HPE Converged Edge Systems Video Analytics and Management Solution

Abstract
Using a converged approach to solve today's video surveillance and business needs
Notices

The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.


Links to third-party websites take you outside the Hewlett Packard Enterprise website. Hewlett Packard Enterprise has no control over and is not responsible for information outside the Hewlett Packard Enterprise website.

Acknowledgments

Intel®, Itanium®, Pentium®, Intel Inside®, and the Intel Inside logo are trademarks of Intel Corporation in the United States and other countries.

Microsoft® and Windows® are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

Adobe® and Acrobat® are trademarks of Adobe Systems Incorporated.

Java® and Oracle® are registered trademarks of Oracle and/or its affiliates.

UNIX® is a registered trademark of The Open Group.
Video data right at the edge can extract critical insights, helping speed up reaction times, reduce the risk of data transfer, and drive better business decisions.

In today’s world of big data, there is an increasing demand for Artificial Intelligence and hardware accelerated analysis (GPU-based analysis). For example, as the amount of data collected increases, the need for real-time, in-depth analysis increases to meet demand right at the data source, a.k.a. the edge. It is important to delegate AI-enabled video analytics at the edge to take action as close to real-time as possible. In the field of video surveillance, the speed, size, and complexity of data grows exponentially with every camera added to a system. AI-enabled Edge Video Analytics also provide other benefits, including less latency, bandwidth, cost, threats, duplication, and provide higher reliability and compliance.

HPE’s Dr. Tom Bradicich clearly states the bandwidth problem for video analytics. In a recent blog post, Dr. Bradicich states,”Sending data from edge devices to the cloud or a data center can use a tremendous amount of bandwidth. Fearing that such devices will be a drag on the system, some have proposed creating a separate network for the IoT. You can greatly curtail that drag by eliminating the need to send data back and forth. Many companies simply cannot handle the bandwidth needs of IoT right now.” - 7 Reasons Why We Need to Compute at the Edge.

You can find the article at: https://news.hpe.com/7-reasons-why-we-need-to-compute-at-the-edge/

AI-enabled Edge Video Analytics present new challenges. For example, the computing resources needed for AI and video analytics traditionally required data-centers for power, cooling, CPUs, GPUs, memory, and storage. To enable the video management system and video analytics components, AI and Edge-enabled video analytics require multiple systems working in unison. These components require data-center grade resources to accomplish their tasks, but there simply is not enough space, storage, power, or cooling at the edge to handle today’s AI and analytics demands using yesterday’s data-center systems.

To solve this problem, one must push all video data to the data center where traditional resources exist, or one must employ a new category of high-performing systems at the edge. HPE’s solution is to enable the edge with AI, hardware accelerated video analytics, and an integrated video management system using the use HPE Converged Edge Systems. HPE Converged Edge Systems provide data-center level CPUs, GPUs, memory, and storage in a fraction of the space, power, and cooling required by traditional servers.

Modern Closed Circuit Television (CCTV) surveillance systems are often mandated to store the captured video using Video Management Systems Software (VMS). This document describes how HPE is integrating the video management system, AI, and hardware accelerated video analytics into a scalable solution built for the edge. The benefits of converged systems include less space, energy, purchase and operational cost, cables, deployment time, improved performance, and adding the converged applications required to power AI-enabled Edge Video Analytics at the edge. In summary, the HPE Converged Edge Systems provide a highly flexible and dense solution for AI and hardware accelerated video analytics by enabling four GPUs within a single rack unit (RU) with varying analytics workloads.
HPE has created an edge solution for AI and hardware accelerated video analytics using a set of scalable building blocks and layers of various supported features. Converging various features into a single platform and using scalable building blocks to provide the highest level of flexibility and convergence for a most efficient solution. Each feature layer of the solution, for example, the VMS, AI, or video analytics, operates independently on separate server cartridges within the same highly available and redundant chassis, or set of chassis. Scalability of each server component is determined independently, based on the type of analytics or AI being performed, and the number of camera streams per server component. Each individual building block is optimized for each individual feature or analytics type within the solution to provide the highest level of hardware efficiency and performance possible.

An aggregation of similar server components provides overall size and scalability for the feature layer. This means that a single feature layer can be contained at the smallest level within a single server, or span across multiple servers and chassis to build a larger homogenous layer. These homogenous feature layers then build on top of each other to provide the most density, flexibility, and variation of AI and video analytics features available today within a single converged platform. Customers and systems integrators can utilize this scalable framework to build, order, and deploy any number of AI and video analytics converged edge solutions. This is accomplished by simply choosing the number of server building blocks needed per feature, and the quantity of HPE Edgeline chassis to build their own converged VMS, AI, and video analytics solution.

This convergence of VMS, AI, and video analytics into the same HPE Edgeline Converged Edge System enables the lowest latency possible between features while providing the flexibility and cost effectiveness of an integrated solution. The convergence of multiple feature layers within the same unified platform in this way also retains the independence needed for highly performing VMS, AI, and hardware accelerated video analytics features.

**Converged Video Management & Analytics Architecture**

![Converged Video Analytics & Video Management](image)

**Figure 1: Converged Video Management and Analytics Architecture**

**Supported solution components**

Use the following configurations for general reference of available feature layers in the solution. Contact HPE Sales support for a complete list of the orderable components and for additional sizing guidelines.
<table>
<thead>
<tr>
<th>Feature layer/type</th>
<th>Integrated software</th>
<th>Camera streams per component unit/server cartridge</th>
<th>Server component building block</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video Management System</td>
<td>Milestone XProtect® Corporate</td>
<td>200 (1080p@30fps)¹</td>
<td>HPE ProLiant m710x</td>
</tr>
<tr>
<td>Al/Machine Learning Anomaly Detection</td>
<td>iCetana®</td>
<td>62 (720p@12fps)</td>
<td>HPE ProLiant m710x</td>
</tr>
<tr>
<td>Facial Recognition</td>
<td>FaceFirst®</td>
<td>10 (1080p@30fps)</td>
<td>HPE ProLiant m510-8 core</td>
</tr>
<tr>
<td>Color Recognition</td>
<td>MicroFocus IDOL®</td>
<td>Contact HPE</td>
<td>HPE ProLiant m510 8-core</td>
</tr>
<tr>
<td>Object Detection</td>
<td>BriefCam®</td>
<td>Contact HPE</td>
<td>HPE ProLiant m510 8-core</td>
</tr>
<tr>
<td>Object Tracking and Pathing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstructured Data Analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object Detection and Classification</td>
<td>XjeraLabs®</td>
<td>9 (1080p@30fps)</td>
<td>HPE ProLiant m510 8-core</td>
</tr>
<tr>
<td>Face Expression Analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatic License Plate Recognition (ALPR)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Default Settings were used when calculating streams per component unit
2. Some software providers have regional support only. Contact HPE Sales for a complete list of software providers and video analytics features.
HPE Converged Edge Systems

HPE Converged Edge Systems are purpose-built edge compute platforms that converge real-time data acquisition, control, enterprise-class analytics, and remote management using HPE Integrated Lights-Out™ technology. HPE Edgeline is the industry's first product in this category. These systems are compact, energy-efficient, and rugged platforms with a broad range of network connectivity, data acquisition, and control options to accommodate almost all edge video management and analytics use cases. Chassis type, number of servers, number of CPU cores, memory, and storage can all be tailored to fit the precise number and quality of video feeds and intensity of processing and analytics needed.

Customers have two chassis options, and make the choice based on the deployment size, location, and space requirements. HPE recommends using the HPE EL1000 Converged Edge System in deployments with a smaller number of cameras. For mid to large-scale deployments and numbers of cameras, HPE recommends the HPE EL4000 Converged Edge System. Both systems are described below along with their capabilities.

- The HPE Edgeline EL1000 Converged Edge System accommodates up to two NVIDIA GPU cards and one server cartridge. The HPE ProLiant m510 Server Cartridge, incorporating up to 16 Intel® Xeon® cores, 128 GB RAM, up to four TB of blazing fast NVMe SSDs, and a dual-port 10GbE NIC. The PCIe slots support PCIe full-height and half-length (FHHL) cards such as the NVIDIA Tesla P4. The HPE EL1000 Converged Edge System can also support two small form-factor drives for expanded storage capabilities. This system also includes a variety of wireless connection options, for example, Wi-Fi or LTE. It is ruggedized to operate in harsh environments across a temperature range of 0°C (32°F) to 55°C (131°F). For a complete list of hardware specifications, see the HPE EL1000 quickspecs at: [http://www.hpe.com/info/edgeline](http://www.hpe.com/info/edgeline)

The NVIDIA Tesla P4 is powered by the NVIDIA Pascal™ architecture. The Tesla P4 is purpose-built to boost efficiency for scale-out servers running deep learning neural networks (DNNs), enabling smart, responsive, and efficient AI-based analytics. It slashes inference latency by 15X in any infrastructure and provides an incredible 60X better energy efficiency than CPUs. This unlocks a new wave of previously impossible AI services due to latency limitations.
Figure 3: HPE Edgeline EL1000 Converged Edge System

- The HPE Edgeline EL4000 Converged Edge System accommodates four NVIDIA GPU cards, four independent server cartridges, including the HPE m510 servers, giving up to an impressive 64 Intel Xeon cores, 512 GB memory, up to 16 Terabytes of SSDs, and eight 10GbE ports in a slim 1U form factor. The HPE Edgeline EL4000 enables the greatest number of GPUs per rack unit available, offering both high availability and redundancy. It can accommodate up to four NVIDIA Tesla P4 GPUs, each connecting to one server cartridge, giving us an incredible amount of processing power, up to 10240 CUDA cores, to quickly run the most demanding video processing and analytics algorithms at the edge. Redundant power supplies, ruggedization (up to MIL-STD through HPE partners), and the backing of industry certifications such as NEBS Level 3, make this a highly reliable system. This Converged Edge System is comparable to the server equipment used by the Telco industry and other mission-critical operators. Both the HPE EL1000 and EL4000 support several mounting options to fit any environment, and any location, for example a telco rack, server rack, wall, desk, or even within a custom enclosure.
There are multiple storage options available within this solution. Before choosing a storage option, it is important to understand the various factors that will affect the storage requirements. In the field of video recording, the variables include:

- Number of cameras,
- Frame size
  - Compression rate (raw, H.264, H.265, MPEG)
  - Resolution (4K, 1080p, 720p, 420)
- Video retention policy
- Number of hours of recording per day
- Percentage of recording time (if using motion detection)

Use following formulas to determine the amount of storage and network bandwidth requirements:

- Total raw storage per day (Total GigaBytes/day) = cameras * ((Frame size in KB * fps * 3600 seconds * # Hours per day) / 1024MB / 1024GB)
- Total raw storage required = (Total GB/day * days retained)
- Total Mbps (Total Megabits per second) = cameras * (Frame size in KB * 1024MB * fps * 8 bits)

Using these formulas, you can calculate the storage and network requirements. For example:

- **Number of cameras** = 50
- **Frame Size in KB** = 200KB
- **Frames per second (fps)** = 30fps
- Number of hours per day of recording * percentage of motion/recording = 8hrs * 100%
- Number of days retained = 15

**Total raw storage size per day (Total GigaBytes per day)**

\[
50 \times \left( \frac{(200\text{KB} \times 30\text{fps} \times 3600\text{s} \times 8\text{hpd} \times 100\%)}{1024\text{MB} / 1024\text{GB}} \right) = 8,239 \text{TGB/day}
\]

**Total raw storage for 15 days (Total TeraBytes / 15 days)**

\[
8,239 \text{TGB/day} \times 15 / 1024\text{TB} = 120.689 \text{TB / 15 days}
\]

**Total Mbps (Total Megabits per second)**

\[
50 \times \left( \frac{200\text{KB} \times 30\text{fps} \times 8\text{bits}}{1024\text{MB}} \right) = 2,343.75 \text{Total Mbps}
\]

You can find a sample online tool at: [https://www.milestonesys.com/support/let-us-help-you/presales-support/Storage-Calculator/](https://www.milestonesys.com/support/let-us-help-you/presales-support/Storage-Calculator/)

The second aspect of selecting the correct storage option is the calculating the amount of storage resiliency desired. The numbers above include the amount of usable storage required. Be sure that any RAID volumes support the amount of usable storage. The amount of RAID storage components, and the number of usable storage components required, determines the amount of total raw storage needed in a solution.

Consider the built-in M.2 storage with NVMe SSD drives on HPE ProLiant Server Cartridges if you use:

- Small deployment of cameras
- Lower frame rates
- Lower retention periods

The embedded storage option, when used for the recording servers, provides a high-performing and dense storage option that requires no external storage. The M.2 storage option supports software-based RAID configurations, up to four internal M.2 slots, and up to 2TB NVMe internal storage drives per slot.

For medium to large deployments of cameras, average frame rates, good data resiliency, and good data throughput, the ProLiant H241 HBA connected to recording servers and an HPE MSA storage system may be considered. The HPE MSA storage system supports hardware-based RAID, and allows for both 12Gb SAS and 10GbE iSCSI data connections from multiple recording servers.

For the most resilient and highest performing storage, one may consider HPE 3PAR StoreServ Storage with the HPE 3PAR File Persona. This option offers a unique solution that incorporates multi-protocol support into the system architecture to deliver a tightly integrated, truly converged solution for provisioning both block storage volumes and file shares from a single storage system. With truly converged, flash-optimized HPE 3PAR StoreServ Storage, you can address a broad spectrum of workloads and data types, today and into the future—from video storage, databases, video analytics models, user shares, content management, collaboration, data preservation, and governance—efficiently, effortlessly, and without compromise. Only HPE 3PAR StoreServ Storage has the ability to host workload-centric Storage Personas directly on a multi-controller architecture with hardware-accelerated data compactiondelivering a high-performance, low-cost, tier-1 storage platform to address a spectrum of workload needs and data types.
Figure 5: Converged storage for spectrum of workloads and solutions
Video Management Systems

Milestone XProtect® Corporate is a powerful Internet Protocol (IP) Video Management software (VMS) designed for large-scale and high-security deployments. Milestone XProtect® Corporate is a fully distributed solution, designed for large multiple site and multiple server installations requiring 24x7 surveillance, with support for devices from different vendors.

The solution offers centralized management of all devices, servers, and users-empowering an extremely flexible rule system driven by schedules and events. The single management interface enables efficient administration of the entire system. This includes cameras and security devices, regardless of the size or if the device is distributed across multiple sites. For systems that demand situational awareness and a precise response to incidents, Milestone XProtect® Corporate features interactive maps linked to alarms and includes an XProtect Smart Wall for security operations centers. Milestone XProtect® Corporate provides the ultimate system reliability for high-security installations. HPE Edgeline Converged Edge Systems provide the enterprise-grade technology to the edge for the storage support and redundant computing needed for recording and management server functions. Using HPE Edgeline ensures that video recording is never interrupted for real-time analytics of AI and hardware accelerated video analytics at the edge.

Milestone XProtect® Corporate allows users to view any camera from a computer with the Milestone Smart Client installed. The system also offers the possibility of including the standalone XProtect Smart Client-Player when exporting video evidence from the Smart Client. The Smart Client-Player allows recipients of video evidence (such as police officers, internal, or external investigators) to browse and play back the exported recordings without having to install any software on their computers. Finally, Milestone XProtect® Corporate handles an unlimited number of cameras, servers, and users-across multiple sites if required. Milestone XProtect® Corporate is capable of handling IPv4 as well as IPv6.

Video recording and archival

The HPE Edgeline EL4000 is perfectly suited for deployment at most edge sites. For example, a traffic junction, telco cabinet or security room. The HPE Edgeline EL4000 provides a variety of mounting options that allow it to also fit into existing sites. Depending on the number of cameras and stream quality, each HPE Edgeline system can be configured with 1 to 4 recording servers. If needed, the event, mobile and management servers can also be hosted at the same edge site and within the same chassis.

The HPE 3PAR Converged block and file solution offers you a variety of scalable, cost-effective storage options, an ideal match to support the multistage archive capability built into Milestone XProtect Corporate. The File Lock feature of HPE 3PAR StoreServ enables data preservation to protect against accidental, premature, or malicious deletion and modification of data. It also provides data retention for any period as required by legal statute or internal policy within your IT department. For more storage sizing guidelines, see HPE Storage.
Figure 6: Architecture diagram

Video and audio recordings, as well as related metadata, is stored in the Milestone dedicated high-performance media database. The database is optimized for storing and retrieving the recordings. The media database supports various unique features, including tiered multistage archiving, video grooming, encryption, and adding a digital signature to the recordings. The media database supports a tiered storage architecture with "Live" recording database and the "Archive" databases distributed across different storage systems and technologies, making it possible to design and optimize the storage solution for performance (recording), size (retention), and cost.

The multistage archiving support, in combination with the grooming feature with converged HPE 3PAR StoreServ Storage, allows recordings to be repeatedly archived to capacity optimized 3PAR File Persona shares. Each archive requires less and less storage capacity over time. Each archive level has a different data retention period. For example, archive level 1 may be 14 days, archive level 2 may be 30 days, and archive level 3 may be 90 days. A file archived from the "Live" database to level 1 will be automatically migrated to level 2 after 14 days. This same file will be moved from level 2 to level 3 after 30 days. Finally, after 90 days, the archived file is removed from the system. As the file moves through each level, it is groomed, reducing the frame rate of the recorded video to save storage space.

The live video streams are stored, in the live video database, on the local drives which is the NVMe SSD on the cartridge. Second level storage, the archive database, is stored on a 3PAR 8200 storage. The target retention period for the live recording data was 1 hour. When camera streams reach the recording server, the recording server will write the media metadata on local disk. After 1 hour retention period, the recording server will delete data to release space on local disk.

Video walls and monitoring

Video walls are often used with CCTV and video management systems to provide the visibility and control for monitoring, management, and operation. HPE’s solution includes Milestone XProtect Smart Client software to provide operations personnel with the required amount of control and visibility needed within the environment. These HPE Edgeline systems connect either directly to multiple monitors using an HPE Edgeline EL1000, or connect via a remote network protocol to provide a seamless and integrated solution with the HPE Edgeline
EL4000. Multiple users and monitors can be supported on the same HPE Edgeline system for added scalability.

Figure 7: Milestone XProtect Smart Client with Smart Wall

With the latest version of Milestone software, HPE Edgeline also provides hardware accelerated H.264 and JPEG video decoding, color correction, and scaling for XProtect Smart Client users and video walls. Hardware acceleration of Milestone XProtect Smart Client is supported with both the built-in Intel® IrisPro GPU on the HPE ProLiant m710x cartridges, or with an NVIDIA PCIe-based GPU on HPE ProLiant m710x and m510 server cartridges. In most cases, hardware acceleration is enabled automatically if a supported GPU is detected within the HPE Edgeline system.

Refer to the following for requirements on hardware acceleration with Milestone XProtect Smart Client: [https://developer.milestonesys.com/s/article/hardware-acceleration-in-Smart-Client-2018-R1-and-newer](https://developer.milestonesys.com/s/article/hardware-acceleration-in-Smart-Client-2018-R1-and-newer)

Integration with video analytics ISVs

Another benefit to enabling VMS within the HPE Converged Edge solution is that modern VMS systems also support integration with third-party video analytics independent software providers (ISVs). With this integration comes the ability to perform AI and hardware assisted video analytics on hours or days of previously recorded videos without the need to duplicate raw video files across the multiple AI and video analytics systems. Thus, both the VMS subsystem, and add-on video analytics subsystem(s) benefit from less storage space required, and less latency for video transfer when both the VMS and video analytics systems are integrated into the same HPE Edgeline Converged Edge System.
Figure 8: BriefCam Video Synopsis Integrated with Milestone XProtect
The field of video analytics includes a wide variety of capabilities, features, and software products to choose from. It can be challenging at the beginning of a journey to try to choose which type of video analytics is needed, as well as the best software product to use. To help clarify within this document, the varying artificial intelligence, machine learning, and video analytics providers are categorized by feature. The following table represents the type of video analytics features currently supported within the HPE AI and video analytics solution. HPE is working with an ecosystem of independent software vendor (ISV) partners to enable one or more to run on HPE Edgeline Converged Edge Systems. For example, Facial Recognition, object/scene recognition and license plate recognition are commonly requested features. One or more of HPE’s partner ISVs may be able to meet the requirements of a given use-case, or specialize in providing one feature or the other. For example license plate recognition can be offered by Microfocus IDOL. Xjera Xhound, OpenALPR or Platesmart etc. and the choice depends on factors including cost, ease of use, and configurability. A vendor ISV can either run bare metal, under hypervisor support, or as a microservice (containers) depending on its licensing terms and support. A single ISV can scale across multiple nodes in an EL4000 chassis providing extra-ordinary density advantage, or distinct ISVs can be run as needed on different nodes of the system. Since AI enabled Video analytics is a new and emerging area, several ISVs provide depth of capabilities; however, there are few ISVs which provide a breadth of capabilities. In case a legacy VMS system is being used on legacy hardware which needs an upgrade, a Video Management System (VMS) recording function can be run on the same hardware external storage, for example, a JBOD, MSA, or a 3Par storage system, depending on the scale. Running a high performance video application at the edge on Edgeline Converged Edge Systems has been proven to deliver extraordinary performance, density (translating into high Streams/Rack Unit or $/Stream).

Please contact HPE Sales support for more information on performance benchmarks in support of this statement; a few representative analytics features are provided below:
**Table 2: List of video analytics features**

<table>
<thead>
<tr>
<th>Component Type / Feature</th>
<th>Integrated Software</th>
<th>Hardware Acceleration</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI/Machine Learning</td>
<td>iCetana®</td>
<td>NVIDIA Tesla P4</td>
</tr>
<tr>
<td>Anomaly Detection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facial Recognition</td>
<td>FaceFirst®</td>
<td>NVIDIA Tesla P4</td>
</tr>
<tr>
<td>Color Recognition</td>
<td>MicroFocus IDOL®</td>
<td>NVIDIA Tesla P4</td>
</tr>
<tr>
<td>Facial Recognition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatic License Plate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognition (ALPR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object Detection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object Tracking and Pathing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstructured Data Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object Detection and Classification</td>
<td>BriefCam®</td>
<td>NVIDIA Tesla P4</td>
</tr>
<tr>
<td>Object Tracking and Pathing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video Synopsis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Face Expression Analysis</td>
<td>XjeraLabs®</td>
<td>NVIDIA Tesla P4</td>
</tr>
<tr>
<td>Facial Recognition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatic License Plate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognition (ALPR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object Detection and Classification</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Some software providers have regional support only. Contact HPE Sales for a complete list of software providers and video analytics features.

The key to delivering real-time actionable results from analog input, in this case video, lies in deep neural networks, artificial intelligence, and machine learning. Today’s advances in deep neural networks have revolutionized the speed and accuracy that computers train, detect, and act on events. For example, accuracy results of up to 97 percent can now be achieved in near real-time for object detection and object recognition. Whether it is a single type of analytics required, or a mix of varying types of analytics, HPE is leading the way to deliver integrated, flexible, and scalable solutions. HPE has partnered with some of the leading companies in the AI and video analytics space in order to create a platform that is performance optimized across the different types of analytics engines, scalable to any size of deployment, and flexible enough to support any mix of analytics feature.

Since there are many types of video analytics features and software products to choose from, the following definitions are available for additional clarification.

**Video anomaly detection**

Video anomaly detection is a form of video analytics that uses self-learning methods to determine the abnormal surveillance events that require attention. This is accomplished by self-learning the "normal" patterns in any given scene and then automatically reporting on the abnormal activity.

For more information, see: [https://icetana.com/the-icetana-solution/](https://icetana.com/the-icetana-solution/)
Face expression analysis

Face expression analysis is an area of research that attempts to identify the emotional state of an individual based on a well-defined set of metrics and muscular changes used to convey emotion. For example, modern computers use face expression analysis metrics to determine a selected individual's emotional state.

For more information see: [https://www.xjeralabs.com/product/Product_Brochure_Xintelligence.pdf](https://www.xjeralabs.com/product/Product_Brochure_Xintelligence.pdf)
Facial recognition

Face recognition describes a biometric technology that attempts to establish an individual's identity, also known as facial recognition or face detection. The process uses a computer application that captures a digital image of an individual's face (sometimes taken from a video frame) and compares it to images in a database of stored records.

For more information see: https://www.facefirst.com/face-recognition-glossary/what-is-face-recognition/
Object detection

Object detection is a category of computer vision that focuses on the detection of different classes of objects, for example, person, vehicle, dog, cat, buildings, roadway, etc.) in a given scene. Object detection strives to identify the various classes of objects, rather than the differences between each object within a given class.

For more information see: https://support.hpe.com/hpsc/doc/public/display?docId=emr_na-c05336736&docLocale=en_US

Object classification

Object classification takes object detection a step further. Object classification characterizes the objects in a given class by a standard set of attributes associated with that class. For example, people could be further
characterized by having attributes such as gender, age, ethnicity, hair length, hair color, clothing type, or clothing color.

For more information see: [https://www.briefcam.com/technology/how-it-works/](https://www.briefcam.com/technology/how-it-works/)

**Object tracking and pathing**

Determining a moving object’s direction, path, and location is a form of video analytics sometimes referred to as object tracking or “pathing”. Simply put, object tracking analyzes an object’s historical movement within a scene or field of view. This is useful when classifying objects based on direction, time spent within a specific location, or path taken within the scene.

For more information see: [https://www.briefcam.com/solutions/planning-research/](https://www.briefcam.com/solutions/planning-research/)

**Automatic license plate recognition (ALPR)**

ALPR is an acronym synonymous with the identification of vehicles based on the unique set of characters or numbers used for vehicle registration. ALPR is useful for automatically searching suspects identified with an
investigation, or with payment systems tied to toll-ways, parking, or access to restricted areas. For more information see: https://www.openalpr.com/forensic_plate_finder.html

Figure 16: Automatic License Plate Recognition by OpenALPR

Video Synopsis® by BriefCam

The ability to review hours of video within minutes or seconds is called Video Synopsis. This capability enables surveillance and security teams to search through and report on video from multiple camera inputs within an accelerated timeframe. Video Synopsis visualizes object characteristics and superimposes the objects onto a stationary background. It also simultaneously displays events occurring at different times, with the ability to link back to original video. For more information see: https://www.briefcam.com/technology/video-synthesis/
Artificial Intelligence

Classical Machine Learning v/s Deep Neural Networks for Computer Vision

Artificial Intelligence refers to the broad field of techniques used to make computers mimic human behavior, or provide human-like responses.

Machine Learning is a sub-field of AI that allows computers to learn from examples, instead of using explicitly programmed rules, by creating models from sample data.
Deep Learning is a subset of machine learning which involves the use of artificial neural networks as models. While classical machine learning approaches might learn from examples, they typically required expertly crafted feature engineering, and the design of classifiers.

"Coming up with features is difficult, time-consuming, requires expert knowledge. (Therefore) Applied machine learning is basically feature engineering." - Andrew Ng, Machine Learning and AI via brain simulations.

In the example shown above, object detection can be accomplished using Haar classifiers and blob detection. Haar features are used together with training samples of a particular type (human faces or cars etc.) to generate thresholds for particular scenarios. This entire process was human mediated and required, at the least, expert involvement in the design, deployment and tuning of the implementation in a given use-case. This resulted in poor scalability and limited accuracy, which was achievable using these techniques (for example up until 2012, 67% accuracy would be considered state of the art).

Since 2012, neural nets began to be re-looked at for such image classification problems. Artificial neural networks are vaguely inspired by biological neurons. Each node of a neural network performs a mathematical operation and can signal adjoining neural networks connected to it. In case of images or video, a convolution operation is typically performed in each node. Between each stage, there might be steps for data reduction (called pooling). When such a mathematical structure is provided positive reinforcement for a certain types of images (the types of objects that we would like to train - say vehicles), the pathways which lead to the correct classification of vehicles are given a boost whereas the others are penalized. This process when used over thousands of images modifies the weights of certain interconnections whereas downplays others. Once the neural network reaches a pre-set threshold on test images (as distinct from training images), such a network is considered trained for vehicle detection and can be tested on unknown images. Networks trained in this manner can easily achieve ~ 97% accuracy and are therefore much more accurate and interesting today.

The process of testing neural networks on unknown images ("new data") is called inferencing. Unlike the training phase which involves iterative back propagations on a neural network and the modification of its weights, inferencing involves only one forward pass through a pre-trained network. In that case, it is important to either be able to process a large number of test images (batch-optimized inferencing) or process a given image with lowest latency (latency-optimized inferencing).

A pre-trained neural network is simply a description of the structure of the network, the weights interconnecting them and can range from a few hundred MB's to few GB's depending on their complexity. Neural networks have been historically studied for decades; however the wide availability of training data, as well as the required computational power to process pixel-by-pixel, has only become mainstream in the last decade leading to their popularity.

Another key advantage with neural networks is that there is no need to do complex feature engineering which is entwined with the implementation of the algorithm. Raw images of various classes can be passed through a common software stack setup for training and inferencing without the need for applying Haar-like features or others to pre-process and prepare images.

This reduces the need to have subject matter experts for each domain such as images, audio or text but require practitioners of DNNs. In many cases, the process of changing hyperparameters is done in an ad-hoc manner; neural network structures which work well for a given class of images are known and only need to be retrained for certain scenarios (cameras, lighting etc.) Further once a neural network has been trained, it is also possible to update only the last few layers to add classifiers for new objects (transfer learning).

**Hardware accelerated video analytics**

The use of Deep Neural Networks for image and vision classification problems comes at the cost of significantly increased computational complexity for both training and inferencing problems. Training is an iterative process which involves both a forward and backward propagation step to update weights. Whereas systems for inferencing have to be optimized to either process a large batch of images or video or do so at acceptable latencies for real-time detection.
Silicon alternatives for DNNs

While it is possible to perform inference for images at sub-second latencies using only a CPU, this is unusable when dealing with video data. For example, video running at 30 fps provides a window of only 1 s / 30 = 33 ms to operate on a frame and make a classification result. Further, a reasonably cost-effective system should not only infer on a given frame, but a plurality of frames (from tens to a few hundreds), each being derived from multiple input streams (cameras). Most video today from CCTV cameras is encoded using block-oriented video compression standards such as H.264 to reduce bandwidth and maintain efficacy in recording, compression, and distribution of video content. This requires hardware decode units to process an incoming stream in real-time before passing the stream on to a Deep Learning Inference System. Finally since DNNs operate at a pixel level (each requiring some convolutional operation), the computational requirement is exponentially larger requiring the use of accelerators.

Various silicon alternatives exist for performing hardware accelerated video analytics. NVIDIA GPUs have now been tuned across multiple generations to be optimal in processing DNN workloads. For example, GPU cards optimized for training and inference are exposed to developers and software stack developers in multiple ways. For example unlike a standard GPU CUDA core which can perform a single multiply-accumulate operation (x+= y*z) in 1 clock cycle; a modern tensor core can perform 1 4x4 matrix-multiply accumulate operation per clock cycle (X+=Y*Z). Internally this is a mixed precision operation since Y,Z can be FP-16 whereas X is FP-32. This helps significantly since all mathematical operations in DNN's can be expressed as matrix operations. Further, it has been recognized that inference operations do not require floating point operations and can be performed using only integer arithmetic.

NVIDIA TensorRT provides an automated, parameter-less method for converting FP32 models into 8 bit fixed point (INT8). This is performed using a quantization operation while still minimizing information loss. NVIDIA GPUs such as the Tesla P4 have included hardware decode (and encode units) which can be used to feed into a Neural network model. NVIDIA's DeepStream SDK make the use of hardware decoders together with CUDA run-time libraries, CUDNN (a library of functions used to run DNNs) etc. within a seamless operation. At a higher level, they provide bindings to commonly used higher level programming models for neural networks such as Tensorflow, Caffe etc.

Field-Programmable Gate Arrays (FPGAs) from Intel and Xilinx are also being used for DNN training and inference such as in Project Brainwave by Microsoft and with Deephi Technologies/Baidu respectively. Neural networks trained in the cloud can be inferred upon at the edge using GPU or FPGA accelerators. The
physical implementation of neural networks in an FPGA is typically in the form of systolic arrays. Unlike GPUs; they can do so at comparatively lower power. Microsoft Azure provides a service to train neural networks in the cloud and deploy them at the edge onto HPE Edgeline converged Edge Systems using Azure IoT Edge. Finally various hard DNN processing units (DPUs) are being designed such as vector processors from Intel Movidius™ and Intel Nervana and Google's Tensor Processing Unit (TPU). While progressing towards the right (FPGAs, hard DPUs and ASICs) may increase efficiency (TOPS/W), they tend to be less flexible or require considerable investment in software engineering resources to create a pipeline including compilers, quantizers, and run-times to map into the appropriate hardware layers.

For an example use-case with hardware accelerated VA, please refer to Delivering accelerated video analytics at the edge for AI Cities located at: https://h20195.www2.hpe.com/V2/getpdf.aspx/a00004240enw.pdf
Conclusion

The key to delivering edge-based and real-time video analytics solutions lies in creating performance optimized, tightly integrated, and scalable components. HPE has done just that by using a set of scalable building blocks and tightly integrated layers of various supported features within the HPE Edgeline Converged Edge System; a solution for video management, AI, and hardware accelerated video analytics. Converging various types of features into a single platform and combining with scalable building blocks provides the highest level of flexibility and convergence for a most efficient solution. Furthermore, the ability to store high volumes of IP video surveillance data efficiently is critical for the overall performance and reliability of the surveillance solution.

HPE Edgeline Converged Edge Systems converge enterprise-grade computer (Intel Xeon CPUs), high-performance storage, high speed networking (up to 10 GbE), data acquisition and control technology, and industry-leading management using HPE Integrated Lights-Out (iLO) into a compact and ruggedized form factor suitable for placement at edge sites. The HPE Edgeline EL4000 is a slim 1U system that can be racked or mounted on a wall but dense enough to support up to four NVIDIA GPUs and four ProLiant servers.

The HPE 3PAR Converged block and file solution offers you a variety of scalable, cost-effective storage options and is an ideal match to support the multistage archive capability built into Milestone XProtect Corporate. It can be seen that HPE Edgeline Converged Edge System, HPE 3PAR, and ecosystem of video management system such as milestone make the video surveillance easy to monitor and scale.

The emergence of new technologies such as deep neural networks (DNNs) and hardware accelerated video analytics tremendously increases computer vision's capabilities of accuracy and timeliness to results. Combining DDNs and hardware accelerated video analytics with edge computing enables truly real-time video analysis for a multitude of analytics features and use-cases.
<table>
<thead>
<tr>
<th>Resource</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Edge Video Analytics - Overview</strong></td>
<td><a href="http://www.hpe.com/support/EL-EVA">http://www.hpe.com/support/EL-EVA</a></td>
</tr>
<tr>
<td><em>Enable Exceptional Scene Analysis &amp; Object recognition performance with IDOL Media Server on HPE Edgeline</em></td>
<td></td>
</tr>
<tr>
<td><strong>Accelerated Edge Video Analytics - Overview</strong></td>
<td><a href="https://www.hpe.com/h20195/V2/getpdf.aspx/a00004240enw.pdf?ver=5.0">https://www.hpe.com/h20195/V2/getpdf.aspx/a00004240enw.pdf?ver=5.0</a></td>
</tr>
<tr>
<td><em>HPE Edgeline and NVIDIA IVA enable rapid time-to-insight for AI capabilities at the edge</em></td>
<td></td>
</tr>
<tr>
<td>Getting Started with IDOL Media Server on HPE Edgeline IoT Systems</td>
<td><a href="https://solutionsdepot.int.hpe.com">https://solutionsdepot.int.hpe.com</a></td>
</tr>
<tr>
<td>Step by step guide to setting up Facial Detection and Recognition on HPE Edgeline 20,1000, 4000</td>
<td></td>
</tr>
<tr>
<td><strong>Delivering accelerated video analytics at the edge for AI Cities</strong></td>
<td><a href="https://h20195.www2.hpe.com/V2/getpdf.aspx/a00004240enw.pdf">https://h20195.www2.hpe.com/V2/getpdf.aspx/a00004240enw.pdf</a></td>
</tr>
<tr>
<td><strong>Milestone Storage Calculator (JPEG)</strong></td>
<td><a href="https://www.milestonesys.com/support/let-us-help-you/presales-support/Storage-Calculator/">https://www.milestonesys.com/support/let-us-help-you/presales-support/Storage-Calculator/</a></td>
</tr>
<tr>
<td><strong>7 Reasons Why We Need to Compute at the Edge</strong></td>
<td><a href="https://news.hpe.com/7-reasons-why-we-need-to-compute-at-the-edge/">https://news.hpe.com/7-reasons-why-we-need-to-compute-at-the-edge/</a></td>
</tr>
</tbody>
</table>