

Choose the Right Hardware

Proposal

Scenario 1: Manufacturing

Client Requirements and Potential Hardware Solution

Look through the scenario and find any relevant client requirements. Then, suggest a potential hardware type and explain how this hardware would satisfy each of the requirements.

Which hardware might be most appropriate for this scenario? (CPU / IGPU / VPU / FPGA)
FPGA (Field Programmable Gate Array)

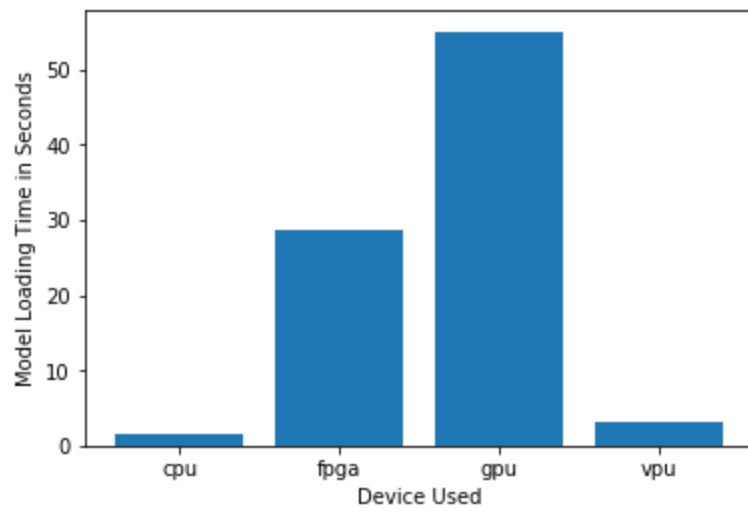
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
The client wants to monitor the number of peoples in the factory line. The installed Vision Cameras can record videos at 30-35 FPS. The client wants to improve the image processing task by upto Five times per second.	FPGAs have lowest latency because of their architecture. FPGAs can be connected directly to inputs and can offer very high bandwidth. FPGAs can parallel process.
The client wants to repurpose the hardware once the above task is completed. The flawed chips are being shipped. The client wants to detect these flaws priorer to the packaging.	FPGAs are field programmable, perfect for this appliation. FPGAs are fast.
The client wants to install a quality system which, as per the requirement, should last for at least five to ten years.	The FPGA cost is under the client requirement. FPGAs have a lifespan of 10 years.

Queue Monitoring Requirements

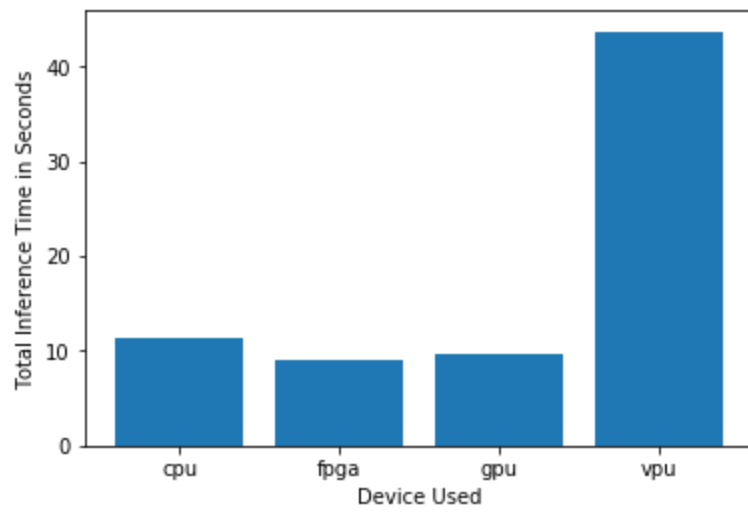
Maximum number of people in the queue	2
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

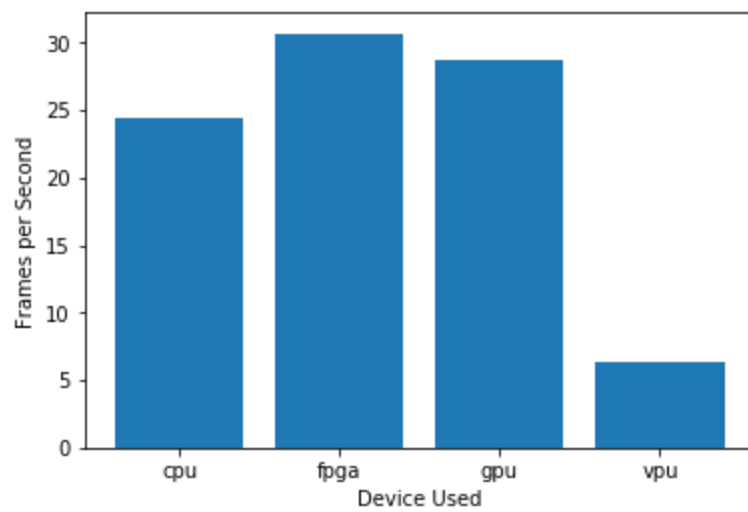
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



FPS

Final Hardware Recommendation

Now synthesize your points from above and provide a brief write-up describing why the chosen hardware is the best choice for this scenario. Be sure to discuss the client's requirements, the test results, and how these relate to one another (e.g., perhaps one of the devices performed better than the rest, but does not meet one of the client's requirements).

Write-up: Final Hardware Recommendation

The above graphs paints a very clear and vivid picture. As you can see, the inference time for the FPGA is the lowest among the CPU, GPU and VPU. The 'Frames Per Second' count is highest for the FPGA. This is as per the client's requirement. The model load time is significantly higher than the CPU and VPU, but, in this scenario, it does not matter. The client wants to invest in a quality system. Thus, it is recommended to go with FPGAs such as Intel Arria 10.

Scenario 2: Retail

Client Requirements and Potential Hardware Solution

Look through the scenario and find any relevant client requirements. Then, suggest a potential hardware type and explain how this hardware would satisfy each of the requirements.

Which hardware might be most appropriate for this scenario? (CPU / GPU / VPU / FPGA)

CPU (Central Processing Unit)

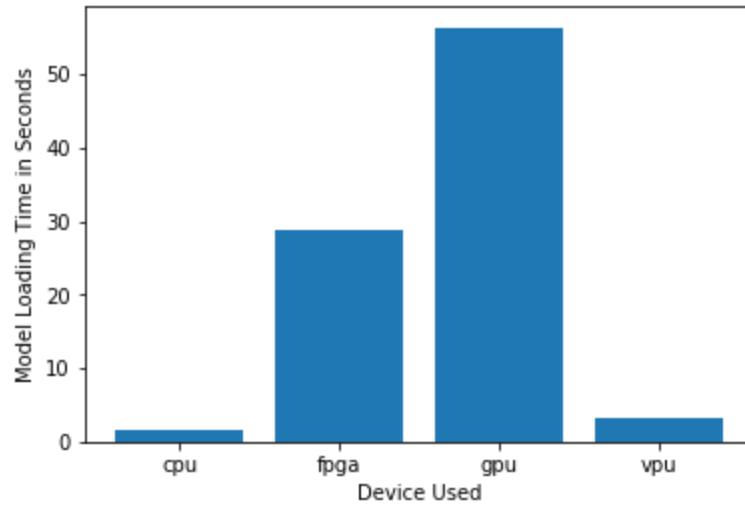
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
The client does not have money to invest in a new hardware.	<i>The existing systems with i7 CPU can be used, thus reducing the cost of buying a new system.</i>
<i>The client has a modern processor at every checkout with an i7 processor installed. These computers are used for small tasks only.</i>	<i>The load on CPU is minimal, hence the CPUs can be used to perform inference.</i>
<i>The client does not want the new hardware to increase his electricity bill.</i>	<i>CPUs take less electricity to run inference. CPUs will not drive up the electricity bill.</i>

Queue Monitoring Requirements

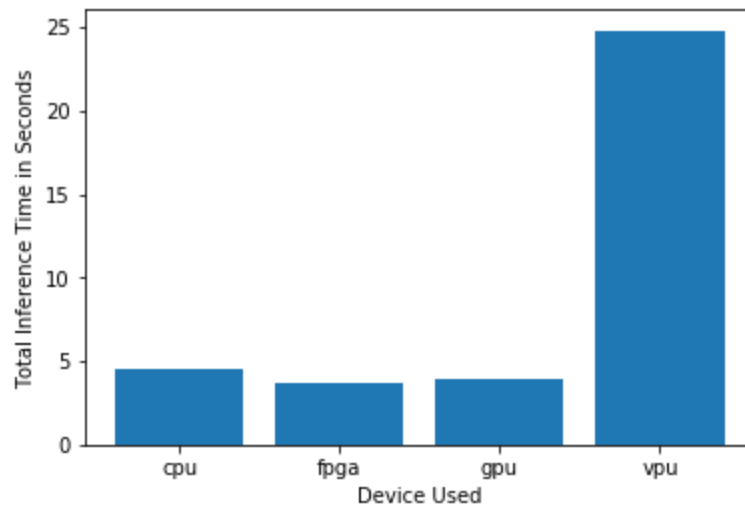
Maximum number of people in the queue	2 to 5
---------------------------------------	--------

Test Results

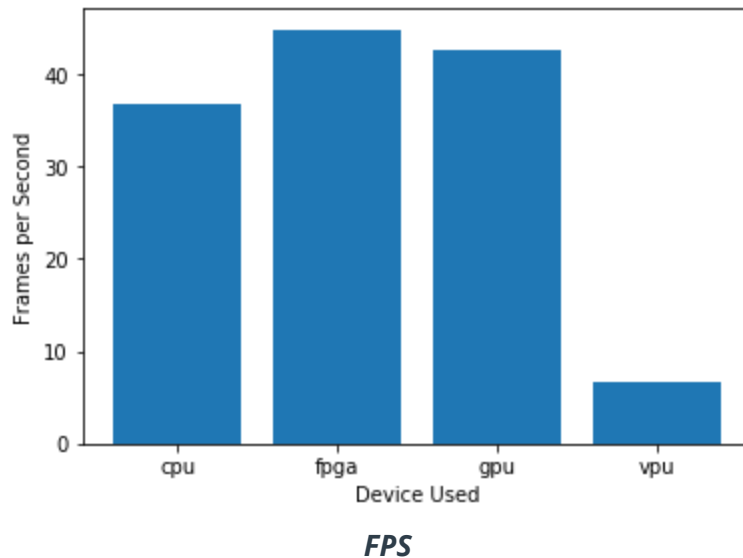
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



Final Hardware Recommendation

Now synthesize your points from above and provide a brief write-up describing why the chosen hardware is the best choice for this scenario. Be sure to discuss the client's requirements, the test results, and how these relate to one another (e.g., perhaps one of the devices performed better than the rest, but does not meet one of the client's requirements).

Write-up: Final Hardware Recommendation

Compared to GPU and VPU, the CPU has comparatively good inference time and FPS count. The CPU has best Model Load time if compared to its counterparts. The client has desktop class CPUs already installed at every checkout. Thus it will not drive up the electricity bill. Hence, CPU is to go-to choice for the edge system.

Scenario 3: Transportation

Client Requirements and Potential Hardware Solution

Look through the scenario and find any relevant client requirements. Then, suggest a potential hardware type and explain how this hardware would satisfy each of the requirements.

Which hardware might be most appropriate for this scenario?
(CPU / IGPU / VPU / FPGA)

VPU (Visual Processing Unit)

Requirement Observed
(Include at least two.)

How does the chosen hardware meet this requirement?

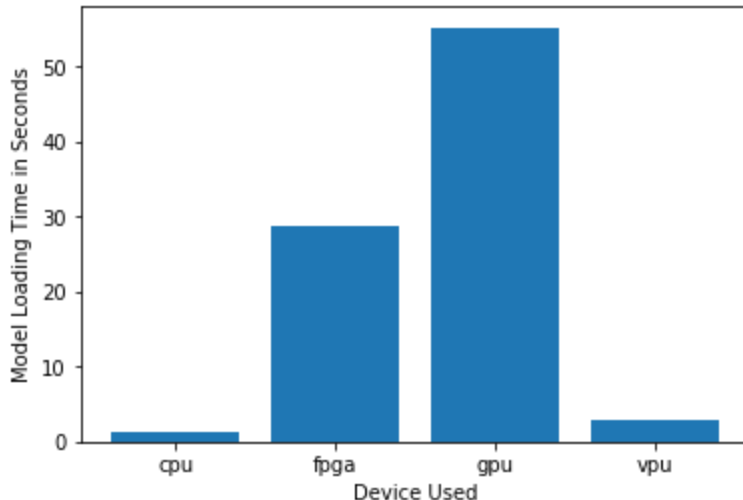
<i>The client has All-In-One PCs located in a nearby security booth. The CPUs in these machines are currently being used to process and view CCTV footage for security purposes and no significant additional processing power is available to run inference.</i>	<i>The VPUs can be plugged into those systems to improve upon the CPU inference.</i>
<i>The client's budget allows for a maximum of \$300 per machine.</i>	<i>The VPUs cost less compared to FPGA and GPUs.</i>
<i>The client would like to save on future power requirements.</i>	The VPUs are very power efficient. The NCS (Neural Comput Stick) 2 can run on less than 1W of power.

Queue Monitoring Requirements

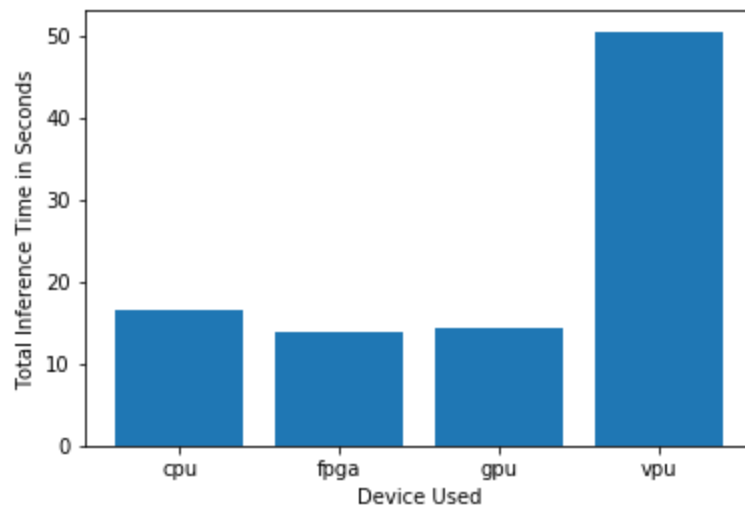
Maximum number of people in the queue	7 to 15
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

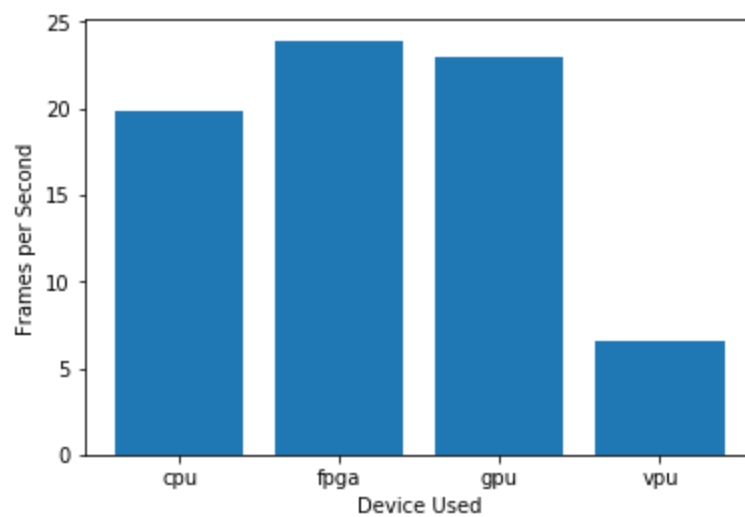
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



FPS

Final Hardware Recommendation

Now synthesize your points from above and provide a brief write-up describing why the chosen hardware is the best choice for this scenario. Be sure to discuss the client's requirements, the test results, and how these relate to one another (e.g., perhaps one of the devices performed better than the rest, but does not meet one of the client's requirements).

Write-up: Final Hardware Recommendation

The VPUs such as Neural Compute Stick, are portable, powerful plug and play machines designed specifically for visual processing. VPU require very less time to load models. When we look at the FPS count, the VPUs fare better than almost every other option. Due to budget and power requirement, Inference time can be ignored for the client's request.