Solution Brief

Intelligent Edge Solution

ENSCONCE for Retail Solutions

The Smart Retail solution from Capgemini Engineering integrates key Intel® hardware and software ingredients to deliver a scalable visual computing edge platform optimized for emerging retail solutions.

Capgemini engineering

Abstract

Visual computing in retail is a hot topic. While visual computing is already finding success in other industries like factory automation, retailers are just beginning to experiment with the full potential of this technology. The combination of visual computing and artificial intelligence will allow video cameras in retail to go from simple surveillance devices that protect against loss to revenue enhancing assets that enable Grab-and-Go payment systems, real-time inventory tracking, and customer behavior analysis to drive improvements to store layouts, product displays, and the overall customer experience.

To meet this growing demand for edge computing capabilities across industries, Capgemini Engineering has developed the Capgemini ENSCONCE edge platform. ENSCONCE is a scalable compute platform that can interface with different networks, from wired Ethernet to Wi-Fi to 5G, that allows cloud-compute capabilities to be located at the edge of the network. By locating compute resources closer to where the data is generated, bandwidth costs are reduced along with latencies, allowing for real-time results based on visual input. By combining scalable computing resources, support for multiple types of networks, and standards-based application orchestration, ENSCONCE provides the foundation for a wide variety of solutions spanning from manufacturing to transportation to government applications.

Capgemini ENSCONCE is built on Intel® Smart Edge Open, which leverages industry-standard containerization and orchestration implementations such as Kubernetes to allow developers to rapidly deploy and scale applications on one or many ENSCONCE platforms. To address visual computing needs, the ENSCONCE platform supports industry-standard Open Visual Cloud Stack and OpenVINO™ toolkits, along with the Visual Cloud Accelerator Card – Analytics (VCAC-A) to deliver hardware-accelerated visual computing to enable high-performance visual computing solutions. ENSCONCE is optimized for the Intel® Xeon® Scalable processor family, enabling a wide range of performance scalability to meet the exact needs of each application.

This paper will review these key ingredients and how they enable Capgemini ENSCONCE edge platform to deliver a comprehensive, scalable visual computing solution for retail.

Latency and Bandwidth Drive the Need for Edge Computing

With the proliferation of devices within a business, there is an increasing need for reliable, cost-effective, low-latency communication between the devices generating data and the application processing that data. Smart factories, automated warehouses, quality and flow control systems all combine the need for fast processing of large amounts of data. Many of these solutions must operate in real-time, requiring tight controls on transmission and computing time. Processing

data in remote data centers can add enough transmission time to break real-time capabilities. As data requirements are increasing with visual computing, bandwidth costs also rise.

While the move to 5G networks for device connectivity increases system flexibility, solutions still need to deal with network latency and bandwidth costs. Capgemini ENSCONCE edge platform addresses these concerns by allowing developers to place compute resources close to where the data is generated, which reduces latency and bandwidth costs. ENSCONCE platform can integrate with multiple networks, from wired Ethernet to 5G, allowing for the optimal combination of data interfaces for a wide range of solutions.

Retail Visual Computing Opportunities

The ENSCONCE edge platform with Intel® Smart Edge Open is well-suited to serve retail and smart city use-cases. The combination of multiple cameras installed throughout a store and visual computing/artificial intelligence applications offers retailers several compelling solutions to improve operations and sales:

- Shelf tracking to optimize item positioning and improve store layouts
- Customer tracking to understand customer behavior and enhance customer engagement
- Heatmaps at most trafficked areas to gain insights
- Checkout queue management to redirect customers to less loaded checkout counters
- Threat identification to help ensure product, customer, and employee safety
- Dashboard analytics to act on insights

ENSCONCE Platform: Architected for Edge Computing

ENSCONCE is a platform-as-a-service solution for deploying latency-sensitive applications at the edge of the network. The solution has been designed based on research with application developers, hardware providers, network operators, and enterprises. It is based on industry standards to allow for flexibility, scalability, and reusability. The ENSCONCE solution has been recently chosen as the finalist in the Leading Lights Award 2020 in the category of "most innovative edge computing strategy". Along with ENSCONCE, Capgemini also brings a rich ecosystem of network equipment and cloud partners, application developer ecosystem and silicon partners, including active participation in GSMA Operator platform, Linux Foundation (LF Edge), TIP and Open Edge Computing.

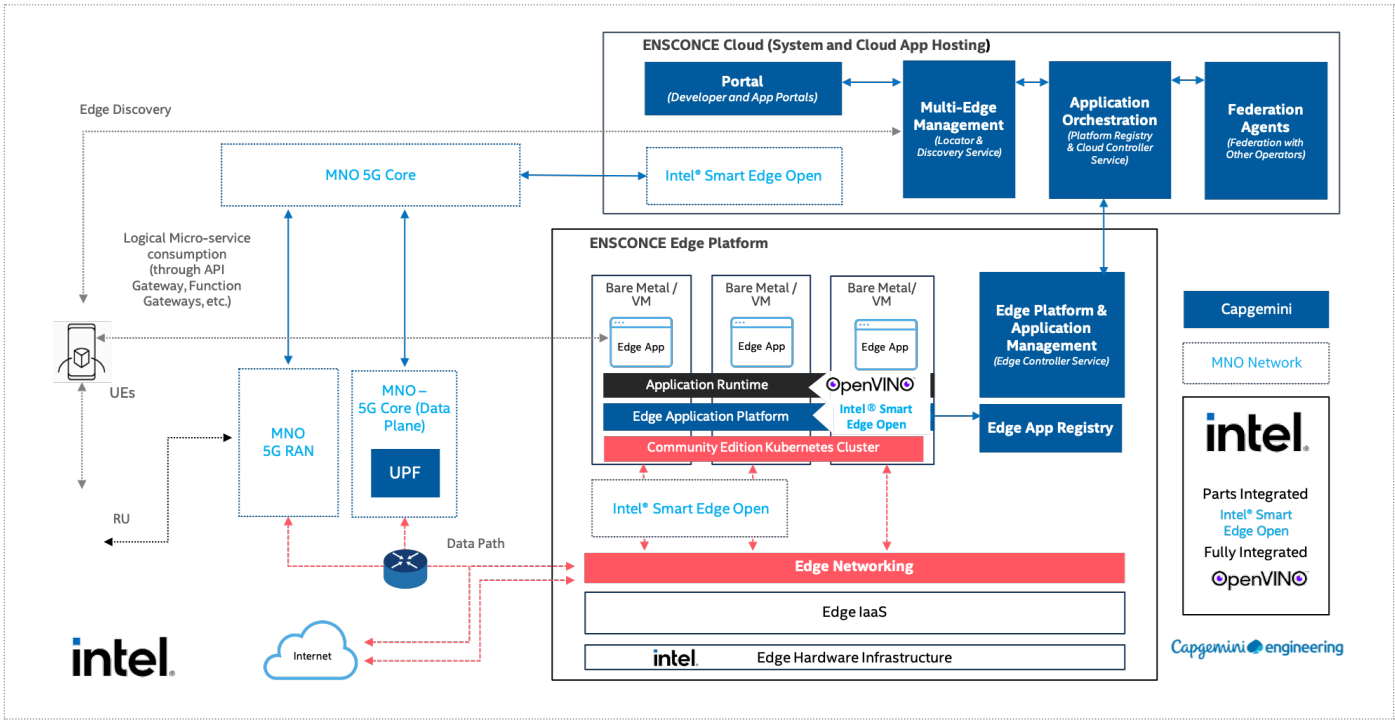

Figure 1. ENSCONCE Cloud and Edge Platform

3rd Gen Intel® Xeon® Scalable Processors

When an ENSCONCE platform is located at the network edge, compute resources are moved close to the data. Being based on Intel Xeon Scalable processor family, the platform can be designed to address the compute requirements for each specific installation, from a compact single processor solution to a rack with dozens of processors. Intel Xeon Scalable processors also include Intel® Deep Learning Boost (Intel® DL Boost) to accelerate AI algorithms. Regardless of whether the retail location is small or large, the ENSCONCE platform can be designed to deliver the optimal compute power.

1 CPU
8 cores total

8 CPUs
244 cores total

Visual Cloud Accelerator Card for Analytics (VCAC-A)

When processing large amounts of video in real time, ENSCONCE supports the use of hardware acceleration cards to further increase performance. The Visual Cloud Accelerator Card for Analytics15 is a high density, cost-optimized acceleration solution for use in network edge servers. Powered by both an Intel® Core™ processor and an Intel® Movidius™ Vision Processing Unit (VPU), VCAC-A includes video encode-decode functions as well as machine learning inference engines to enable a high-performance, power-efficient, cost-effective media analytics solution.

VCAC-A runs a stand-alone operating system, based on Ubuntu 18 LTS, and is managed by Kubernetes as a standalone node, rather than as a peripheral. Kubernetes can then schedule workloads directly to the VCAC-A.

Intel® Smart Edge Open

Intel Smart Edge Open is a software toolkit that enables scalable edge platforms to manage and orchestrate applications and network functions with cloud-like agility across any type of network. It is built on standardized APIs and open source software tools and enables solutions based on the ENSCONCE platform to:

- Easily migrate applications from the cloud to the edge while abstracting network complexity.
- Deliver more secure on-boarding and management of applications with an intuitive web-based GUI.
- Leverage standards-based building blocks for functions such as access termination, traffic steering, multi-tenancy, authentication, telemetry, and appliance discovery and control.

Open Visual Cloud and OpenVINO toolkit

Open Visual Cloud is a set of open source software stacks for visual computing, including transcoding and analytics. Open Visual Cloud software stacks are delivered as ready-to-use Docker images for easy deployment. Images have been validated for deployment on hardware platforms based on Intel® architecture processors.

ENSCONCE platform leverages the OpenVINO toolkit for standardized algorithms to deploy inference-based applications, such as computer vision, speech recognition, and natural language processing. It supports the latest innovations in neural networks, including convolutional neural networks, and recurrent and attention-based networks. The toolkit provides APIs that are supported across Intel® CPUs, GPUs, FPGAs and VPUs, enabling solutions to easily scale performance as required.

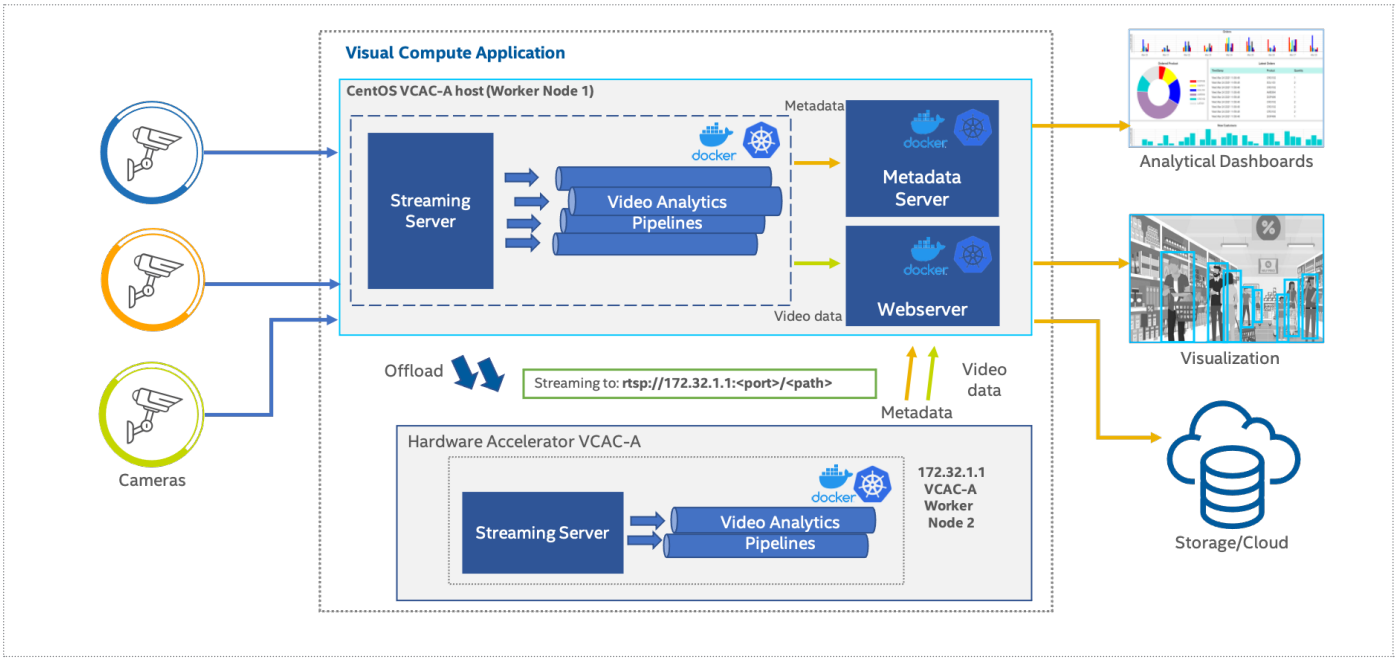

Figure 2. Visual Compute Application Architecture

Smart Retail Visual Computing Solution

To address the unique requirements of the retail sector, Capgemini has developed the Smart Retail visual computing solution. The Smart Retail solution accommodates video feeds from multiple cameras positioned throughout a store to monitor shelves, aisles, and checkout counters, and uses the ENSCONCE platform to receive and process those video streams to address specific retail needs. For example, the shelf camera feeds are used for inventory tracking whereas the aisle camera feeds are used for proximity detection and customer behavior analysis. The combination of visual computing and machine learning applied to these video feeds allows for levels of information and automation that are unprecedented in the retail sector.

The Smart Retail solution is based on a Visual Compute architecture integrated into ENSCONCE.

Streaming Server: The streaming server accepts connections from devices streaming video data to the server using RTSP protocol. This handles all requests coming from the clients and is also responsible for invoking video analytics pipelines on the Intel Xeon Scalable processor. For platforms using hardware acceleration for high performance, the client streams are sent to the VCAC-A node, which then allocates video streams between the CPU and VPU for the most efficient processing. By offloading video processing to the VPU, this frees up CPU resources to focus on packet processing and other network access functions to help support consistent, reliable connections.

Video Analytics Pipelines: The processing of incoming video frames is carried out by the video analytics pipelines, where operations such as decoding, pre-processing, and inference are performed on the video frames. Video analytics plugins from the OpenVINO toolkit are used to carry out different inference functions, including detection and classification. Camera feeds received from different locations in the retail store are processed in parallel, allowing for real-time analytics combining data from multiple cameras.

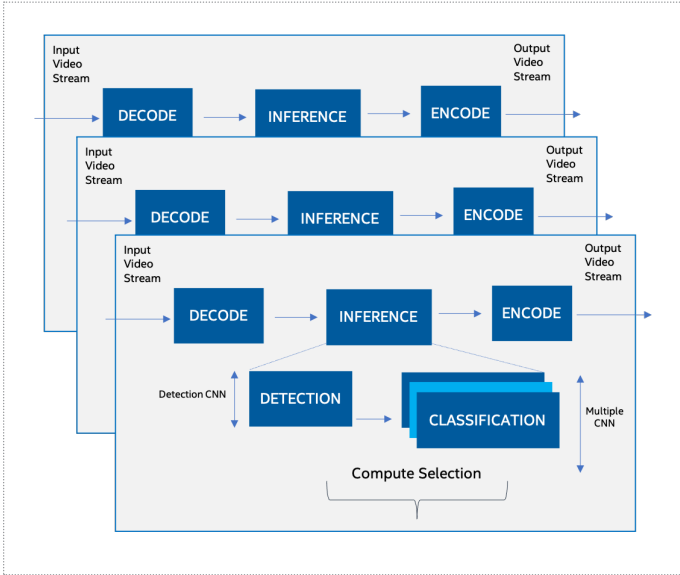

Figure 3. Video Analytics Pipelines

Metadata server: Metadata derived from the video streams is published to analytical dashboards and allows display of real-time statistics from the video streams. The metadata generated can additionally be used to trigger actions and alerts. For example, the output from shelf analytics can be used to trigger an alert to restock the inventory.

Webserver: The webserver provides a backend for visualizing annotated video streams in a web browser, for example by an administrator for monitoring the surveillance camera feeds. If required, post-processed video streams can be picked from the webserver for storage and offline analytics to the central cloud.

Performance Benefits

The Visual Compute architecture on the ENSCONCE edge platform is designed to maximize visual processing at the edge. Pre-processing of video streams is performed on the VCAC-A accelerator card. Intel Xeon Scalable processor resources are utilized for decoding and post-processing of the video streams, while inference is offloaded to the VPU. With this hardware accelerated design, video analytics consumes minimal host CPU resources, freeing up Intel Xeon Scalable processor cores on the ENSCONCE platform to perform other application processing.

As an example, a 500-square-foot retail store may require 20 cameras to track customers and inventory. By utilizing VCAC-A hardware accelerators, the ENSCONCE platform can double the number of video streams that can be processed while maintaining application compatibility. By combining standards-based computing with hardware acceleration, this Visual Compute architecture enables ENSCONCE platform to offer a high-degree of scalability for visual computing solutions.

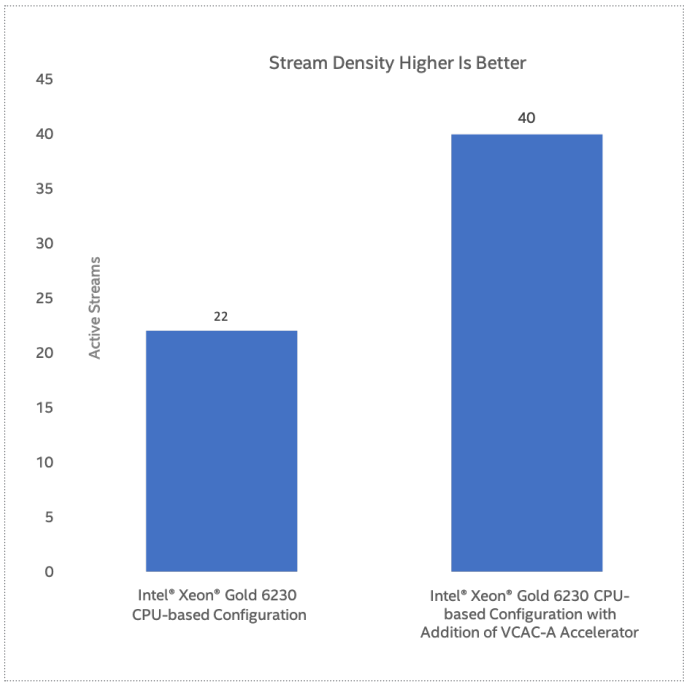

Figure 4. Performance Benefits of Intel® Xeon® Gold 6230 CPU-based Configuration with Addition of VCAC-A Accelerator[1]

Conclusion

ENSCONCE Smart Retail solution delivers an end-to-end visual computing edge platform, enabling new capabilities such as automated inventory tracking and customer counting. The visual compute architecture integrated into ENSCONCE platform delivers powerful visual computing capabilities to developers, enabling innovative solutions such as Smart Retail across a wide range of industries. If you'd like to learn how ENSCONCE platform and its Visual Compute architecture can help address your needs, please visit here.

[1] Testing done by Capgemini Engineering in December 2020. ENSCONCE Edge platform configuration included two Intel® Xeon® Gold 6230 processors (microcode 0x5002f00) running at 2.1 GHz. Intel® Hyper-Threading and Intel® Turbo Boost Technology were both enabled. BIOS version: 2.6.4. The platform included 256 GB DRAM running at 2933 MHz and Intel® Ethernet Server Adapter I350 with quad gigabit Ethernet ports. The platform software stack was based on Ubuntu 18.04.5 LTS. OpenVINO toolkit 2021.1.110 and DLstreamer, Gstreamer, and Python libraries were also used. VCAC-A accelerator card included Intel® Core™ i3-7100U processor, 8GB DDR4 DRAM, twelve Intel Movidius™ Myriad™ X VPUs. VCAC-A host app accelerator configuration also ran Ubuntu 18.04.5 LTS, OpenVINO 2021.1.110, and DLstreamer, Gstreamer, and Python libraries.

Performance varies by use, configuration and other factors. Learn more at www.Intel.com/PerformanceIndex.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

Intel is committed to respecting human rights and avoiding complicity in human rights abuses. See Intel's Global Human Rights Principles. Intel's products and software are intended only to be used in applications that do not cause or contribute to a violation of an internationally recognized human right.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

0721/FZ/ICKK/PDF 347519-001 EN