Federated Edge Computing System Architectures

Vaidy Krishnamoorthy, Intel Corporation
September 2020
Notice and Disclaimers

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY
ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN
INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL
DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY
OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT,
COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

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.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other
names and brands may be claimed as the property of others.
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

EDGE

**EDGE REQUIREMENTS**
- Latency
- Real-time
- Bandwidth
- Security
- Connectivity

DEVICES / THINGS
ON PREM EDGE

NETWORK EDGE, REGIONAL CLOUD
CORE NETWORK
CLOUD DATA CENTER
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

K8s Controller
- API Server
- Replication Controller
- etcd
- Container Runtime
- Storage
- CSP Runtime

K8s Worker Node
- Pod
- Container Runtime
- Storage
- Kubelet

OS (Linux)
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
EDGE DISTRIBUTION IS UNDERWAY

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

VALUE DRIVERS

TOWARDS... A MORE DYNAMIC EDGE
Intel Products addressing Edge Inferencing
Intel Products

SOLUTIONS
SOFTWARE

Workload Breadth

CPU + GPU + FPGA

Multi-Purpose AI Foundation
AI, HPC, Media & Graphics
Real-Time & Multi-Use DL Inference

Edge DL Inference
Data Center DL Inference (Goya)
Data Center DL Training (Gaudi)

AI Specific

STORE
MOVE

Intel XEON inside
Intel CORE i7 10TH GEN
Intel XE
Intel STRATIX ‘10 inside
Intel MOVIDIUS inside

Optane Memory
Optane SSD
Intel 3D NAND SSD

Silicon Photonics
Ethernet
barefoot

Intel Products
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
2) Kube Edge - https://www.kubeedge.io
4) Intel OpenVINO - https://docs.openvinotoolkit.org/