

Choose the Right Hardware

By: Shahzeb Anwar

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

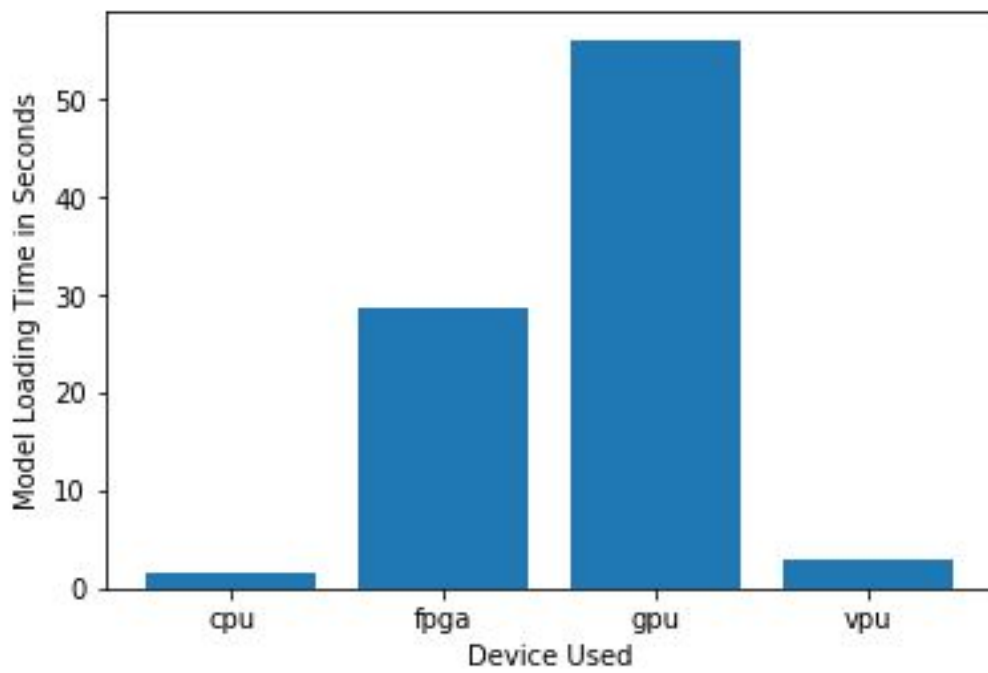
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Factory has vision camera installed with frame rate between 30-35 FPS on each belt	Due to high frame rate, latency will be a major factor of concern therefore using an FPGA can do inference a lot faster since there's less hardware latency involved as it doesn't send the output back to CPU via PCI bus which saves time.
Client is financially well and focuses more on reliability than cost	Although FPGA are costly but seeing the financial status of the client, it would be suitable to with FPGA
Flexible and long lasting(5 to 10 years) solution is required as the client wants to repurpose the system to solve other issues as well	FPGAs are highly flexible due to their reprogrammability and have a long life.

Queue Monitoring Requirements

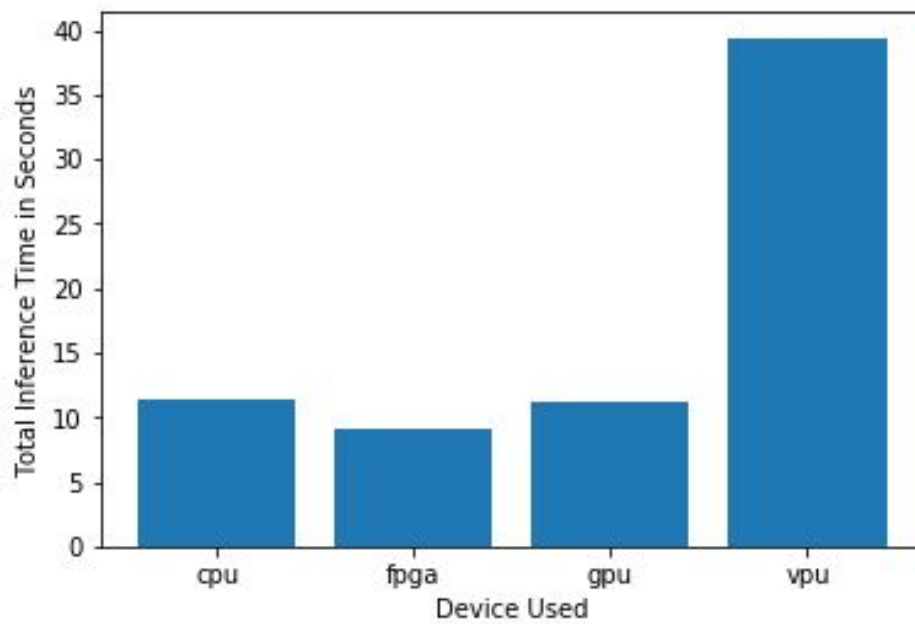
Maximum number of people in the queue	2
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

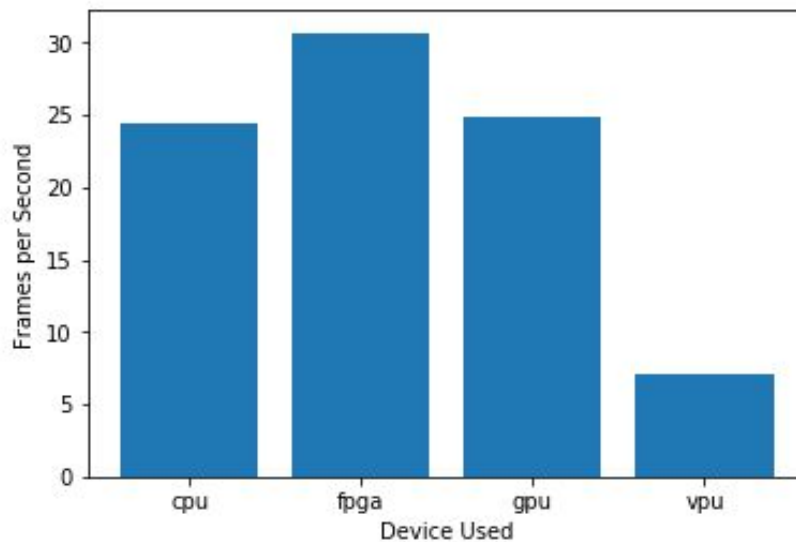
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

- Since FPGA has the fastest inference time as evident from the inference time graph hence it will best fulfill the client's demand.
- Client needs to inference at 25-30 fps which can be addressed by FPGA
- FPGA is most flexible devices that can be repurposed since it is reprogrammable

Final Recommendation: **FPGA**

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 / IGPU / VPU / FPGA)

CPU

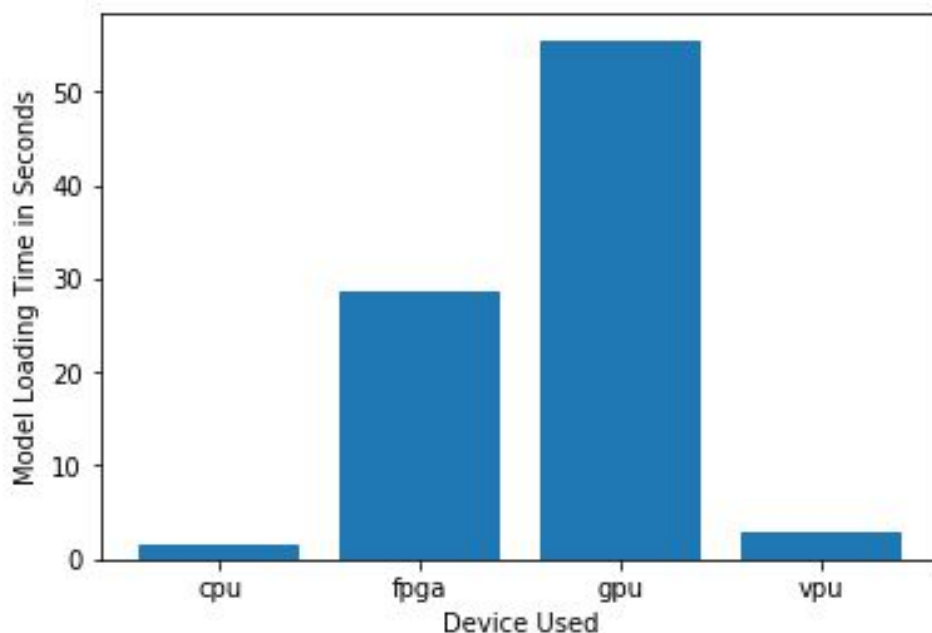
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>Checkout counters are already equipped with modern core i7 processors</i>	<i>Since checkout counters are already equipped with i7 processors and are not performing computationally expensive task, it would be good to use them for inference</i>
<i>Client can't invest much money</i>	<i>Since client can't spend much on the system therefore this would be good if we can use the i7 processors of the existing systems at checkouts</i>

Queue Monitoring Requirements

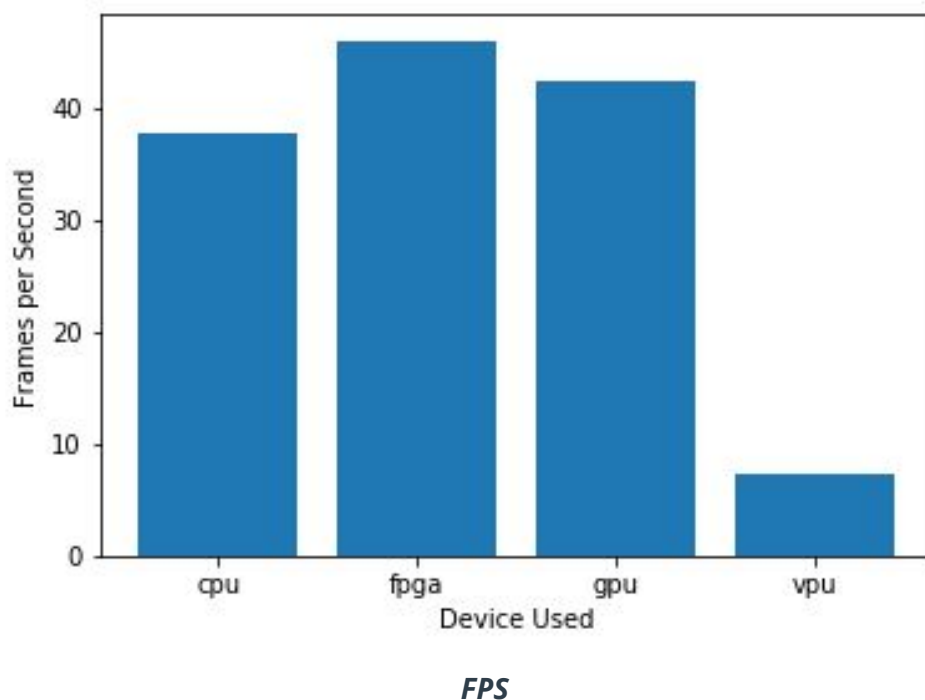
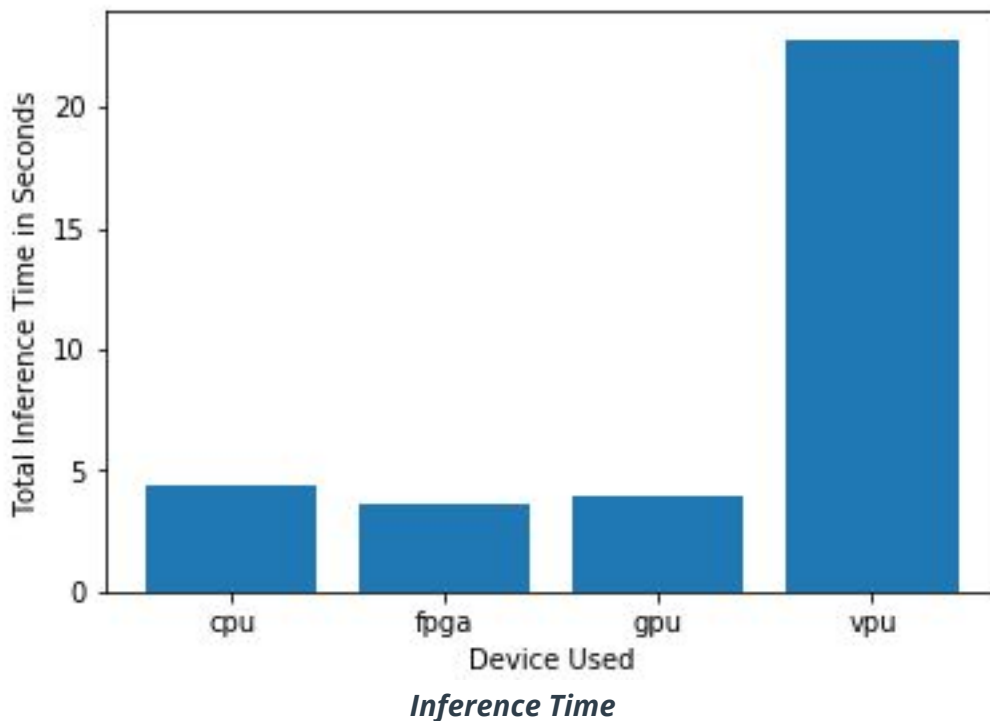
Maximum number of people in the queue	2-5
Model precision chosen (FP32, FP16, or Int8)	FP32

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



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

- CPU is given inference time comparable to GPU and a lot faster inference than a VPU and that too without costing any extra penny in processing hardware on the other hand it can infer on much higher fps and slightly less fps than GPU therefore CPU will be the best choice in this scenario

Final Recommendation: **CPU**

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

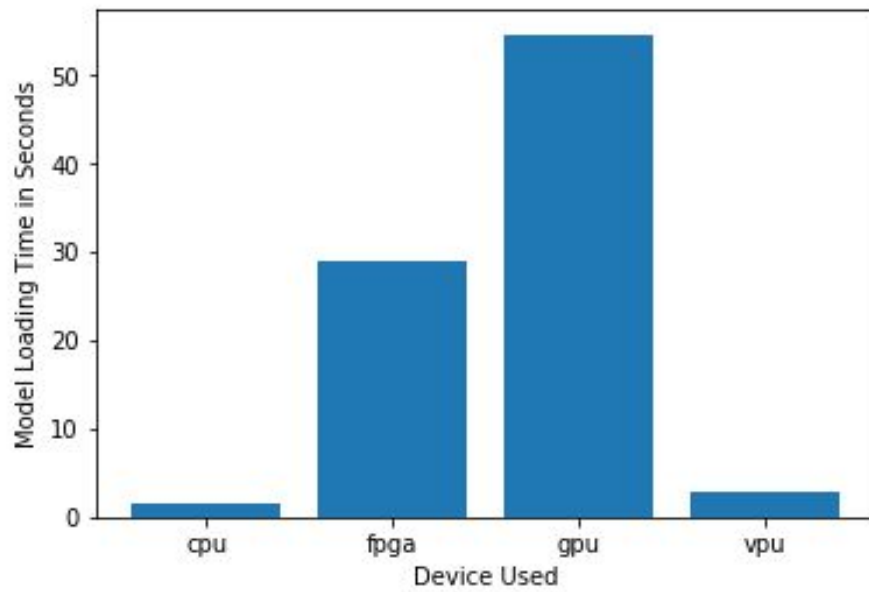
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
CPU in the installed system is preoccupied with CCTV jobs and there is no extra computing power is available to process the queue	Since the CPU is already processing CCTV jobs, the queue management can be done by attaching an NCS-2 stick which can be used on USB ports. Infact the optimal results will be obtained if we use multiple NCS-2 sticks (03 NCS-2 is recommended)
300\$ budget is available	The only devices that falls within this range is NCS-2 which costs around 70\$ all other CPUs, GPUs and FPGA falls outside of the budget

Queue Monitoring Requirements

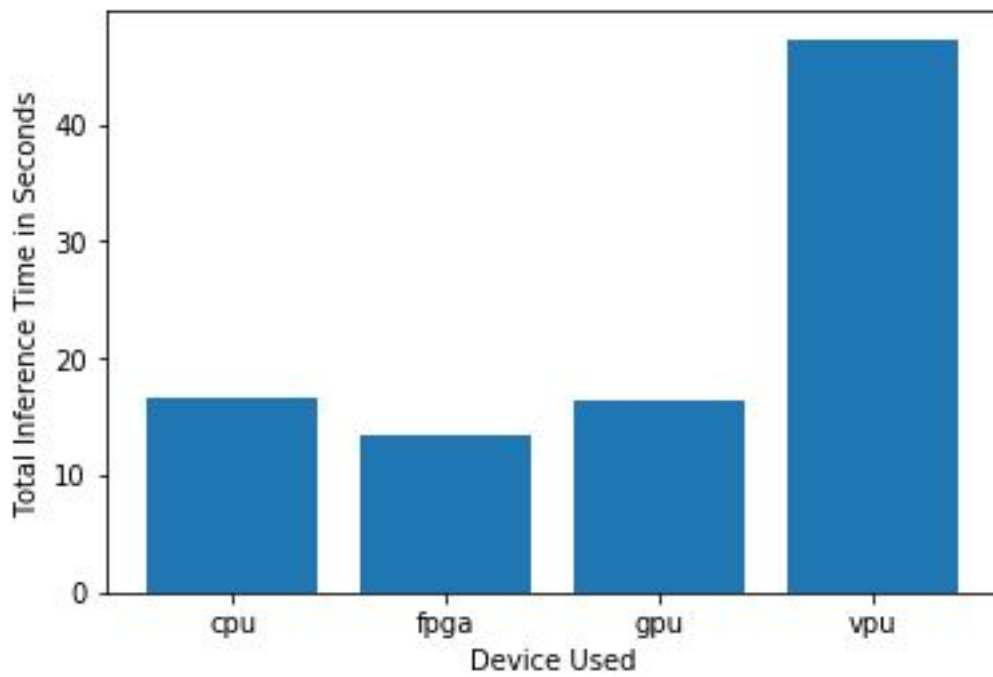
Maximum number of people in the queue	7-15
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

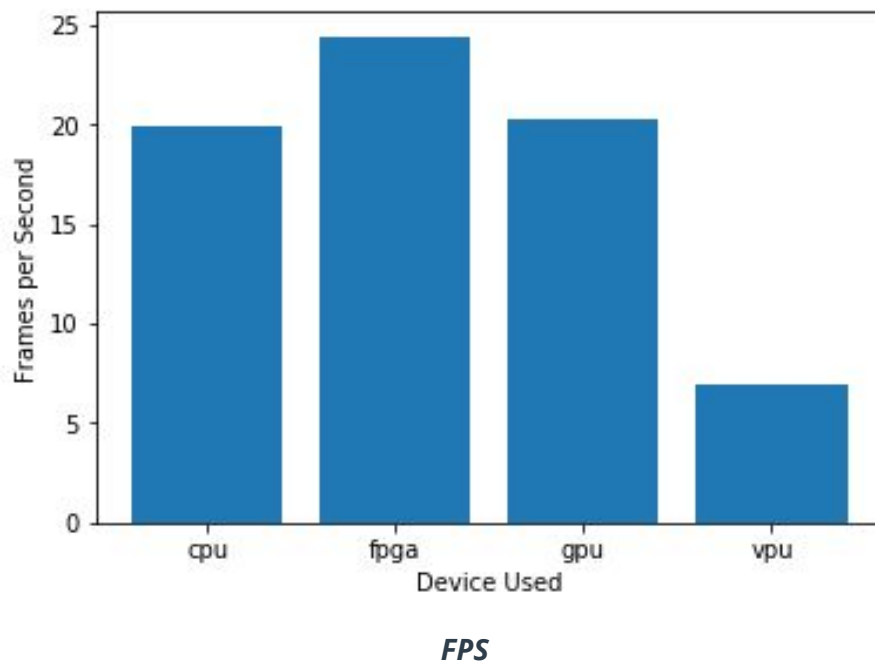
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



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

- VPU is given taking longer inference time as well as low FPS as compared to CPU & GPU
- In ideal scenario, A GPU or GPU will be good to go with but seeing the cost factor, VPU is the only option left
- VPU performance can be further increased if multiple NCS-2 are used in Parallel

*Final Recommendation: **VPU(e.g. NCS-2)***