EPIC iO Performance and Validation Report for OpenVINO™ Integration on Dell Technologies PowerEdge* R750

Report

August 2022

Document Number: 739130-1.0

You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a non-exclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Copies of documents which have an order number and are referenced in this document may be obtained by calling 1-800-548-4725 or visit www.intel.com/design/literature.htm.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No product or component can be absolutely secure. Check with your system manufacturer or retailer or learn more at intel.com.

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® Turbo Boost Technology requires a PC with a processor with Intel Turbo Boost Technology capability. Intel Turbo Boost Technology performance varies depending on hardware, software and overall system configuration. Check with your PC manufacturer on whether your system delivers Intel Turbo Boost Technology. For more information, see http://www.intel.com/technology/turboboost

Intel, the Intel logo, OpenVINO and the OpenVINO logo are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.

Contents

Figures

Tables

Revision History

Date	Revision	Description
August 2022	1.0	Initial release.

§

1.0 Overview

This document provides an overview and results for validation of an EPIC iO* AI OpenVINO™ Video Analytics Algorithm port running on an enterprise server solution (Dell Technologies PowerEdge* R750) installed with a production version of EPIC iO AI Platform optimized for Intel's OpenVINO™ Acceleration.

Video Analytics algorithm (VA algorithm) can be performed either on CPU, or on a special High-Density Deep Learning (HDDL) acceleration card. The focus of this report will be running the VA algorithm on the CPU.

Configuration for multi-stream in-process analytics (no video stored) includes a pipeline process of video decode, video analytics via AI model with video analytics metadata creation, and injection of metadata into reporting and visualization platform.

1.1 Objective

The objective of the validation process is to:

i. Validate and Size the system configuration for concurrent multi-stream video analytics.
ii. Ensure that Thermal and Cooling setting provides long hours sustainable performance.
iii. Validate that VA algorithms' payload is evenly distributed across all compute units:
    - For CPU VA acceleration – the balancing load across all CPU cores is validated
iv. Confirm that maximum video analytics channel density is achieved at 90-95% of maximum compute capacity:
    - For CPU Video analytics, VA channels are added till CPU reaches 90% load
v. Confirm that overall software/hardware solution is steady and operates without fail(s) for at least 24 hours.
vi. Measure and log key system running parameters:
    - Overall system CPU load: average and standard deviation. Sampling every 1-sec cadence
    - System inlet and outlet air temperature (iDRAC)
    - System fan RPMs and cooling configuration (iDRAC)
    - System average power consumption (iDRAC)
    - Video analytic inference performance in frames per second: Average and deviation

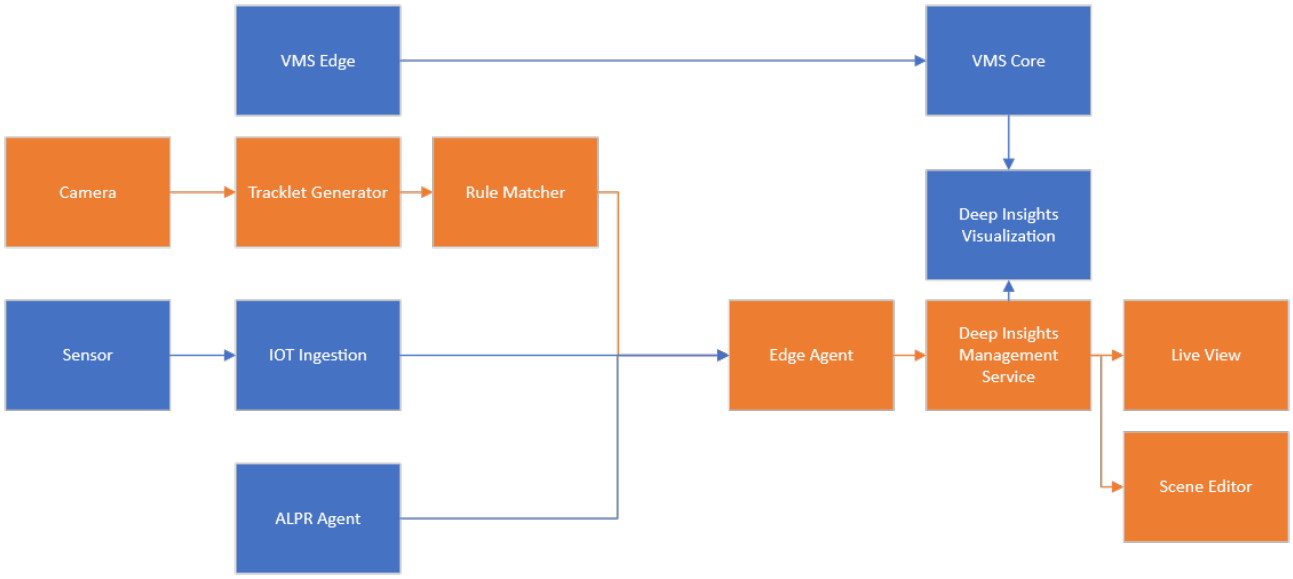
1.2 EPIC iO Architecture Overview

The EPIC iO platform is designed as a distributed microservice architecture. This system supports IoT, AI, video retention, and analytics dashboards.

For the purpose of this report, we'll focus on the AI pipeline (marked in orange in Figure 1).

1. The video is generated by the camera, in this case we are using a camera emulator that generates a H.264 stream based on a video file.
2. The video is ingested by the Tracklet Generator service using a camera reader. Tracklet Generator performs detection, classification and tracking of multiple objects using a CNN. The collection of instances of the same object across multiple frames (Tracklet) is consumed by Rule Matcher for evaluation. Each camera has its own Tracklet Generator configured to use different algorithms and settings.
3. Rule Matcher can be configured to process multiple rules based on the movement pattern of the objects and generate telemetry based on the events in the video. Rule Matcher is configured using the EPIC iO Scene Editor, which pushes scene elements (lines, zones and routes) and the rules associated to each of them to the Rule Matcher.
4. The telemetry generated by the Rule Matcher is consumed by the Edge Agent and pushed to the Deep Insights Management Service (DIMS) for visualization and analysis.

Figure 1. The EPIC iO Platform



1.3 Intel® Distribution of OpenVINO™ Toolkit Overview

OpenVINO™ is an open-source toolkit for optimizing and deploying AI inference.

1. Boost deep learning performance in computer vision, automatic speech recognition, natural language processing and other common tasks.
2. Use models trained with popular frameworks like TensorFlow*, PyTorch* and more.
3. Reduce resource demands and efficiently deploy on a range of Intel® platforms from edge to cloud.

Click here to learn more about the Intel® Distribution of OpenVINO™ Toolkit.

§

Reference: https://docs.openvino.ai/latest/index.html

2.0 System Configuration

Table 1. System Configuration

Components	Version
Hardware	
Chassis	Dell Technologies PowerEdge R750
CPU	2x Intel® Xeon® Gold 6338N CPU @ 2.20 GHz, 32 Core(s), 64 Logical Processor(s)
Memory	Installed Physical Memory (RAM) of 256 GB
Hard drives	512GB Total Storage but not leveraged for Storage
HDDL card	None
Network card	Intel® Ethernet Network Adapter E810-DA4 QP 25GbE SFP28 OCP 3.0
Others	N/A
Software	
BIOS	Dell Inc. 1.4.4
iDRAC	5.00.20.00 (Build 22)
Operating System	Ubuntu 20.04.3 LTS
VA Software	EPIC iO AI Platform with Multiple Models
OpenVINO™	OpenVINO™ Toolkit v.2021.3 LTS
Others	Hyper Threading (Logical Processor in BIOS)
Others	Enabled dynamic CPU frequency

2.1 Processor Details

Name	Processor	Version	Current Speed	Core Count
CPU1 Status	Intel® Xeon® Gold 6338N CPU @ 2.20 GHz	Model 106 Stepping 6	2.20 GHz	32
CPU2 Status	Intel® Xeon® Gold 6338N CPU @ 2.20 GHz	Model 106 Stepping 6	2.20 GHz	32

2.2 Dell iDRAC Processor Settings

Dell iDRAC Processor Settings	
Logical Processor	Enabled
CPU Interconnect Speed	Maximum data rate
Virtualization Technology	Enabled
Kernel DMA Protection	Disabled
Directory Mode	Enabled
Adjacent Cache Line Prefetch	Enabled
Hardware Prefetcher	Enabled
DCU Streamer Prefetcher	Enabled
DCU IP Prefetcher	Enabled
Sub NUMA Cluster	Disabled
MADT Core Enumeration	Round Robin
UPI Prefetch	Enabled
XPT Prefetch	Enabled
LLC Prefetch	Enabled
Dead Line LLC Alloc	Enabled
Directory AtoS	Disabled
Logical Processor Idling	Disabled
AVX P1	Normal
Intel SST-BF	Disabled
Intel SST-CP	Disabled
x2APIC Mode	Enabled
AVX ICCP Pre-Grant License	Disabled
AVX ICCP Pre-Grant Level	128 Heavy
Number of Cores per Processor	All
Processor Core Speed	2.20 GHz
Processor Bus Speed	11.2 GT/s
Local Machine Check Exception	Disabled
Family-Model-Stepping	6-6A-6
Brand	Intel® Xeon® Gold 6338N CPU @ 2.20GHz
Level 2 Cache	32x1280 KB
Level 3 Cache	48 MB
Number of Cores	32
Maximum Memory Capacity	6 TB
Microcode	0xD000311

Family-Model-Stepping	6-6A-6
Brand	Intel® Xeon® Gold 6338N CPU @ 2.20GHz
Level 2 Cache	32x1280 KB
Level 3 Cache	48 MB
Number of Cores	32
Maximum Memory Capacity	6 TB
Microcode	0xD000311

Settings	Current Value
System Profile	Performance Per Watt (DAPC)
CPU Power Management	System DBPM (DAPC)
Memory Frequency	Maximum Performance
Turbo Boost	Enabled
C1E	Enabled
C States	Enabled
Memory Patrol Scrub	Standard
Memory Refresh Rate	1x
Uncore Frequency	Dynamic
Energy Efficient Policy	Balanced Performance
Monitor/Mwait	Enabled
Workload Profile	Not configured
CPU Interconnect Bus Link Power Management	Enabled
PCI ASPM L1 Link Power Management	Enabled
OS ACPI Cx	OS Cx C2
GPSS Timer	500 us
CPU C1 Auto Demotion	Disabled
CPU C1 Auto UnDemotion	Disabled
Workload Configuration	Balance
Dynamic L1	Disabled
Package C States	Enabled
Package C State Latency Negotiation	Disabled
Power and System Criteria for Package C State	Disabled

NOTES:

1. iDRAC = Integrated Dell Remote Access Controller.

3.0 EPIC iO System Configuration

3.1 Video Stream Configuration

Component	Settings	Comments
Video Analytic Input video stream parameters	1920x1080@30fps (1080p)	High-resolution video stream
Number of input video streams for analytics (virtual cameras)	1-80	Each virtual camera stream has high-resolution and low-resolution videos
Video analytic inference framerate per video channel	Variable based on the load	Each AI service is set to process the max amount
Number of active video analytics streams at maximum testing	80 max (refer to *Performance Test Results* for details)	Maximum Number of Streams where video analytics were applied

3.2 Video Analytics Algorithm Description and Parameters

Item	Vehicles Detection and Tracking – Intel® Xeon® Gold CPU
Description	Video ingestion, multi class vehicles detection and tracking.
Detection CNN Architecture	YOLOX
Feature extraction CNN Architecture	SqueezeNet
Service Name	TrackletGenerator
Version	2.6.1
Date	Jul-22

§

4.0 Profiling

4.1 Validation Steps

1. Deploy and Configure Dell Technologies* PowerEdge* R750 Server.
2. Install Ubuntu* Operating System and Analytics Platform with Testing Criteria.
    a. Set up maximum virtual video streams with specified video sources for high-resolution streams.
    b. Set up EPIC iO* video analytics to process the virtual video streams.
3. Run the profiler tools to record hardware usage and other metrics over a given period of time.
4. Process results to generate tabulated data using multiple readings.
5. Analyze results and report.

4.2 Checklist for Results Validation

i. EPIC iO is utilizing the maximum amount of CPU without compromising the system accuracy.

ii. Processing frame rate is matching the expectations.

iii. CPU usage and Memory consumption values are consistent during the test.

iv. Archive is being recorded constantly during the test.

§

5.0 Performance Test Results

To measure system scalability, we sequentially increased the number of streams being processed in parallel while keeping records about hardware utilization and processing time for each stream.

5.1 Analysis

The total memory used by the system grows almost linearly based on the number of streams. Certain services are shared among all AI services, which explains why the growth is not fully linear. In the case of a single stream, the memory usage increment produced by a single AI service is insignificant compared to the idle system.

The memory utilization per stream is almost constant and the small variations are due to the dynamic memory allocation as the system processes the streams.

The total CPU usage reassembles a logarithmic function with a limit based on the maximum possible CPU utilization. The increment in the CPU is almost linear between 1 to 45 streams and starts to slow down after 45 streams.

See Sections 2 and 3 for workloads and configurations. Results may vary.

5.2 CPU Temperature and Frequency per Stream

Number of streams	CPU 1 temperature [°C]	CPU 2 temperature [°C]	Average CPU frequency [MHz]
Idle	31	31	811.97
1	32	33	873.86
5	36	38	1131.79
15	47	51	1564.83
30	69	73	2169.16
45	77	80	2184.07
60	78	82	2314.79
80 [1]	77	81	2335.49

See Sections 2 and 3 for workloads and configurations. Results may vary.

§

[1] Used for test purposes only. Not to be used for sizing.

6.0 Conclusion

Based on the analysis in this report, we can define the specifications required per stream/camera to be deployed using the Dell Technologies PowerEdge R750 with the dual socket Intel® Xeon® Gold 6338N CPU.

6.1.1 Recommended Specifications per Stream

Cores	Speed	Memory	Recommended Camera Capacity
3 cores	2.2GHz	2.2 GB	45

Please consult with EPIC IO regarding specific requirements. [2].

6.1.2 Minimum Specifications per Stream

Cores	Speed	Memory	Maximum Camera Capacity
2 cores	2.2GHz	2.2 GB	60

§

[2] Depending on the use case scenario, the system may need to run with 3 cores per stream, or with a minimum of 2 cores per stream. The main driver to decide which spec to use is based on the speed and number of the objects to be detected.