

TietoEVERY Scalable Edge Reference Platform

September 2021



Table of contents

I. Introduction	3
II. MEC platforms transformation – from dedicated embedded solution to multi-workload platform	5
III. Building flexible edge platform using Openness	6
IV. Deploying and configuring Smart City safety solution (TietoEVERY Automatic Pedestrian Alert System reference application)	8
V. Edge Deployment	14
VI. Total Cost of ownership (TCO) considerations	14
VII. Summary	16
VIII. Resources	17
Contact information	18
About TietoEVERY	18

I. Introduction

OVERVIEW

TietoEVERY edge reference platform, the open-source based solution demonstrated in this paper, is a viable alternative to existing commercial edge solutions to improve total cost of ownership (TCO) without compromising on performance. The platform, due to its flexibility, scalability, cloud-native characteristic and Intel technology is powerful for emerging edge use cases across various industries such as Telecom, Industry, Enterprise, IoT, SmartCities, SmartHomes, Automotive or MedTech.

This paper is divided into 3 major parts.

- First (chapters I, II) discussing the problem, MEC (multi-access edge computing) evolution and the use case itself
- Second (chapters III and IV) explaining technical solution, deployment, benchmarking, and scalability
- Third (chapters V to VII) describing the advantages solution brings to the industry and business



MEC (Multi-access edge computing) and what it brings to the industry

Today's booming demand for low latency edge applications creates the need for highly flexible, scalable, and automated solutions. Edge computing [1] is transforming the way data is being handled, processed, and delivered from millions of devices globally. The explosive growth of internet-connected devices, fueled by fast networking (such as 5G), accelerate the creation of new use cases, such as video analytics, self-driving cars, artificial intelligence, robotics, and others.



Multi-workload scalable platform

The goal of this paper is to exemplify the application of TietoEVERY (which the company provides to the industry), based on extensive technology experience and strong partnerships, in building cutting edge solutions powered by Intel technology.

We selected the Open-source Intel Smart Edge Open (formerly known and referred to in this paper as OpenNESS) platform to create a reference implementation of a scalable edge computing platform. The example deployment comprises of real-time, AI inference video analytics solution for Smart Cities implementing an Automated Pedestrian Alert System (APAS).

OpenNESS exposes Intel hardware features to the Kubernetes based, containerized Edge environment and enables easy deployment and optimized orchestration (using OpenVINO™, Data Plane Development Kit (DPDK), Real-time kernel etc.) of various edge use cases such as media analytics, Content Delivery Network (CDN) up to 5G access, and core network functions.



Smart City – pedestrian safety use case

Urbanization increases the number of people and vehicles on the move in metropolitan areas [2]. As road traffic increases, so does the risk of traffic accidents, especially at intersections. The risk of injury and death is especially high for pedestrians. APAS is an IoT solution to improve pedestrian safety when crossing (or intending to cross) the street by alerting incoming cars about the potential risk of a person on the crossroad via the car's cockpit warnings. This is especially crucial for multi-lane crossroads (where stopped cars may limit another driver's zebra crossing visibility) or in bad weather conditions. [3]



Edge ready

The successful implementation of open and intelligent edge computing solutions relies on the deployment of commercial-off-the-shelf (COTS), white-box hardware closer to sensors, devices, and end-users. These edge computing platforms need to deliver high-performance processing while providing sustainable operations in outdoor or semi-outdoor environments, keeping a constant eye on power consumption. Advantech 5G Edge Servers balance the latest generation Intel® Xeon® Scalable processors with advanced AI and video acceleration in optimized, high density, high-reliability platforms designed to withstand edge environmental conditions.

II. MEC platforms transformation – from dedicated embedded solution to multi-workload platform

The edge computing concept has been used in the communication industry for years, with proven benefits both for end-users and service providers. Many former deployments were based on dedicated function embedded boxes, introducing vendor lock-in, and constrained potential for scaling and with long Time-To-Market (TTM) for new services.

The OpenNESS based scalable edge platform presented in this paper removes the majority of these problems. Opensource, container-based, running on off-the-shelf hardware dramatically reducing TTM for new services, improves scaling, removing vendor lock-in as it can be run on standard x86 based devices, making updates/upgrades fast and easy, and of course, reducing Total Cost of Ownership (TCO) (both on CAPEX and OPEX side)

The platform built for this case is based on Advantech SKY-8000 Series of 5G Edge Servers integrating the Intel® Xeon® Gold 5218R CPU and Intel® Movidius™ Vision Processing Unit (VPU) acceleration, providing significant improvement of performance-to-cost ratio for our selected use case.







Considering the platform's flexibility and OpenNESS, Kubernetes based multi-node orchestration, new microservices in the form of cloud-native application can be run depending upon the need within seconds on any edge node of the operator choice. OpenNESS package provides several potential deployment flavors which are optimized for specific use case types, like 5G, video analytics and many others. Based on Cloud Native principles the platform can easily host a combination of workloads coming from O-RAN, intelligent edge or i.e., enterprise CPE use-cases. Additionally, the platform itself can provide additional performance/functionality boost by simply adding Intel's based PCIe smart NIC offload card, which is natively supported by OpenNESS based Converged Edge Reference Architecture (CERA) distribution.

III. Building flexible edge platform using Openness

OpenNESS provides open-source, Kubernetes based, free to use, software toolkit to deploy applications on the edge. In addition to network services, OpenNESS provides telemetry and life cycle management for application services.

The edge reference platform solution described in this paper is using OpenNESS (CERA Media analytics flavor) on SKY-8101 [4] server from Advantech. TietoEVRY APAS reference application is deployed on top of that making best use of OPENVINO, Intel® Deep Learning Boost (Intel® DL Boost) [5] and Intel Movidius VPU features [6].

Edge Server Bill of Materials (BoM) options

System component/ Capacity	Test Platform	Minimum configuration to serve 44 camera streams	Minimum configuration to serve 24 camera streams with VPU	Minimum configuration serving 34 camera streams with pure CPU processing
	Advantech SKY-8101 Edge Server			
	Intel® Xeon® Gold 5218R CPU @2.10GHz	Intel® Xeon® Gold 5218R CPU @2.10GHz	i.e. Intel® Xeon® Silver 4210R CPU @2.40GHz	Intel® Xeon® Gold 5218R CPU @2.10GHz
	Advantech VEGA-340 PCIe x4 with 8x Intel® Movidius™ Myriad™ cores	Advantech VEGA-340 PCIe x4 with 8x Intel® Movidius™ Myriad™ cores	Advantech VEGA-340 PCIe x4 with 8x Intel® Movidius™ Myriad™ cores	
	192GB DDR-4 2666 (32G per socket)	min 56GB RAM	min 40GB RAM	min 50GB RAM
	480GB, SSD	min 92GB, SSD	min 92GB, SSD	min 92GB, SSD
	Intel Corporation I210 Gigabit Network Connection driver compatible card			

Software Stack BoM

- Operating System: CentOS 7.6 Linux distribution
- Intel Distribution of OpenNESS 20.09.01
- OpenVINO™ 2021.2.0-1877
- TietoEVERY APAS application (containerized)

Deployment, orchestration, operation & maintenance

Step by step deployment instruction of OpenNESS Cluster can be found in docs within openness.org. Several flavors (optimized deployment scenarios) are available in the package, that makes OpenNESS easily deployable with all Intel-based hardware optimizations enabled for a wide number of use cases. There are some pre-installation tasks to be done on all nodes, where OpenNESS needs to be deployed and then the existing deployment ansible scripts OpenNESS deployment can be used.

The OpenNESS platform is straightforward to use and well documented. Deployment instructions are easy to follow with minimal experience with scripting and k8s itself.

IV. Deploying and configuring Smart City safety solution (TietoEVERY Automatic Pedestrian Alert System reference application)

The Automatic Pedestrian Alert System application (APAS) is a system to warn drivers about pedestrians potentially crossing the roads, especially crucial in difficult driving conditions or multi-lane roads.



Figure #1 - Video inferencing snapshot with pedestrian recognition results

APAS application is running in the container, which is registered to OpenNESS EAA Microservice and is processing input video stream from single-camera source. The result of such inference is further processed by a neural network for pedestrian detection.

TietoEVERY team has conducted several accuracy tests to find optimal neural processing algorithm for the pedestrian recognition use case for APAS application to find the best performance/accuracy tradeoff. SSD (Single Shot MultiBox Detector) deep learning model for a neural network with FP16 precision was selected with a frame of 512x512 pixels, which is providing good accuracy during the processing of inference requests in various weather conditions. Not the whole frames (camera output) are processed but each crossing or camera output has its configuration file, determining 2 areas of interest for pedestrian detection (blue rectangles).

APAS serves as a producer of signals for car systems and consumer purchasing APAS service is subscribed to all relevant APAS signaling streams. For the demonstration, consumer is presented in a form of android application simulating car's digital cockpit.

The following figure shows an architecture of APAS application integration into OpenNESS based edge platform.

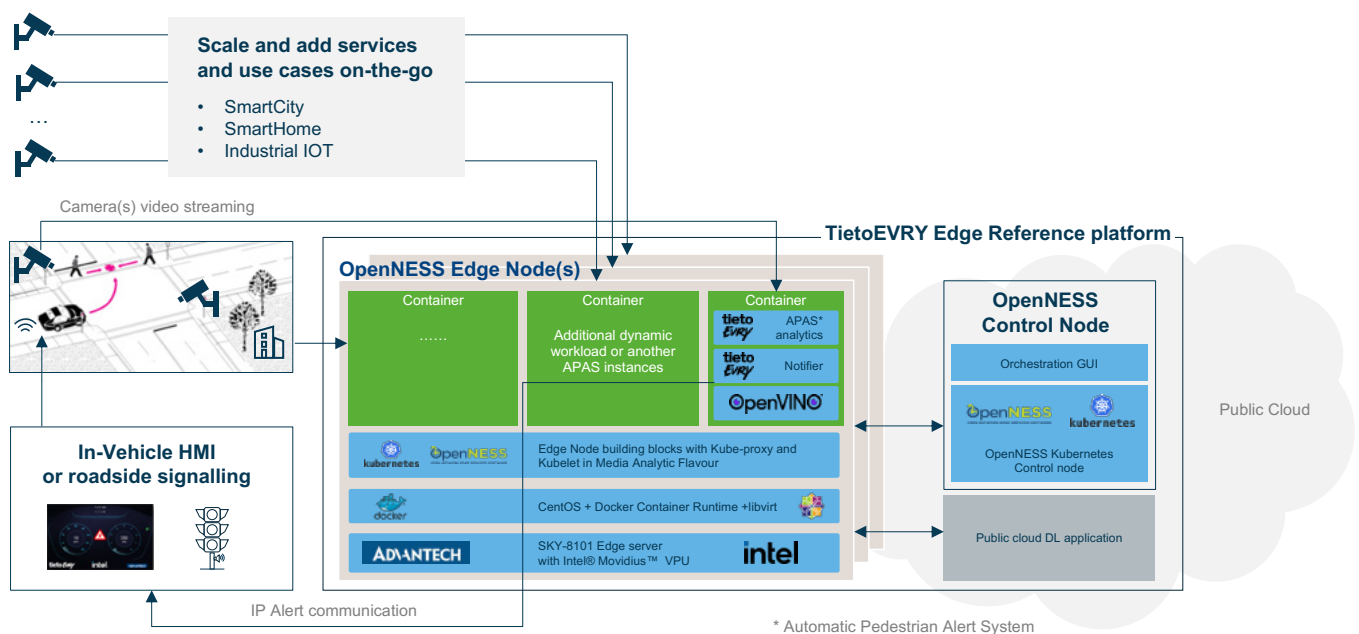


Figure #2 – Use case topology

Deployment and use

Video processing in the APAS application is done via BASH script as DaemonSets (Kubernetes objects). Once the objects are started, relevant camera streams are started to be processed.

In multimode deployment, DaemonSets can select nodes with the right hardware using matching Kubernetes labels (e.g., if just some of the nodes have VPUs). When DaemonSets are started, the application will open port and listen for incoming video stream to start processing it. As visible from the screenshot below, passing cars are simulated by a digital cockpit app, which is receiving alerts about pedestrians from APAS.



Figure #3 – TietoEVRY digital cockpit demo app

Performance and Capacity

Based on conducted research, to safely recognize a pedestrian approaching crosswalk (also known as a zebra crossing) it needs to be ensured that system can process each camera stream feed with minimum of 3 frames per second (fps). For this use case, the main key performance indicator (KPI) determining platform scalability/performance is number of camera streams with 3fps processing speed system can process. That KPI is indicating how many different camera streams can be processed on a single low/mid cost edge platform. Performance tests done on the system proved the linear dependency between camera fps

and maximum number of streams, which can be launched in parallel before the system becomes overloaded (cannot keep up with target fps) e.g. with 6fps the system was able to handle 22 streams.

System scalability and Intel technology accelerations (Intel Deep Learning Boost, Intel Movidius Myriad X Vision Processing Unit)

2nd Generation Intel Xeon Scalable processor Intel® Xeon® Gold 5218R CPU delivers dedicated instructions set (Intel® AVX-512) significantly improving AI deep learning and inferencing process with INT8 precision. This combined with Intel Movidius Myriad X VPU and OpenVINO libraries, optimized to use all technological advances of both technologies enables our use-case to be easily scalable to handle top of 44 camera streams inferencing in parallel, keeping the performance on target level (3fps) with a single, Intel Xeon based SKY-8000 Advantech edge platform. Combining the above technology with OpenNESS creates an easily orchestrated and scalable solution.

Benchmarking results

Workload parameter	Scenario/target	Description
Input stream	All scenarios	SSD model, FP16 for VPU & INT8 for CPU*, person-detection-0102.xml, 2x detection areas (512 x 512 each)
Number of camera streams	Scenario 1	34 camera streams processed exclusively by CPU
	Scenario 2	24 camera streams processed exclusively by VPU (VEGA-340)
	Scenario 3	44 camera streams processed by combination of VPU + CPU (24+20)
Target frames per second (fps)	3fps	We are aiming to get 3fps throughput from all camera streams. This is considered as minimal safe throughput for good pedestrian movement detection.

* INT8 precision model was used to leverage Intel® Deep Learning Boost technology

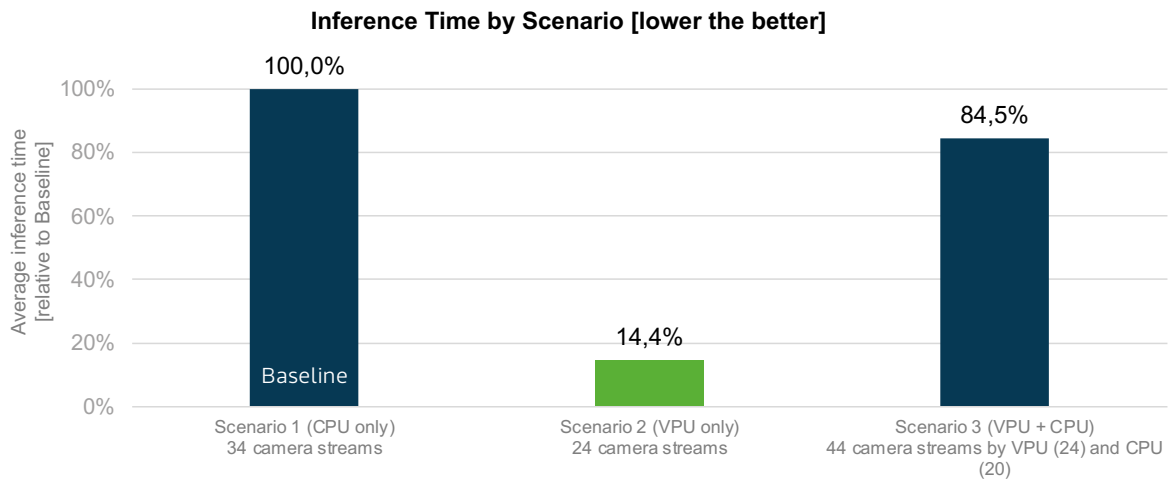
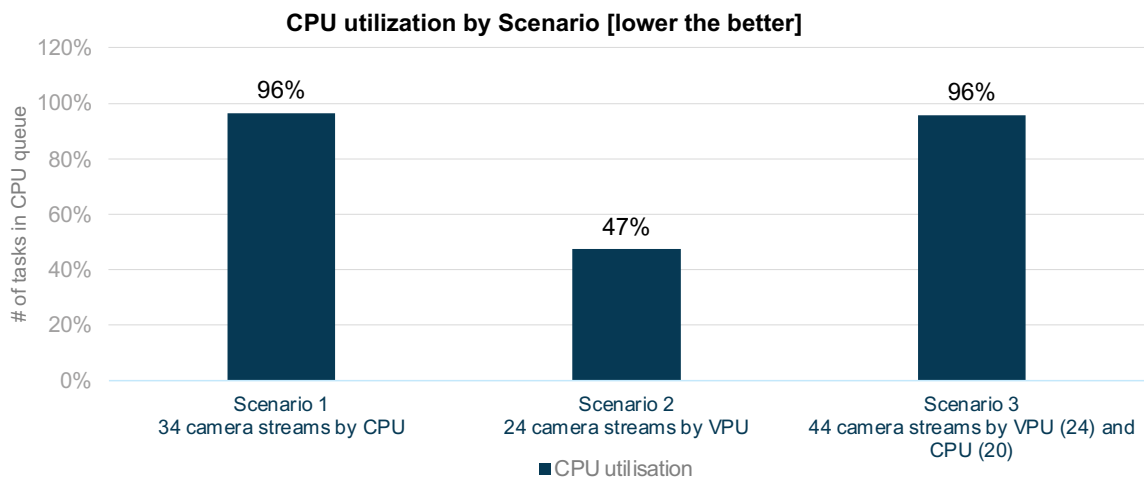
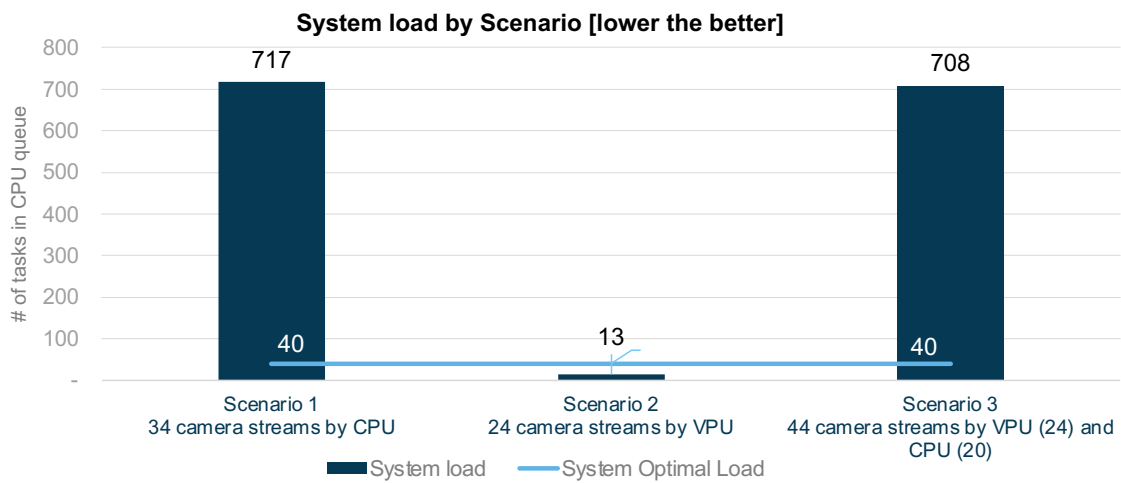


Figure #4 – Video inference performance chart



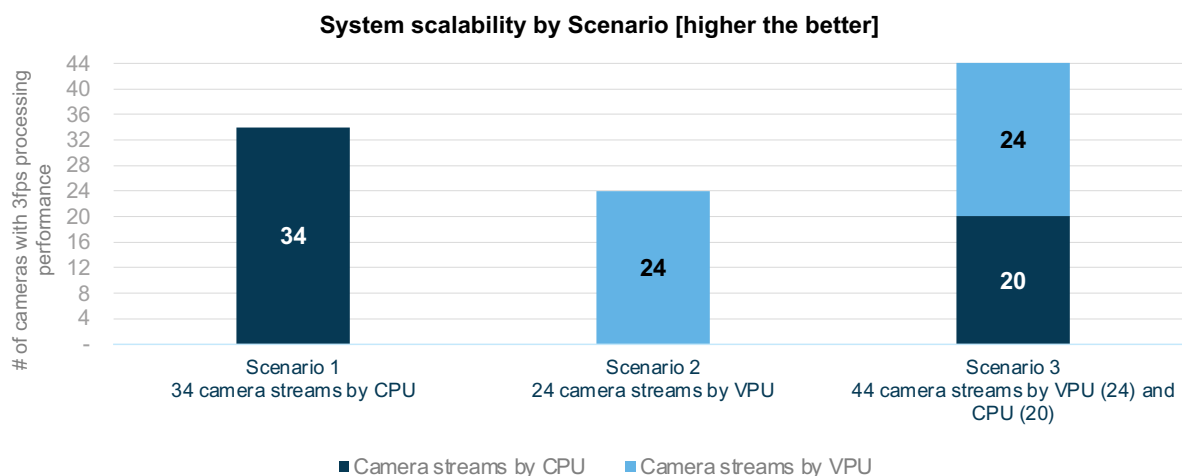


Figure #5 – System & CPU load and system scalability

Benchmarking scenarios are to check system from a performance and scaling capability perspective. The tested system can process 44 parallel camera streams total, however, the system is then in heavy load condition, which is causing occasionally extra-long inference requests (~1000ms). This issue can be resolved by decreasing the number of camera streams to (35-40) or optimizing inference even further, what will put system load on more healthy level. The tested setup is equipped with more than enough RAM memory and storage. To make sure of optimized solution cost, the minimum HW requirements have been specified with the assumption that Edge node under test is processing only a single workload (APAS) here [\[link\]](#). To make sure the system under test is running only workloads of interest OpenNESS control plane as well as traffic generator (camera streams) and analyzer is running on separated server.

V. Edge Deployment

The realization of smart city solutions such as the one presented in this paper relies on the deployment of a white-box edge infrastructure able to deliver sustained processing performance outside of controlled data center environments. The integration of COTS hardware in such solutions doesn't mean that standard IT servers can be used. On the contrary, special attention needs to be paid to critical hardware specifications such as density, reliability, serviceability, mechanical and environmental features to avoid greater inefficiencies in terms of service interruptions, onsite interventions, and energy consumption.

The selected Advantech SKY-8000 range has been designed to maximize throughput and AI acceleration density required by intelligent edge computing workloads. These carrier-grade servers can operate at extreme temperatures in both clean and dusty environments and can be deployed in IP65 pole mount, street, or roadside cabinets of special interest for smart city use cases. Enhanced reliability features include support of single failures for critical components such as fans, advanced platform management and redundant BIOS and FW images that not only provide a safe way to recover from component failures but also offer remote fail-safe update capabilities. All this combined with Advantech global services and support network that provides that extra peace of mind when deploying mission-critical solutions.

VI. Total Cost of ownership (TCO) considerations

The solution presented in this paper brings number of technological advancements which have positive impact on Operator's TCO. Using a commercial-off-the-shelf edge server platform instead of a dedicated function box makes upgrades and overall CAPEX go down. The use of optimized, high density and highly reliable edge servers such as the Advantech SKY-8000 maximizes energy efficiency while minimizing system downtime and on-site interventions, further reducing OPEX. In addition, by leveraging open architectures, operators can avoid vendor lock-in, building a seamless white-box edge infrastructure ready to host new and innovative smart city services.

Taking advantage of Intel's technology, especially important for this case Intel DL-Boost and Intel Movidius VPU acceleration pushes the overall CAPEX down even more. Considering benchmarking results from previous chapters this is evident that Operator can easily scale down the solution (and HW cost) depends on the real use cases.

Combining the HW versatility with OpenNESS, Kubernetes and Cloud-Native based opensource platform saves a lot of Operation and Maintenance (O&M) OPEX costs and enables operator to scale up and down solution easily as well as use its future proof capabilities to run new services (new Cloud Native Functions) in a time of seconds, combining them under one system, instead of deploying number of new dedicated boxes. This flexibility enables Operators to dramatically decrease TTM (Time to Market) for new services needed now and in the future.



VII. Summary

This paper outlines integration of multi-workload edge platform based on Open-source software and Advantech's off-the-shelf carrier-grade edge servers based on Intel Architecture. Pre-integration, optimization and offloading of most compute hungry application functions demonstrate a big potential for scaling and use the Cloud Native based platform to deploy various services (Networking, AI, IoT, ORAN etc.) with reduced deployment time down to minutes, from days or weeks what usually took to deploy new dedicated box.

Talking about the overall advantages of the platform we can list several main ones like:

- **Open-source software platform based on OpenNESS and Kubernetes – enables easy Edge Cloud-Native Application deployment and eases of building multi-functional box by running CNF of your choice. Additional Open-source nature of the platform gives more cost-efficiency compared to commercial products on the market**
- **Pre-integrated, optimized and offloaded (by Intel Movidius VPU) application serves as an example of smooth orchestration, deployment, and optimization, especially powered by OpenVINO libraries, Intel Deep Learning Boost and Intel Xeon Scalable processor features.**
- **White-box platform based on COTS is highly reliable edge hardware that reduces TCO while avoiding vendor lock-in over a future proof, versatile edge infrastructure ready to host future smart city services.**

Summarizing, all the benefits above the scalable edge platform outlined in this document creates viable alternative to other commercial products on the market.

Considering the open-source platform deployments, the owner needs to consider also productization, new feature addition or maintenance of the software. Even though open sources communities provide a lot of advantages, quite often the components communities are working on are still to be productized and adopted to the specific use case requirements.

TietoEVERY is excellent partner to help with that process. Considering the company's wide experience in RAN, Cloud and Edge solutions we can Integrate, validate, customize, maintain, and help to deploy End to End Edge solutions whether comprised on only open-source components or by partnering with number of ISVs or OSVs to utilize potentially also commercial applications or platform components. Presented reference architecture creates strong baseline for onboarding variety of workloads with required function mix, including SmartCity (presented in this paper), Telco, Automotive and Enterprise workloads.

Additionally, partnering with technology leaders like Intel and Advantech enables TietoEVERY to build and integrate a best-in-class AI solution to meet our customers' needs. Advantech SKY-8000 Series of 5G Edge Servers enables access to the latest Intel processing technology in highly efficient edge computing platforms designed to accelerate critical workloads in smart city environments.

VIII. Resources

1. <https://www.networkworld.com/article/3224893/what-is-edge-computing-and-how-it-s-changing-the-network.html>
2. <https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html>
3. <https://www.tietoevery.com/en/success-stories/2019/the-city-of-tampere-pedestrian-traffic-safety>
4. www.advantech-5G.com
5. <https://www.intel.com/content/www/us/en/artificial-intelligence/deep-learning-boost.html>
6. https://www.advantech.com/products/3d060f1e-e73e-460d-b38c-c69f76312c91/vega-340/mod_1c16f479-aeaa-4177-867d-ddee9d692fdb

Contact information



Marcin Nicpon,
Telco Ecosystem Solutions Director
TietoEVERY, Product Development Services
marcin.nicpon@tietoevry.com,

About TietoEVERY

TietoEVERY creates digital advantage for businesses and society. We are a leading digital services and software company with local presence and global capabilities. Our Nordic values and heritage steer our success.

Headquartered in Finland, TietoEVERY employs around 24 000 experts globally. The company serves thousands of enterprise and public sector customers in more than 90 countries. TietoEVERY's annual turnover is approximately EUR 3 billion and its shares are listed on the NASDAQ in Helsinki and Stockholm as well as on the Oslo Børs.

www.tietoevry.com