Solution Brief

Smart City, Critical Infrastructure

NEXCOM Servers Provide Edge Video AI Analytics and Processing

Intel® Edge Video Infrastructure reference design provides smart city edge video processing running on NEXCOM NSA 7160R; tests of the solution show it meets Intel® EVI performance metrics for AI and storage performance

Communities around the world are turning to technology and wireless networks for smart city projects meant to enhance the quality of life for citizens. By integrating smart city solutions with existing physical infrastructure, urban leaders are creating a positive impact on various aspects of city life, including safety, affordability, traffic, environment, health, and connectivity.

A wide range of advanced technologies play a role in effective smart city projects (see Figure 1) including edge computing, cloud computing, IoT/sensing, artificial intelligence (AI), and pervasive computing.

One of the foreseeable challenges in implementing smart city projects is managing the substantial growth in visual data traffic and the necessity to process this data at network edge locations.

To process the vast increase in video data traffic that has resulted in much higher compute workloads, smart city solution operators are embracing edge server solutions instead of, or in addition to, using the cloud. These servers are outside of a data center and are located right where data is generated. So, they need the capability and capacity to locally store, manage, structure and analyze video data. All of this processing is best done on the edge because it reduces latency and doesn't overload backbone data networks.

In addition to a high performance hardware foundation, these servers need software that provides containerization, edge networking services, video core services and the video applications that provide the requested services.

To simplify new edge video server solution creation and help speed the time to market, Intel has developed the **Intel® Edge Video Infrastructure (Intel® EVI)** reference design that provides hardware specifications combined with a complete software stack of all the foundational edge video services.

Intel® Network Builders Community partner, NEXCOM has tested the Intel EVI 2.0 software (that is included in the Intel EVI reference design) on its NSA 7160R, a powerful three-in-one server equipped with dual 4th Gen Intel® Xeon® Scalable processors for high performance video processing and AI inference, high-bandwidth LAN modules, and a high-capacity NVMe storage module. As the tests in this paper will show, the NEXCOM NSA 7160R with the Intel EVI reference design creates a system that is capable of efficiently processing edge video server workloads.

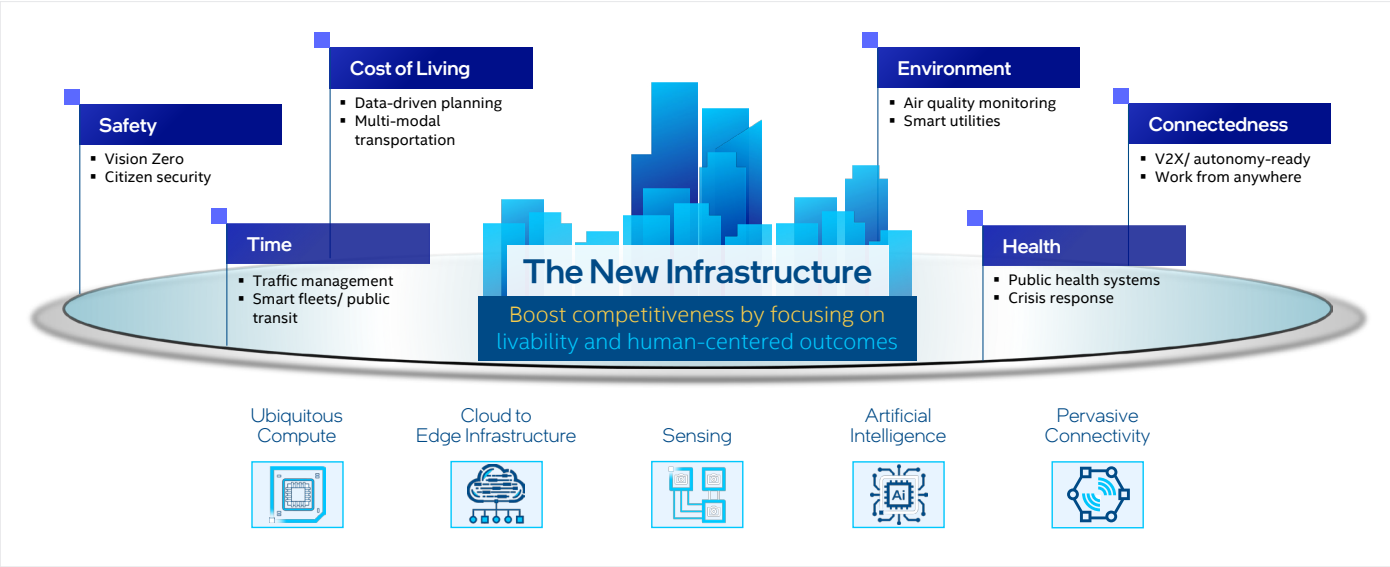

Figure 1. Smart city projects address urban living challenges using networking and technology.

Intel Edge Video Infrastructure

The Intel EVI reference solution outlines all the hardware and software components required for high-performance server video processing, specifically including:

- **Hardware platform:** based on Intel® Xeon® Scalable processors to enable compatibility and software portability to facilitate a variety of server form factors that fit various edge locations.
- **Cloud-native containerization:** The reference design specifies Kubernetes with Intel® Smart Edge Open edge services for seamless deployment of containerized microservices in edge servers, clusters, or the cloud for scalability and flexibility.
- **Core video services:** Provide video processing, storage, AI inference, feature matching, clustering and archiving services necessary for third-party video applications.

Figure 2 shows how the Intel EVI reference design comes together.

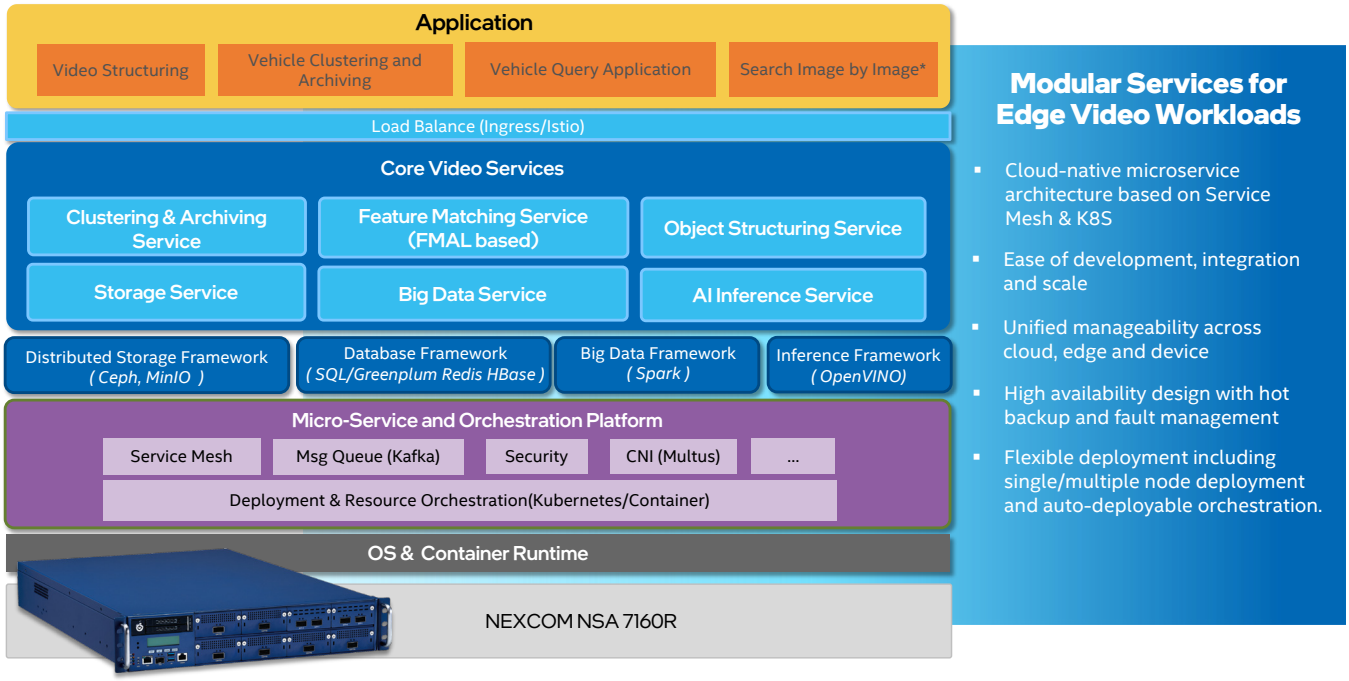

Figure 2. Intel Edge Video Infrastructure block diagram.

The hardware foundation used for the testing is the NEXCOM NSA 7160R server family based on 4th Gen Intel Xeon Scalable processors. These processors offer scalability and performance for compute, storage, network, and security applications. They also feature a well-balanced and efficient architecture that increases core performance, memory, and I/O bandwidth to accelerate diverse workloads from the data center to the network edge.

The processor used in the testing was the Intel Xeon Gold 6430 with 32 high-performance compute cores and up to four built-in accelerators, including the Advanced Matrix Extensions (AMX), which speeds processing of the AI inference workloads that are vital to the Intel EVI reference design.

The operating system and Kubernetes container runtime are foundational layers in the Intel EVI stack and interact with edge services provided by Intel Smart Edge Open which makes it easy, resilient and secure to deploy and manage container-based workloads at the edge.

Intel Smart Edge Open services are integrated with a Kubernetes distribution and features multiple standalone building blocks that provide specific functionality within the overall architecture, including service mesh, network functions, security, observability and storage.

For the core video services, the AI inferencing software is Intel-optimized OpenVINO™ Toolkit. This software is designed to accelerate the inferencing and development of machine learning solutions including video processing and AI analytics. The OpenVINO toolkit seamlessly works with popular AI frameworks, including TensorFlow and PyTorch, and offers a variety of model compression techniques to optimize inference performance along with a common API to enable deployment.

NEXCOM NSA 7160R Multi-Purpose Servers with High Performance

The NEXCOM NSA 7160R is a high-performance, multi-purpose edge server powered by dual 4th Gen Intel Xeon Scalable processors connected by four Ultra Path Interconnects. The system design is also compatible with an upgrade to dual 5th Gen Intel® Xeon® Scalable processors.

This server family is designed to deliver the very powerful computing for demanding edge workloads. This is why the servers meet the requirements of the Intel EVI reference design.

The CPU performance is matched by fast memory speeds with the server supporting up to 16 DDR5 4800/5600 ECC RDIMMS with a per RDIMM capacity of 1024GB. For data storage, the server features two slots for 2.5" swappable SSD/HDD drives.

The flexible design supports eight PCIe 5 slots for networking LAN modules (including 100GbE Intel® Ethernet Network Adapter E810, Intel® Ethernet Network Adapter X710, Intel® Ethernet Controller I226 or Intel® Ethernet Server Adapter I350-based 1~10GbE LAN modules) or other expansion cards (including a storage adapter with up to 4 slots for NVMe, wireless adaptor for 5G and Wi-Fi modules). The design also features redundant power supplies to maximize uptime.

Testing Set Up

Testing conducted using the Intel EVI 2.0 image demonstrating the performance of the NSA 7160R server in a series of tests that measure storage and retrieval of image and video files along with advanced AI capabilities including inference, feature matching and clustering.

The DUT and workload specifications are detailed in the footnotes[1]. Here is a summary description of the NSA 7160R DUT hardware:

- 2 x Intel Xeon Gold 6430 per unit
- 16 (8+8) x 32G DDR5 4800 RDIMMs per unit
- 1 x 128G M.2 2280 SSD for OS
- 1 x 2.5" 512GB SSD SATA III
- 1 x NX 120F-OS – 10GbE LAN module -> slot 1
  - 2 x 10GbE SFP+ ports, Intel® Ethernet Controller X710-BM2
- 1 x NSK 6404-NVMe – Storage adapter -> slot 2
  - 4 x M.2 2280 PCIe4 ×4 4TB NVMe modules

Testing Results

The test results used Intel EVI 2.0 test protocols to examine the throughput of the NEXCOM solution across four workloads that are important for the performance of computer vision applications. The server's performance met all of the performance expectations specified by the Intel EVI reference platform.

The tests include:

Image/Video Storage and Retrieval

These tests used the Intel EVI test tool, which places images and video into and out of storage in order to measure the performance of the storage. The image storage tests demonstrate the throughput the server provides for each action (add, read, remove) using a 2.66 GB image file. A second video storage test focused on upload performance was run using a 1.49 GB video file.

Test results are averages of three results. Tests delivered results in megabits per second, which were multiplied by eight to get the megabytes per second metric.

Tests	Actions	Results
Image storage and retrieval tests	Avg. Add throughput	24.08 MB/s 192.64 Mbps
	Avg. Read throughput	69.04 MB/s 552.32 Mbps
	Avg. Remove throughput	133.32 MB/s 1066.56 Mbps
Video storage and retrieval tests	Avg. Add Throughput	690.54 MB/s 5524.32 Mbps

Table 1. Image storage and retrieval test results.

AI Inferencing (Image/Video)

In these tests, the YOLOv5s object detection model was used to test how many images/video frames can be queried and processed in a second. The image test was configured to serve up three images per query. Test results are an average of three requests.

Tests	Content Type	Results
AI inference	Image	155.48 queries per second 466.44 images per second
	Video	22.35 FPS per stream

Table 2. AI inferencing test results.

Feature Matching

Image descriptors (vectors) are compared across multiple images to identify similar features and allow the comparison of two images or videos. This is an important task in many computer vision applications. In the tests, similar requests were batched in groups of one and 32. For feature matching (batch=1), results are the average of 61 requests with results from the first request eliminated due to time needed to load data to memory. For feature matching (batch=32), results are the average of 11 requests with results from the first request eliminated due to time needed to load data to memory.

Tests	Batch Size	Results
Feature Matching	Batch = 32	31.77 billion vector matching per second
	Batch = 1	1.83 billion vector matching per second

Table 3. Feature matching test results.

Clustering

Clustering is an Intel EVI big data use case designed for vehicle monitoring applications. All incoming new videos and images that are captured and sent by an organization's network of hundreds or more surveillance cameras are retrieved in clusters at a scheduled time so they can be efficiently analyzed and archived.

To get the average performance, 11 test requests were made with the first request not included in the average to account for data loading time. The archiving time was calculated by adding together the HDFS time consumed, and the PostgreSQL time consumed.

Tests	Actions	Results
Clustering	Throughput	1275.82 million
	Archiving time	9.53 seconds

Table 4. Clustering test results.

Conclusion

Intel EVI is an important reference implementation that speeds to market edge servers. With commercial partners it can provide the performance and feature set needed to deliver a complete edge video analytics solution. This is essential for many smart city applications that use security cameras to ensure public safety and other services.

Intel EVI is a reference solution that anticipates the needs of smart cities and others and enables NEXCOM to develop a server that has the compute performance that enables the NSA 7160R to rapidly process both image and video data and execute AI models. The results of the four storage and AI feature tests show that the NEXCOM NSA 7160R delivers the performance needed to meet the requirements of the reference design.

Learn More

NEXCOM

NEXCOM NSA 7160R

4th Gen Intel Xeon Scalable processors

5th Gen Intel Xeon Scalable processors

Intel OpenVINO Toolkit

Intel Smart Edge Open

Intel® Network Builders

Notices & Disclaimers

[1] NSA 7160R device under test: 1-node, 2x Intel Xeon Gold 6430 processors with 252 GB (16 slots/ 32GB/ 4800) total DDR5 memory, microcode 0x2b0004d0, HT on, Turbo on, OS: Ubuntu 22.04, kernel v5.19, storage for boot512GB; storage for application drives is 1.2TB. Workload: EVI 2.0 with FMAL and OpenVINO libraries. Test conducted by NEXCOM in January-February 2024.

Performance varies by use, configuration and other factors. Learn more on the Performance Index site.
Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. No product or component can be absolutely secure.
Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.
Intel technologies may require enabled hardware, software or service activation.
No product or component can be absolutely secure.
Your costs and results may vary.

0424/DC/DJA/PDF ♻Please Recycle 359813-001US