White Paper

Smart Cities & Intelligent Transportation
Converged Edge Solutions

Composable Architectures for a Sustainable Edge

An energy-efficient solution that optimizes the total cost of ownership (TCO) for end-to-end architectures that support edge computing.

Authors:

Francesc Guim,
Marek Piotrowski,
Suman Sehra
Intel®

David Carrera,
Ferran Torredeflot,
Oriol Marí
Nearbycomputing

Judit Bastida,
Raul Gonzalez
Cellnex

Edgar Ramon,
Daniel Pope,
Adrian Whelan
Submer

Alexandru Nastas,
Antonio Bocigas
Lenovo

Cities produce massive amounts of data from clients and IoT devices, which can be overwhelming for legacy networks, centralized datacenters and traditional cloud computing infrastructure. Real time decision making is key, without incurring high network roundtrip cost, or rapidly driving demand for localized data analytics storage and processing that cannot be met by legacy architectures.

Edge computing aims to extend data center and cloud computing to efficiently process, store and navigate relevant data across both data center and IoT models. Forward-thinking leaders predict those key technologies powered by the Internet of Things (IoT), artificial intelligence (AI), and everywhere connectivity can help to build a safe, clean, and inclusive future for smart cities and transport.

Current edge architectures need to support multiple deployment models that contain numerous compute tiers and support different workload types.

As shown in figure 1, edge deployments can start closer to the data source that needs to be processed or served (e.g., camera streams or the VR application user) or, stored in the cloud for less latency sensitive/compute intensive big data use cases.

Tiers	On-Premise Edge: Intelligent Sensor/GW	On-Premise Edge: Intelligent Edge	Network Edge	DC Edge	Public Cloud
Network Latencies (Wire Round trip)	< 1ms	< 1ms	1-5ms	5ms + 1-2 ms (every 100kms)	5ms + 5ms (every 400kms)
Deployment Requirements	**Compute Available Power:** < 50 W **Form Factor:** Small Box **Thermals:** NEBs **Mgmt.:** Remote	**Compute Available Power:** ~10KW **Form Factor:** Rack(s) **Thermals:** NEBS or Standard DC **Mgmt.:** Remote	**Compute Available Power:** <600 W **Form Factor:** Pizza box **Thermals:** NEBS **Mgmt.:** Remote	**Compute Available Power:** 9KW/rack / 1KW sqm **Form Factor:** Rack(s) **Thermals:** NEBS or Standard DC **Mgmt.:** Remote	**Standard Data Center (DC)**

Where, What & Why

Intelligent Sensor/GW Use Case	KPI	Intelligent Edge Use Case	KPI	Network Edge Use Case	KPI	DC Edge Use Case	KPI	Public Cloud Use Case	KPI
Intelligent Transportation	Data Privacy, Backhaul Traffic Savings, Reliability	AR/VR	Latency. Backhaul Traffic Savings, Network scalability	Intelligent Transportation	Data Privacy, Backhaul Traffic Savings, Reliability throughput Lantency	Intelligent Transportation	Data Privacy	CDN	Backhaul traffic savings, Throughput
V2V	Latency	Retail	Data Privacy, Backhaul Traffic Savings, Reliability	V2V	Lantency	Video Analytics	Same as Int. Transp.	Storage GW	Same as CDN
Retail	Same as Int. Transp.	RT Streaming Healthcare	Same as AR/VR Access to services	Video Analytics	Same as Int. Transp.	Drone/IoT	Same as Int. Transp.		
Video Analytics	Same as Int. Transp.			Drone/IoT	Same as Int. Transp.	Healthcare	Access to services		
				Rural	Access to services	CDN & Storage	Backhaul Traffic Savings Throughtput		
						Faas	Latency		
						AR/VR/MR	Latency		

Figure 1. End-to-end system architectures.

End-to-End Edge Orchestration – Nearbyone			
		Network Edge	DC Edge
	Restrictions	**Compute Available Power:** <600 W **Form Factor:** Pizza box **Thermals:** NEBS **Security:** No physical surveillance **Mgmt.:** Remote	**Compute Available Power:** 9kw/Rack - 1kw Sqm **Form Factor:** Rack(S) **Thermals:** Nebs Or Standard Dc **Security:** Physical An No-Physical **Mgmt.:** Remote
Use Case 1: V2X/V2V	Workloads	Road Safety Analytics 5G and Core	Advanced Video Analytics Orchestration Control Plan
Use Case 1: V2X/V2V	Chasis and Systems Arch.	Zero Emision Site Cabinet (Cellnex) 1u System Fan Cooling (Lenovo Se340)	2u System (Lenovo Sr650) Micropod (Submer)
Use Case 1: V2X/V2V	Compute Tecnology	Intel® Xeon® D Processor	Iintel® Xeon® Scalable Processor Gen 3 Intel® Movidius™ S Vpu-Based Vision Accelerator Card
Use Case 2: Intelligent Transportation	Workloads	Intersection Traffic Management	Time Series Data Base Orchestration Control Plan
Use Case 2: Intelligent Transportation	Chasis and Systems Arch.	City Cabinet 2u System Fan Cooling (Lenovo Se450)	2u System (Lenovo Sr650) Micropod (Submer)
Use Case 2: Intelligent Transportation	Compute Tecnology	Intel® Xeon® Scalable Processor Gen 3 Intel® Movidius™ S Vpu-Based Vision Accelerator Card	Intel® Xeon® Scalable Processor

Figure 2. Deployment setup.

Mapping use cases to specific edge locations depends on:

1. Use case requirements (e.g., reliability or data privacy)
2. Deployment requirements (e.g., the number of camera streams to be processed)
3. Customer requirements (e.g., data sovereignty and security, where the compute needs to be placed).

These deployment models require architecture that can provide high service density close to the data source; perform adaptive and proactive service life cycle management, depending on the unique requirements at the edge; operate without backhaul connectivity; and with power efficiency to be sustainable.

In this white paper, we present an architecture created in collaboration with Intel's industry partners Cellnex Telecom, Submer, Nearby Computing, and Lenovo, that addresses the above requirements. Key features of this platform architecture include:

Advanced Compute Technologies:

- Intel® Xeon® Scalable Processors
- Intel 3rd gen Intel® Movidius™ S VPU Based Vision Accelerator Card
- Lenovo ThinkSystem SE450 and SE350 edge servers

Sustainable Power Sources:

- Edge systems powered by Cellnex Zero Emissions renewable energy
- Submer Immersion cooling technologies that allow higher service density with reduced power consumption

A Scalable Multi-Tenant Solution Stack with Automated Manageability:

- NearbyOne end-to-end orchestration solution
- Multiple services can run simultaneously and independently reducing the total cost of ownership (TCO)
- Automated deployments and operation with simplified services deployments

High Accuracy Geolocation:

- Albora's AlbaSpot services for high-accuracy geolocation augmentation
- Software-based geolocation technology provide cm-level accuracy
- Standard protocols and APIs allow interoperability and hardware sensor agnosticism

Use cases

This white paper will outline the joint end-to-end system architecture created by the partners involved to demonstrate how the technologies support sustainable, dynamic, and service dense edge deployments.

Figure 2 sets out the two deployment models and use cases that are included as part of this white paper, as well as the different technologies and workload mappings. The next section will detail the interplay between technologies in the deployment model. The architecture aims to satisfy edge services requirements in terms of:

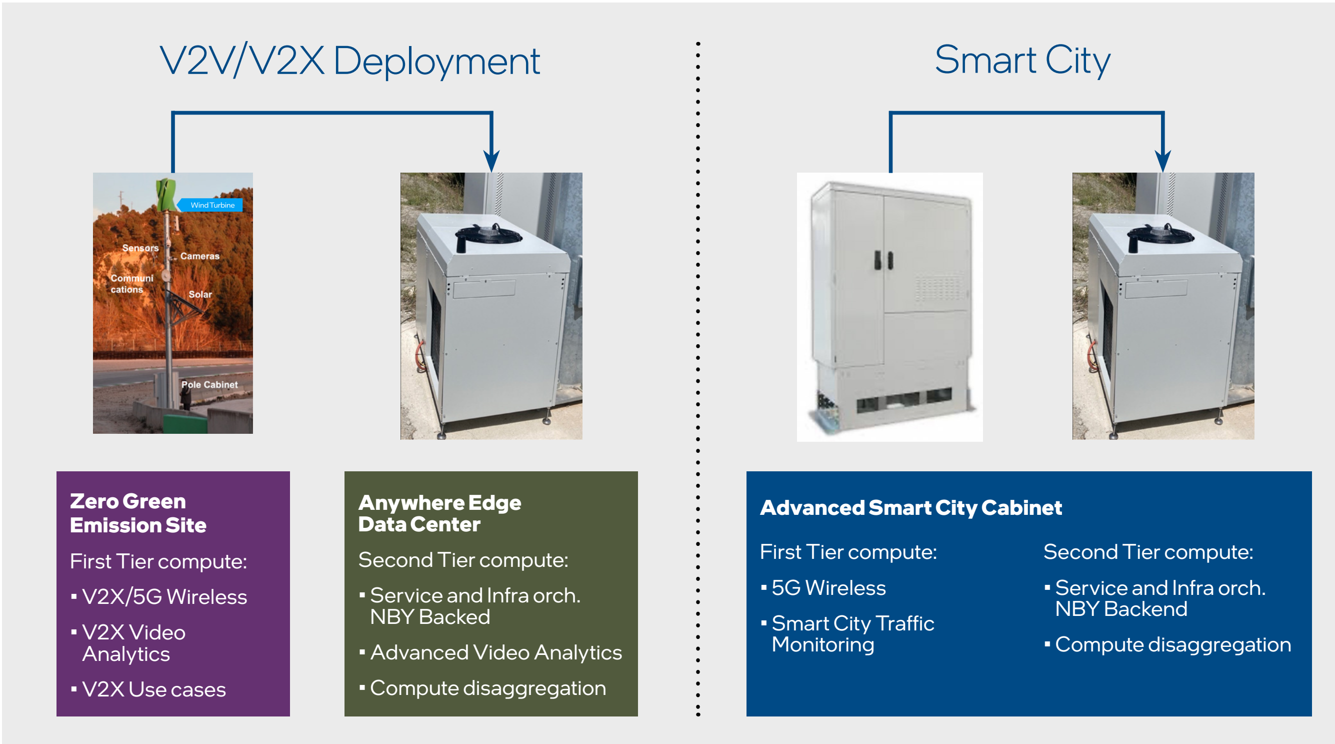

Figure 3. Use case mapping and deployment architectures.

1. Priority (performance) and QoS – For example, traffic for a connected vehicle (CV) application will take higher priority than a temperature sensor in terms of response time.
2. Reliability and resiliency – Some input streams need to be acted upon and the traffic routed with mission-critical reliability, whereas with other input streams, it may be tolerable to accept an occasional failure—it depends on the application.
3. Thermal, electrical, mechanical, and industrial design constraints.

Use case implementation

This section will describe the two key use cases in the white paper and their reliance on the various tier and respective technologies. In both use cases, the workloads that run in each location are organized in a holistic, end-to-end way by the NearbyOne orchestrator, provided by Nearby Computing. NearbyOne is responsible for life cycle management and provisioning of the micro-services that support the use cases. This includes migrating services from one tier to another when required - for example, when the planner on the Zero Green Emission site detects potential problems with battery duration.

Use case 1: V2X and V2V

The Zero Green Emission Site use case demonstrates the implementation of scalable architecture that is both wireless, and energy efficient. This is achieved by deploying high energy, efficient architecture in the first tier of the network using the Zero Emission Site together with Intel compute (Intel® Xeon® D processor) in a Lenovo edge chassis (SE350), optimized for low power consumption and advanced power management capabilities. In this first layer of compute, two different workloads are hosted: 5G connectivity and video analytics for transportation safety and high-accuracy geolocation augmentation for accurate vehicle positioning and tracking, provided by Albora services. The later workloads are deployed in the first tier to provide low latency and low footprint AI analytics to identify potential objects on the road. The later workloads are deployed in the first tier to provide low latency and low footprint AI analytics to identify potential objects on the road. These workloads use simple AI pipelines to reduce the amount of power consumption and allowing batteries last during the whole day.

This first tier of compute relies on the second tier of compute that is based on edge immersion liquid based micro-data center that can be deployed anywhere. This second tier, connected to the power grid, is responsible for performing lifecycle management of all the various Zero Emission sites at the first tier. It also provides extra compute capacity to perform more advanced video analytics workloads and navigation data analytics.

These workloads are defined with more complex AI pipelines that require acceleration to achieve higher stream density. This is achieved by using Intel's edge AI accelerator, the 3rd generation Intel® Movidius™ S VPU based Vision Accelerator Card.

Use case 2: Smart City and Intelligent Transportation

The second use case described in the white paper aims to show case how the same type of technologies can be used for smart city deployments, including intelligent transportation use cases such as intersection traffic management. In these scenarios, the first level of compute is hosted in a roadside edge cabinet that is connected to the grid. Therefore, power availability is not a problem and compute density and complexity of the workloads executed in this first tier can be increased by deploying more powerful solutions. In this use case, the first tier includes high-end compute (Intel® Xeon® Scalable processor) platforms house in a new chassis (Lenovo SE450) design aiming these deployment models together with the 3rd generation Intel® Movidius™ S VPU based Vision Accelerator Card.

The workloads deployed in this first use are advanced video analytics pipelines meant to perform traffic AI analytics on many camera streams in order to implement intelligent transportation/road infrastructure use cases such as traffic management and safety.

Like the previous use cases, the architecture proposes to use liquid-based micro data centers, hosted anywhere, that are responsible for the life-cycle management of each of those roadside edge cabinets as well as hosting data store microservices that are responsible for recording and correlating events.

Figure 4. Zero Emission Site.

Technologies of the architecture

This section describes all the various technologies that have been used as part of the architecture that was described in the previous section. The goal is to describe their major characteristics and the benefits that they provide.

Zero Emission Green Site

The Zero Emission Green Site is developed to acquire the lowest possible energy consumption. This sustainable node harnesses its energy from solar panels and wind turbines, as shown in figure 4 on the next page. Specific algorithms allow the site to manage energy efficiently while displaying the consumption of each connected device to the node (camera, RU, compute servers etc.).

Every node is composed of radios, IoT sensors, cameras, and backhaul antennas. Each of these components are connected through switches, and servers in the edge cabinet. The batteries stored in the cabinet also support those mechanisms which predict energy behaviour, and optimise the energy consumption of connected devices. This means that the entire system is optimised, and can survive 24 hours adapting to the energy it has, and can expect, to harvest.

The Zero Emission Green Site provides a new vision for telecommunication towers. With this type of sustainable technology, we can bring 5G connectivity to rural areas, last mile deployments, and zones without fiber connection. Further evolution of this sustainable technology will benefit both society and nature: it can be deployed easily in a cost -effective manner.

Submer: Submer immersion cooling solution for edge environments

Submer is an industry leader in immersion cooling technology. The MicroPod is an edge-ready, plug-and-play, datacenter-in-a-box solution that delivers high density and efficient infrastructure, anywhere from factory floors to the heart of a city, to remote areas with harsh climate conditions. Submer technology, and its ability to extract energy generated as high-grade temperature liquid, enables new business models such as “Heat Reuse” and circular applications aligned with environmental, social, and governance policies.

Orchestration and SW architecture: Nearby One and Smart Edge Open platform

The ecosystem of technologies enabling the edge computing paradigm is broad and covers many aspects including hardware components, service orchestration, standards and protocols, applications, virtualization technologies, and others. The diversity of these technologies can pose a complex challenge for edge providers, trying to select the correct combination of solutions.
The Smart Edge Open platform helps to alleviate this challenge and is based on three fundamental aspects combined to create a unique end-to-end solution:

1. It uses a cloud native based approach. Therefore, use cases can be developed on top of common microservices that allow better reuse, lower system utilization and reduced time to market.

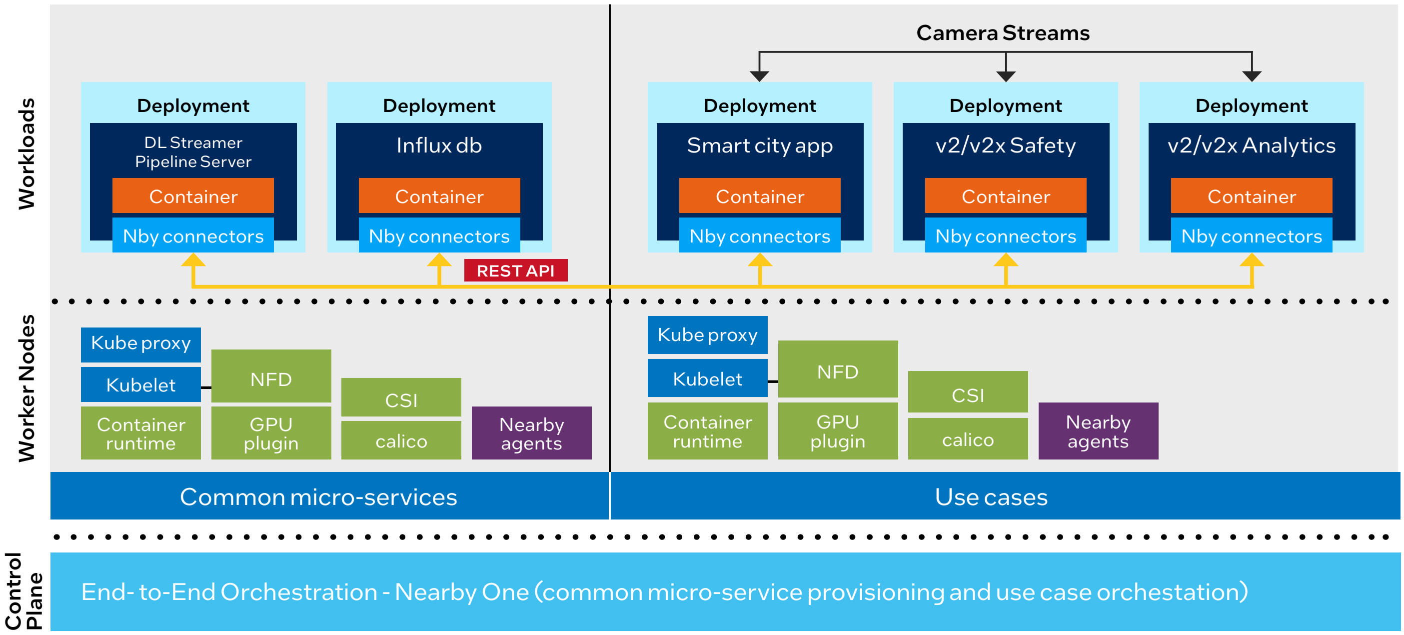

Figure 5. Microservice-based platform.

2. It is based on standards and interoperable components to enable integrating, testing and validating different technologies from different vendors. This enables a rich and competitive composable architecture that is not coupled to a single technology/vendor, and can be scaled based on customer requirements.
3. It unifies different verticals into a single edge architecture (Telecommunications, Internet of Things, Enterprise and Smart Cities/Transportation) which traditionally have been implemented with different technologies and unique architectures. Thanks to this convergence, the overall system can be managed in a more efficient manner, drastically reducing the total cost of ownership.
4. It exposes a set of resources, software development kits and application programming interfaces which are easy to develop, test and deploy applications optimized and designed for the edge.

The architecture, as shown in figure 5 on the next page, supports a set of micro-services that are shared across the various use cases and applications deployed in an edge environment. For example, the Intel DL Streamer Pipeline Server allows registering AI pipelines, that can be executed transparently, to multiple different hardware elements (e.g., Intel® processors, 3rd gen Intel® Movidius™ S VPU Based Vision Accelerator Cards, etc.) using the Intel® Distribution of OpenVINO™ Toolkit.

Nearby One was used in this deployment as the edge orchestration and automation engine to manage the deployment of services and perform end-to-end service assurance. This solution uses a subset of Intel Smart Edge Open, Intel® Deep Learning Streamer Pipeline Server and OpenVINO™ building blocks to realize the service and infrastructure life cycle orchestration. It is composed of:

- The Nearby Orchestration Platform is a centralized controller, providing the logic and control loops for all tasks related to the orchestration of the applications, network and compute infrastructure.
- The Nearby Blocks are distributed components that encapsulate logic and code for different application-specific functionalities to extract application KPIs and expose them to the closed control loops implemented in the orchestration platform.

The Smart Cities and Intelligent Transportation apps and the V2X applications were onboarded, associated with the right placement policies and pushed into Nearby One application marketplace. Furthermore, the common micro-services that are used by V2X applications are deployed as dynamic dependencies and connected to the new use cases using Nearby Connectors. Through the Service Designer, they were deployed to the right locations and parametrized accordingly. The edge orchestration engine resolved the right cameras to be connected to the applications, based on the infrastructure available on-site, and dynamically rendered the configurations at runtime.

The video analytics application used in this PoC was powered by the Intel® Distribution of OpenVINO™ toolkit and Intel® DL Streamer Pipeline Server and was onboarded and encapsulated as a Nearby Block which contained the application logic, as well as a set of auxiliary components that provided the means for the application to be effectively managed, including:

- Application performance KPIs for continuous SLA assessment (not only network-centric metrics).
- Application health/status KPIs for continuous lifecycle monitoring and management.

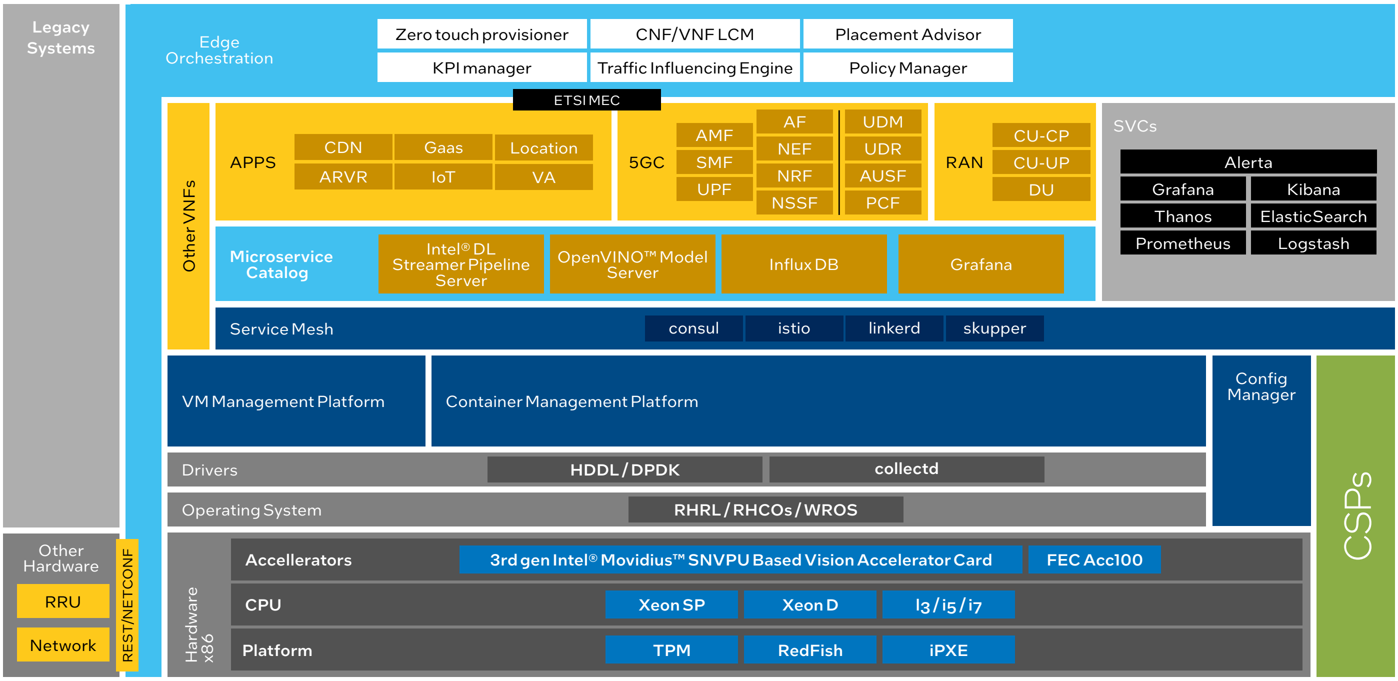

Figure 6. The Nearby One platform.

- Application capabilities dynamically exposed at deployment time.
- Correlation of application KPIs and platform (processor, accelerators, memory, storage) for more effective edge platform management.

The information about the platform and the software components is processed by Nearby One, injecting KPIs and telemetry into the automation control loops of the platform. Therefore, the SLAs of the different applications can be continuously observed and operated to ensure the desired service levels.

Nearby One has bare metal as-a-service capabilities, split between a bare metal descriptor and a set of ansible roles to configure the software stack. In this context, the necessary configurations to enable the use of the 3rd gen Intel® Movidius™ S VPU Based Vision Accelerator Card were enforced to enable the capabilities of the AI accelerator. It was used afterwards by the containers of the video analytics application deployed.

The solution was deployed following a full model, where all the components ran in the Zero Emissions Site and roadside edge Street cabinet. The Nearby Controller was deployed in the MicroPOD and connected by VPN.

Edge appliances: Lenovo Edge platforms with Intel compute technologies

As part of its Edge Infrastructure solutions, it has come up with a rich Edge Computing portfolio accommodating different locations, form-factors, and use-cases.

This paper will focus on two servers, designed to accommodate difficult near and far-edge environments.

- The ThinkEdge SE450, a 1-socket, 2U height, 3rd Gen Intel® Xeon® Platinum processor-based server
- The ThinkSystem SE350, an Intel® Xeon® D-2100 based1U height, half-width, and short-depth Edge server

ThinkEdge SE450

ThinkSystem SE350

Both these servers can be deployed in various deployment modes: hung on a wall, stacked on a shelf, or mounted in a rack. They are ruggedized, can handle continuous operating temperatures from 5°C and 45°C; and some configurations are designed to meet NEBS Level-3 and ETSI requirements for 96 hours operating excursions from -5°C and 55°C as well as tolerance to locations with high-dust and vibration.

Lenovo's innovation makes the **ThinkEdge** SE450 one of the most powerful, yet acoustically quiet servers on the market, allowing it to be deployed next to end users without disturbing the workplace. In Smart City use case, it's being deployed with the 3rd gen Intel® Movidius™ S VPU Based Vision Accelerator Card.

Most general-purpose GPUs provide inferencing but when needed alongside media processing they take a huge portion of the resources in the host device. Intel's VPU with integrated sparsity technology simultaneously accelerates inference and media processing for industry leading end-to-end performance and power efficiency, freeing host resources for other value-added tasks. It is integrated into an energy efficient 75W low profile (HHHL) PCIe accelerator card form factor.

For the Far Edge locations, where the form-factor is crucial, Lenovo has developed the **ThinkSystem** SE350. With its small form-factor (half-width, 1U height).

Both servers have been specifically designed to run edge applications and provide processing power, storage, workload acceleration and networking techniques required for today's edge workloads. From a security standpoint they are equipped with key-encrypted storage, secure BIOS and physical capabilities such as a locking bezel, intrusion, and tamper protection mechanisms.

Lenovo also provides software solutions for Enterprises and Cloud Providers that dramatically accelerate deployment of virtualized cloud infrastructure for both Information Technology (IT) and Operational Technology (OT).

Conclusions and next steps

This work demonstrated the feasibility of implementing sustainable distributed converged edge reference architecture, based on the principles described earlier in this whitepaper allowing for the seamless convergence of virtualized services and VNFs on general-purpose Intel® Xeon® Scalable Processor processor-based edge servers and Intel vision accelerators. Alongside an advanced orchestration engine that can coordinate the actions of VNFs and edge applications.

The architecture, deployed at Cellnex, comprehends end-to-end system design that includes multiple advanced technologies that are orchestrated together to provide a advanced system design: (1) Compute and AI Acceleration that provide flexibility, high service density and scalability; (2) Zero emission sites that allows re-use of solar energy allowing high deployment flexibility and sustainability; (3) Immersion cooling technologies that deploy high end compute resources anywhere at the edge; (3) End-to-end orchestration that provides multi-tier service and infrastructure orchestration.

Adopting this approach will lead to business value realization at the heart of its strategy to use the technology as an enabler to offer new digital services, leveraging advanced computer platforms and new immersion cooling schemes that democratize edge deployments. The deployment utilizes each of the components above which enables new business models and allows to create scalable and sustainable end-to-end.

Today, the technology is deployed in Castellolí Parc[1], and will also be used in various other edge use cases. Sustainability is a key factor when we talk about Cellnex's growth—their vision is not only to offer better performance and support innovative projects—but also ensure that they support efficient resource use. Being able to use such powerful architecture in our zero emission site, and demonstrate the smart city and V2X use cases, exemplifies how restrictions are just challenges to overcome.

Learn More

- **Cellnex**
- **Submer**
- **Nearbycomputing**
- **Lenovo**
- **Intel® Smart Edge Open**
- **Intel® Distribution of OpenVINO™ Toolkit**
- **Intel® Movidius™ Vision Processing Units (VPUs)**

Find the solution that is right for your organization. Contact your Intel representative.

[1] Source: https://www.cellnextelecom.com/en/sustainable-management/

0322/JW/CAT/PDF ♻ Please Recycle 350427-001EN.