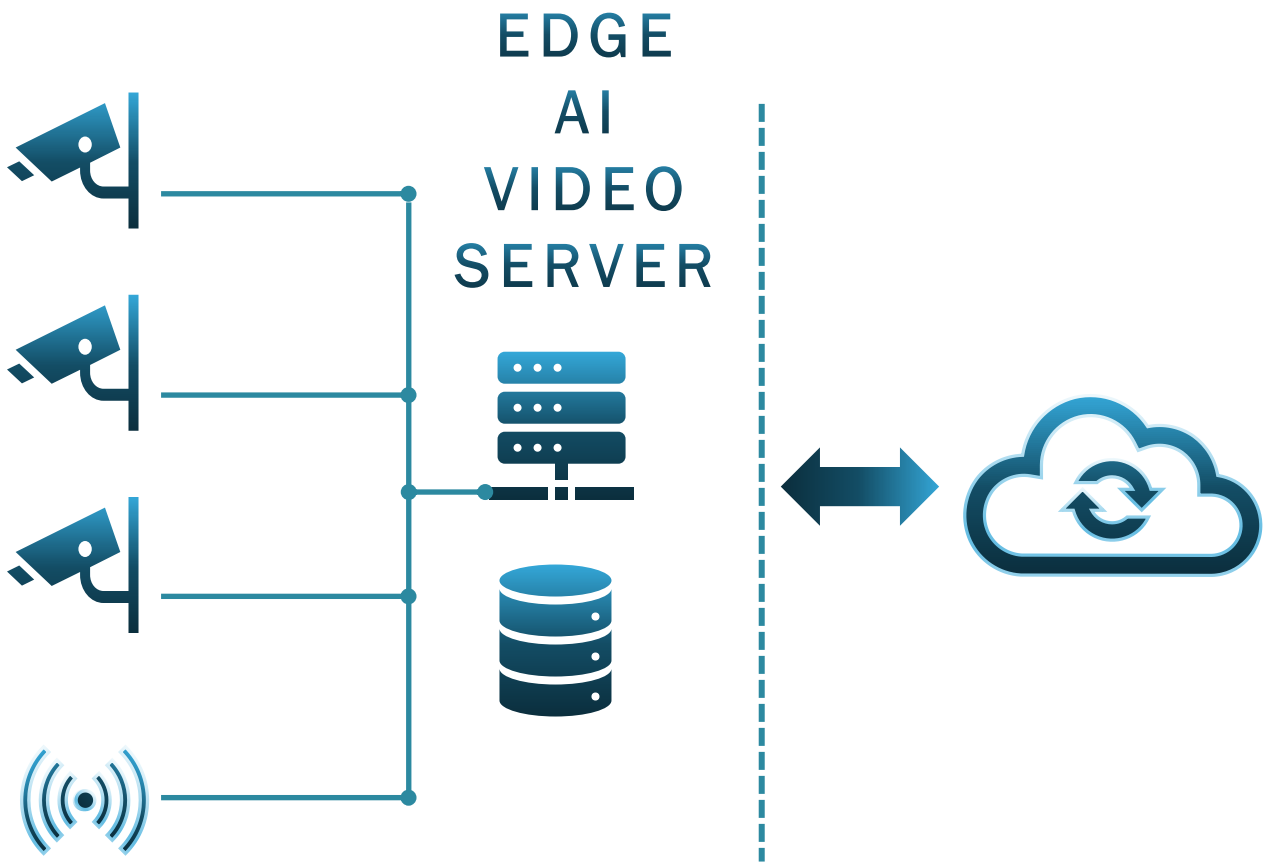
Federated Edge Inferencing Architectures

Residential and Small Retail



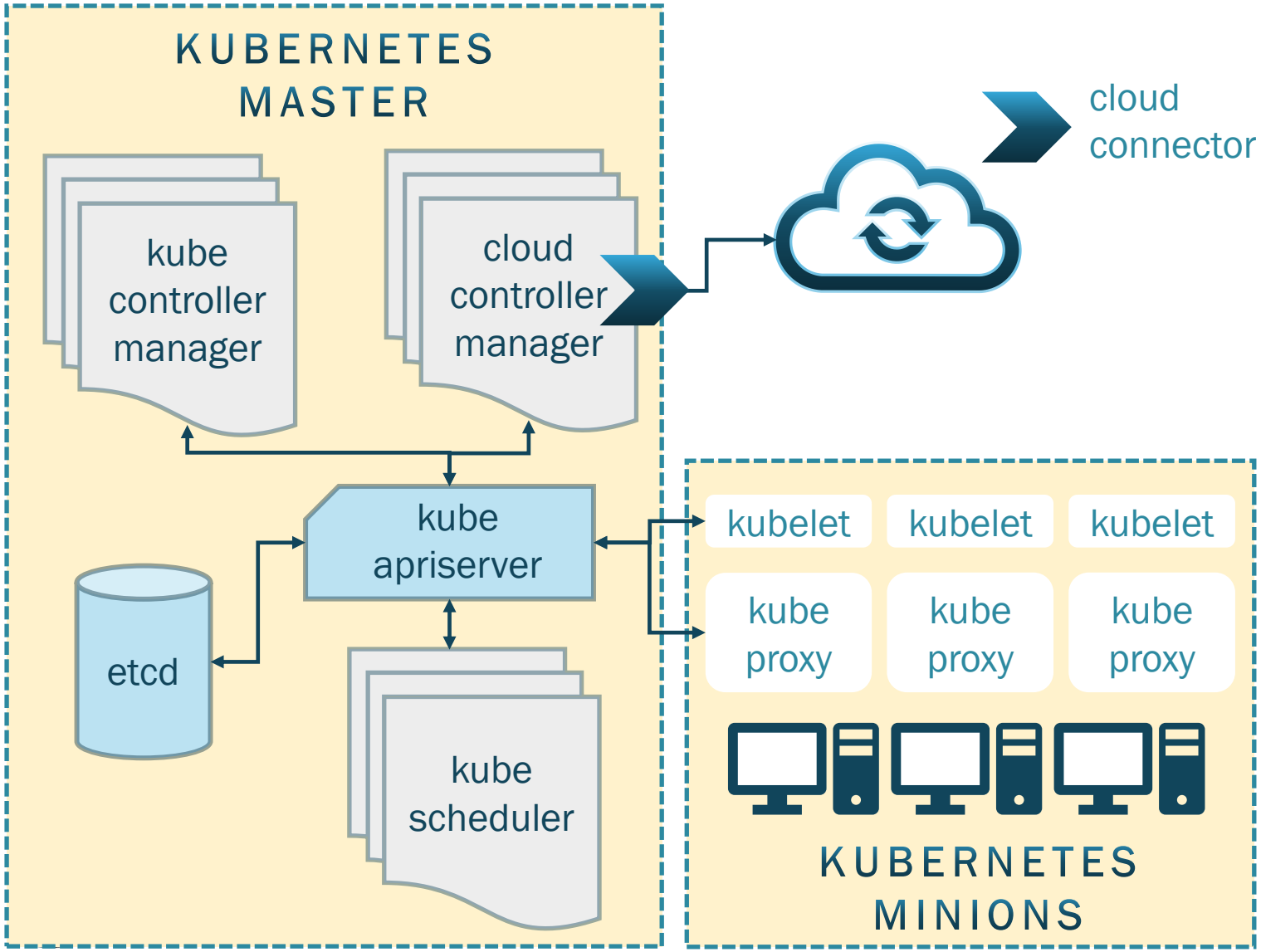
- IP Cameras connected to Cloud directly
- Limited Cameras
- Video as a service Business Model possible
- Use-cases:
  - Person Detection
  - People Counting
- Challenges:
  - Latency Critical usages
  - Device Manageability

Large Retail



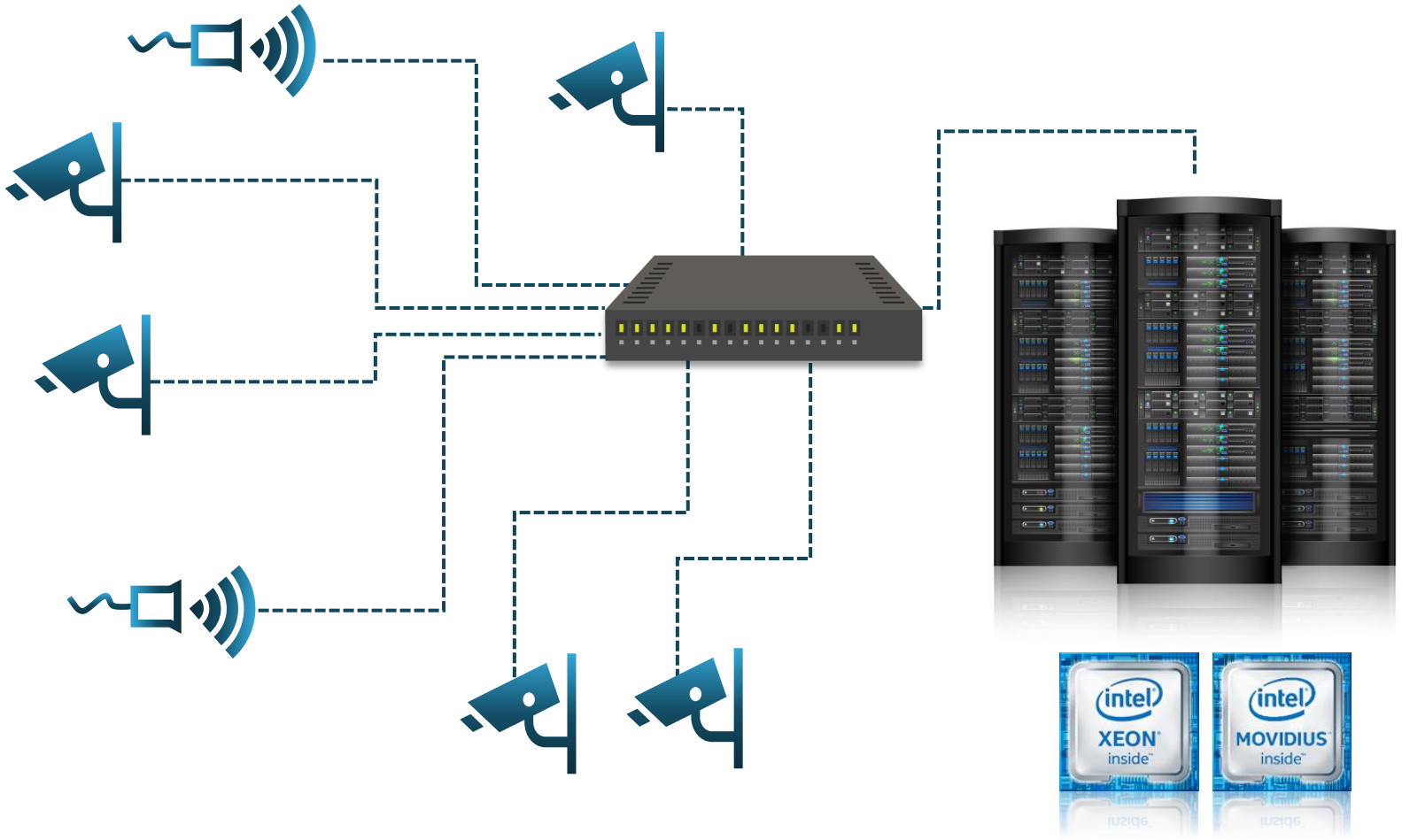
- IP Cameras and other sensors connected to Edge AI servers
- Use-cases:
  - Person-of-Interest
  - People Counting
  - Demographics
  - Social distancing monitoring
  - Shelf Monitoring
- Scale Edge AI server based on workload
- High Availability and Reliability using Cloud-native stacks
- Ability to federate workloads in Edge AI servers using Cloud stacks
- Work even with intermittent connection interruption with the cloud
- Possibility to add AI Cameras

Edge Cloud Orchestration



- Kubernetes provides Cloud Orchestration and High Availability
- Edge Cloud scaling based on workload
- Cloud used for:
  - Orchestration of Edge servers and analytics
  - Retraining to improve model accuracies
  - Long-term Storage
  - Manageability and Software Upgrades

Edge with Server+ AI Accelerator K8s Cluster



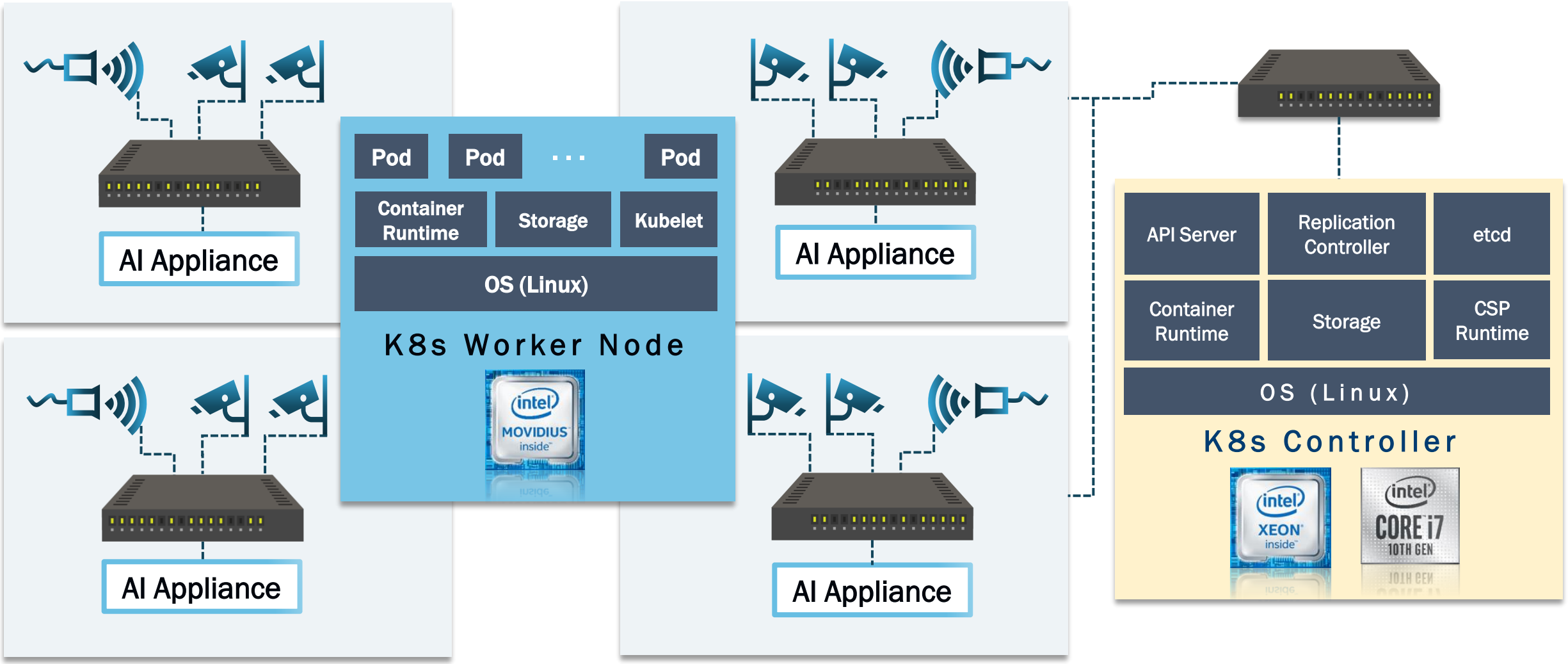
API Server	Replication Controller	etcd
Container Runtime	Storage	CSP Runtime
OS (Linux)		

K8s Controller

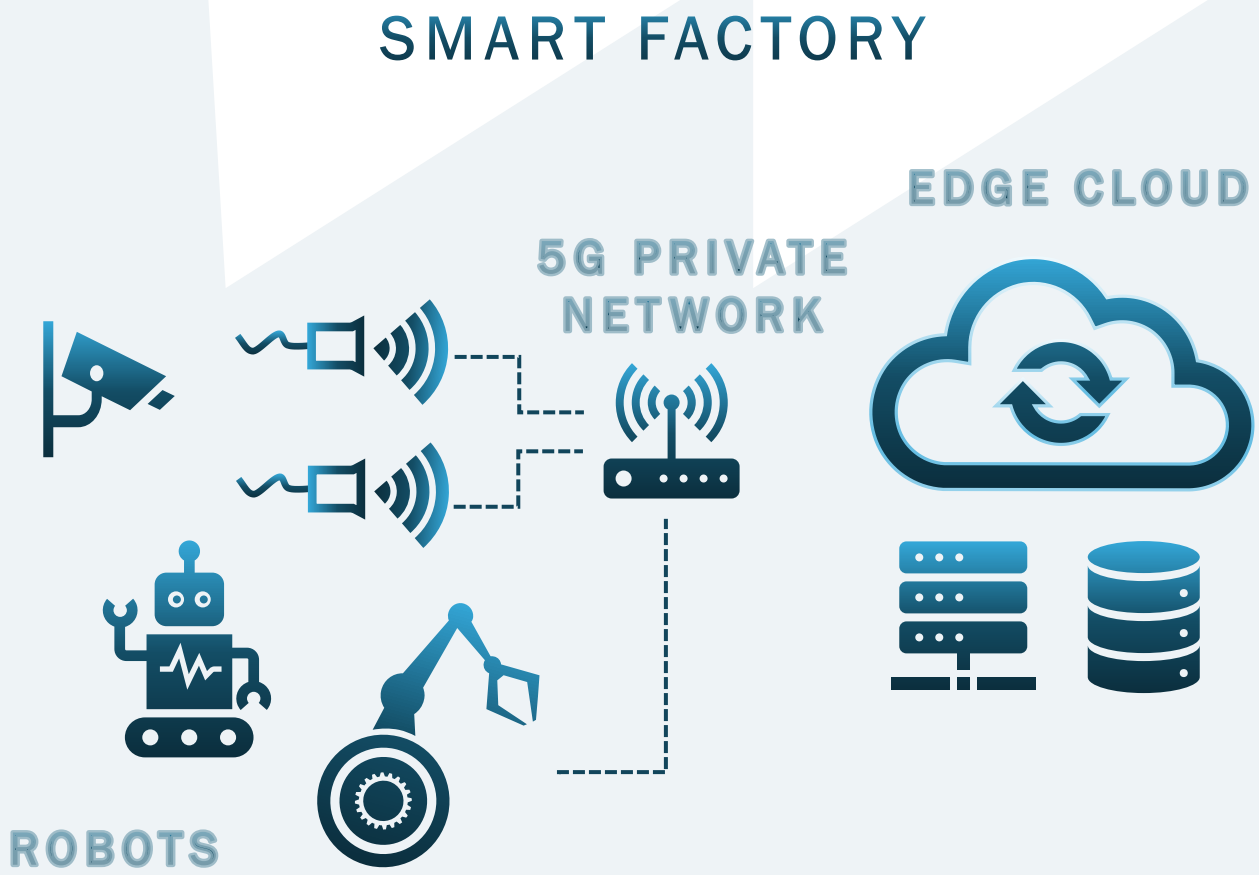
Pod	Pod	...	Pod
Container Runtime	Storage	Kubelet	
OS (Linux)			

K8s Worker Node

Edge with AI Appliance K8s Cluster



5G Private Network Cloud for Industrial



- Edge cloud supporting 5G RAN
- Autonomous Systems needing low latency
- Industrial use-cases
  - Fault Detection
  - Robot Monitoring
  - Emergency stop systems
- Private / Hybrid Cloud implementation
- Scale Edge Cloud based on Workload
- Orchestration using Cloud native frameworks

Telco Cloud (Future)

2020 embedded VISION summit

EDGE DISTRIBUTION IS UNDERWAY

VALUE DRIVERS

UP UNTIL NOW

LATENCY BW, SECURITY PRIVACY

EDGE

CLIENT

NEXT STEP EVOLUTION

LOWER TCO

EDGE

CLIENT

FUTURE EDGE

Decentralized, Stateful, Discoverable, Collaborative Edge

TELCO CLOUD

EDGE

CLIENT

TOWARDS… A MORE DYNAMIC EDGE

intel

Intel Products addressing Edge Inferencing

Intel Products

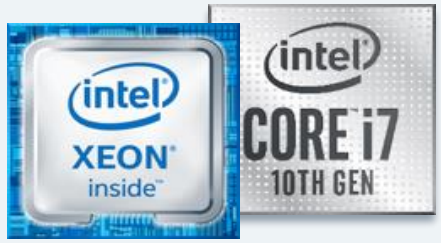
SOLUTIONS

SOFTWARE

PROCESS

Workload Breadth — AI Specific

CPU + GPU — FPGA — DEDICATED



Multi-Purpose AI Foundation

AI, HPC, Media & Graphics

Real-Time & Multi-Use DL Inference

Edge DL Inference

habana An Intel Company

Data Center DL Inference (Goya)

Data Center DL Training (Gaudi)

STORE

intel OPTANE MEMORY

intel OPTANE SSD

INTEL 3D NAND SSD

MOVE

intel Silicon Photonics

intel Ethernet

BAREFOOT NETWORKS | an Intel company

Recap

Recap

- AI Inferencing drives Edge Deployments
- Cloud-Native frameworks @ the Edge
- Intel – an Active Contributor to Cloud-Native Frameworks
- Different Intel Products scaling to needs @ the Edge

Resources

1) Kubernetes - https://www.Kubernetes.io
2) Kube Edge - https://www.kubeedge.io
3) Intel Movidius Products -

   https://www.intel.com/content/www/us/en/products/processors/movidius-vpu.html

4) Intel OpenVINO - https://docs.openvinotoolkit.org/