

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

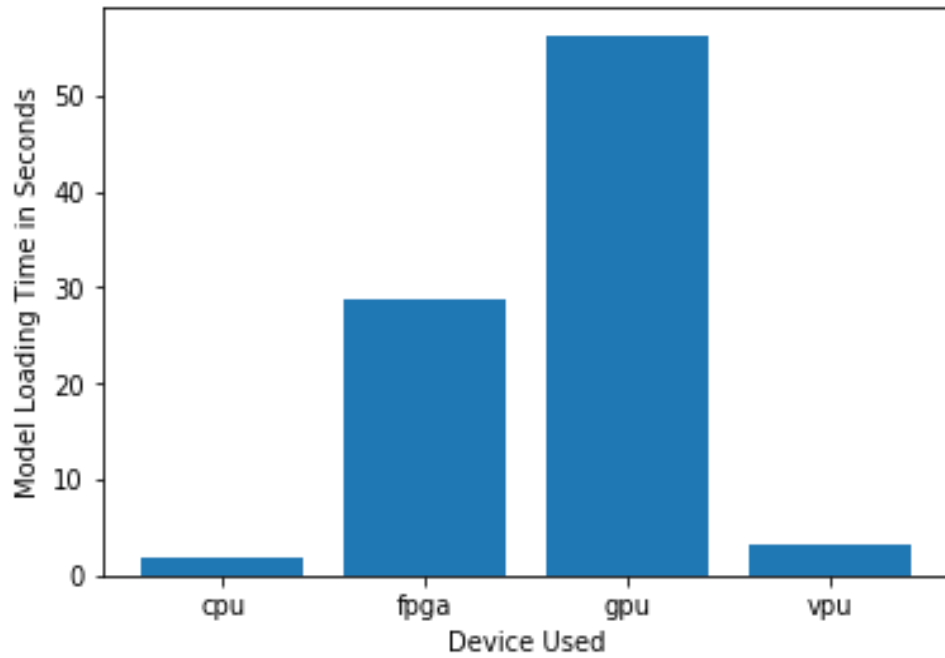
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device.	<i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
Based on client's financial position, the client can easily afford an expensive but reliable and efficient device to run inferences on the video frames.	Truly the FPGAs are much more expensive devices than their counterparts but are best suitable in this kind of scenario
Due to flexible reprogramming capability of the FPGAs, the client's business need of using the same device to address the defects in the product itself can be catered easily via introducing newer models.	Due to the usage of bitstream, the model can be reconfigured and will allow testing the defects products in the manufacturing line.

Queue Monitoring Requirements

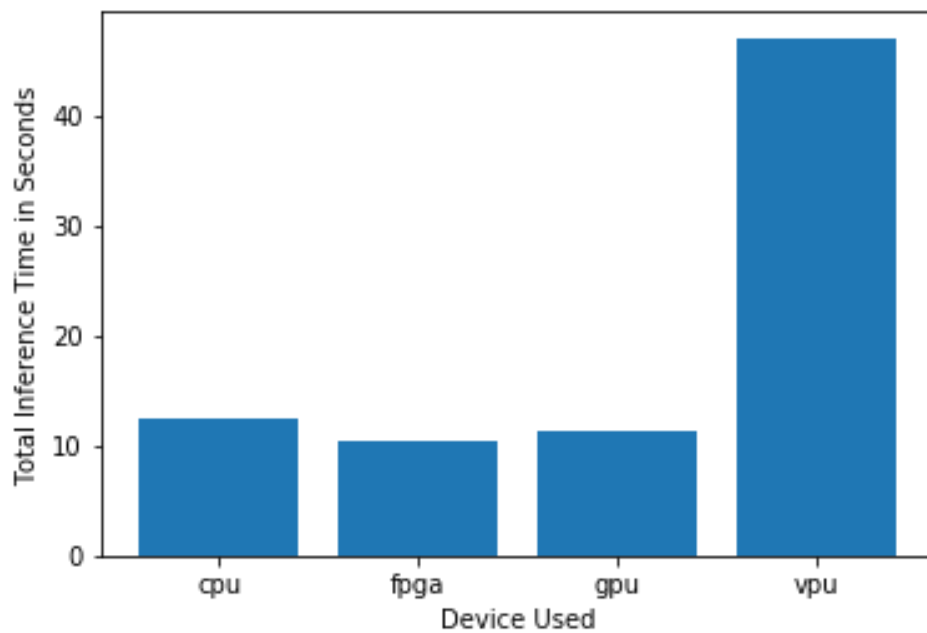
Maximum number of people in the queue	4
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

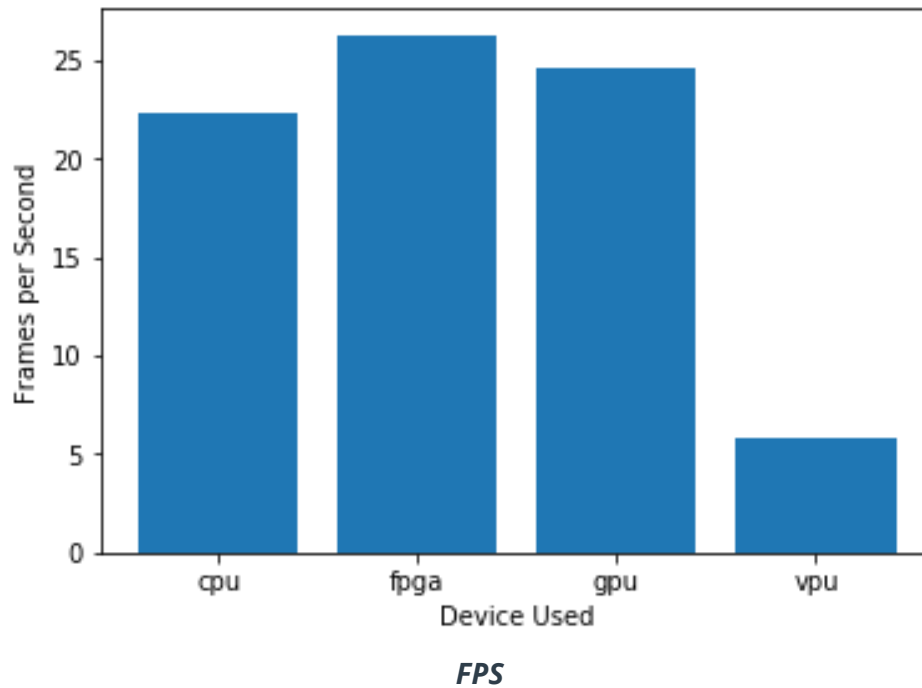
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

*-Final Hardware Recommendation: **FPGA***

-The reasons for selecting and recommending the proposed hardware are defined below .

*-After successfully executing the model on FPGA hardware, the comparative analysis showed the least **Inference Time in seconds** in comparison with other counterparts as it is primarily configured and suitable for the model and the case study. Though the cost of the hardware is indeed expensive but the service is definitely going to justify its importance and need for the tasks in hand and considering the client's strong financial position and long-term plans, the client seems to be in a good position to afford the required hardware investment.*

*-Lastly the **fps** is higher than all of the other hardware under consideration which means the final throughput will be the best in comparison with others.*

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)
<i>CPU</i>

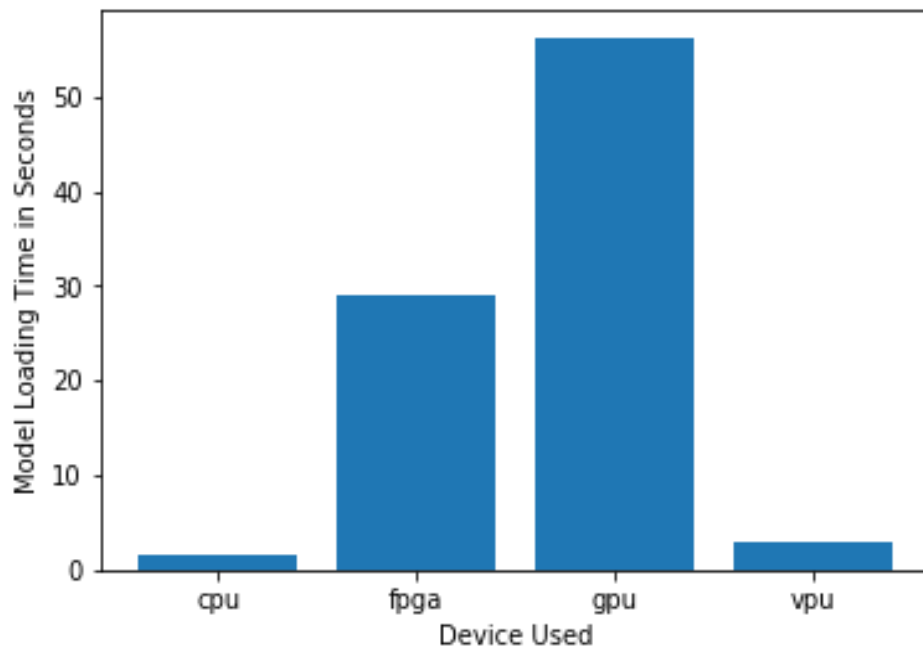
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Due to weak purchasing power, the client is not in a position to afford to invest in new hardware.	Since the machines at the checkout counters are already using i7 core processors, the CPU will be a good fit for running inferences.
Comparatively the queue lengths are relatively lesser than the average number of customers that visit the store.	Considering the lower computational need, CPU will be a good choice for solving the problem at hand and if in case the shorter processing time needs will rise, then GPU can be used for shorter inference time in comparison with CPU.

Queue Monitoring Requirements

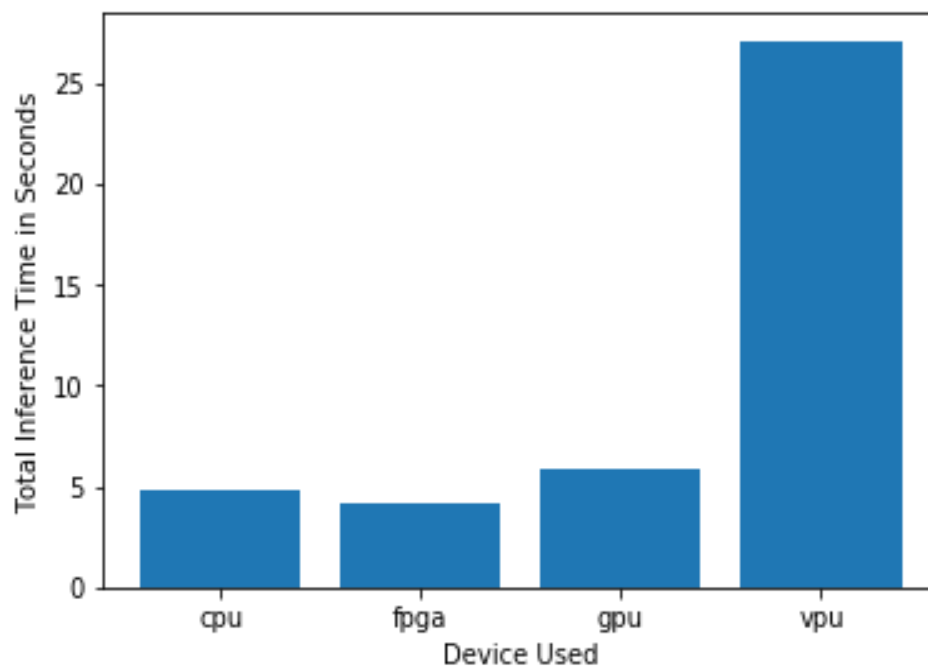
Maximum number of people in the queue	2
Model precision chosen (FP32, FP16, or Int8)	FP32

Test Results

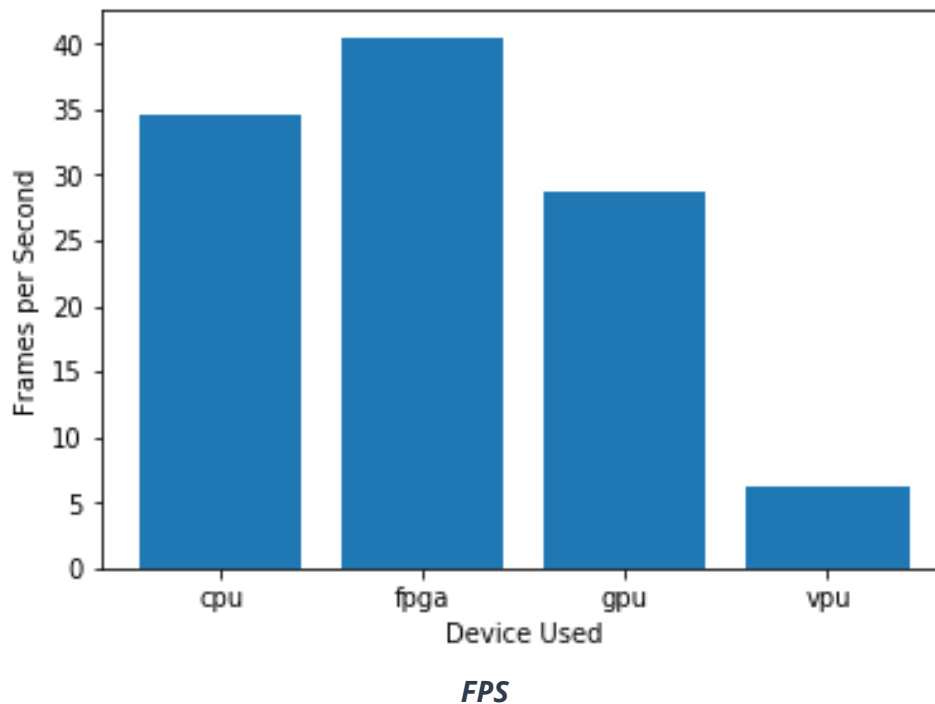
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



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

*-Final Hardware Recommendation: **CPU***

-Since the computers at the checkout counters already possess i7 core processors which is a high-end modern processor series and also currently there is no other use of the computational power, the CPU is the best choice considering the client's computational and low-budget requirements. .

*- CPU also enjoys high **frames per second** and is just next to FPGA. The inference time and loading time of the model are also relatively low.*

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

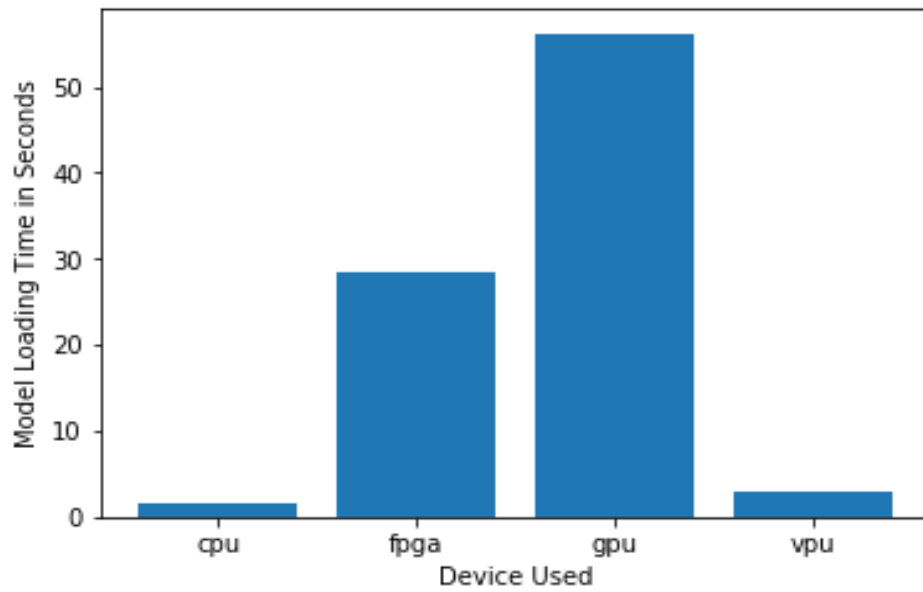
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
The project budget is \$300 per machine.	The price range of Intel's NCS2 is within the budget of the client as it costs around \$100.
Primarily the All in One PC processors are used for gathering the information without any additional hardware to run inference.	Since CPUs are not optimal hardware choice to cater the need of multiple frames processing therefore, multiple MYRIAD processors i.e NCS2 Sticks can be used as a suitable alternative, keeping in mind the need of minimum 3 sticks usage per machine to run inference with optimal results.
3 NCS2 Sticks	To run Inference on input frames coming from 7 CCTV cameras, Multi Device Plugin can be used. Since, relying on one won't be an optimal choice.

Queue Monitoring Requirements

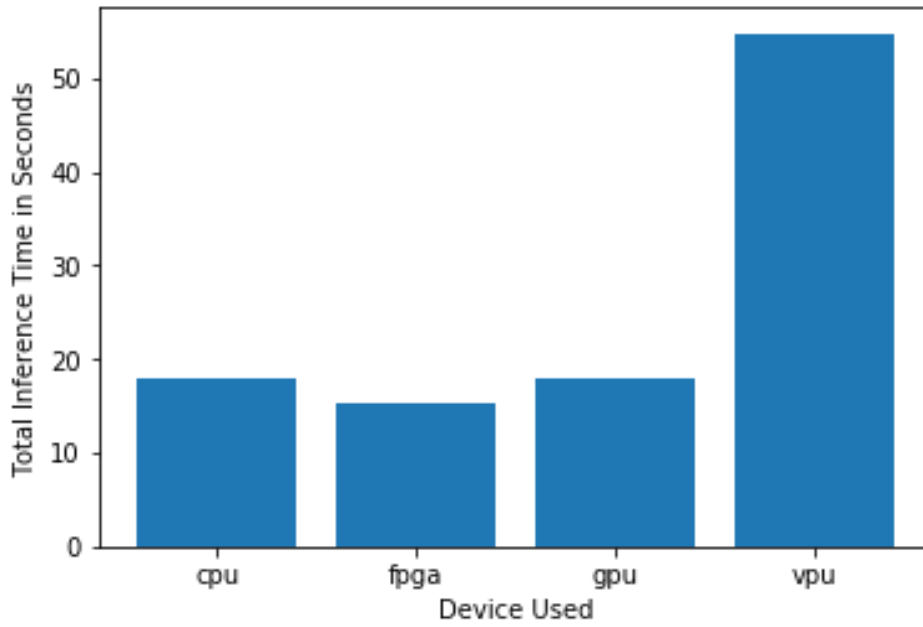
Maximum number of people in the queue	10
Model precision chosen (FP32, FP16, or Int8)	FP32

Test Results

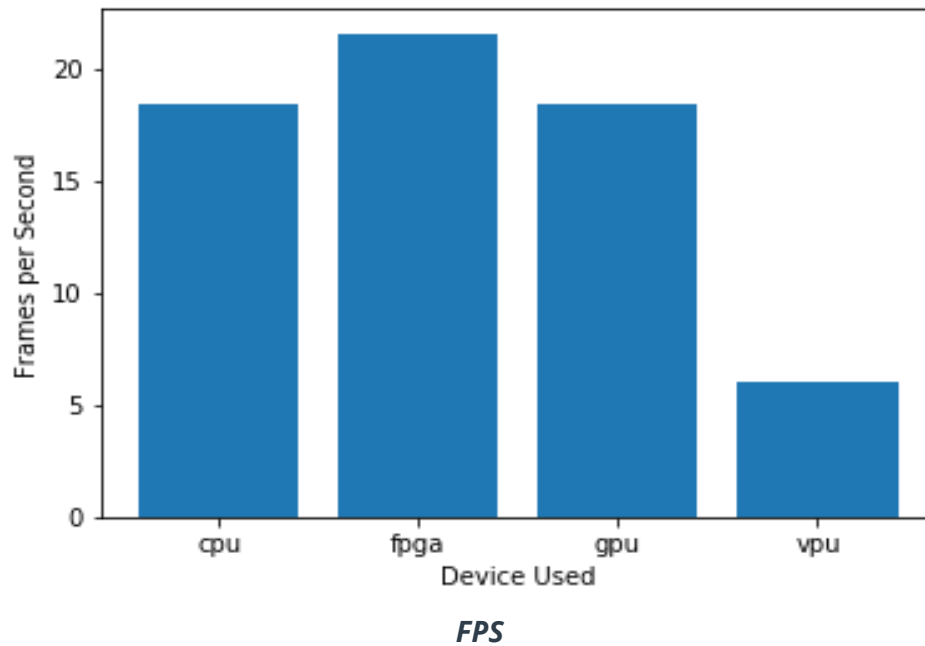
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



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

*-Final Recommendation: **VPU***

-Considering the resultant VPU throughput alone isn't as good as other hardware counterparts in terms of inference time and frames per second but in case of multiple VPUs usage to run inference this problem can be handled effectively and without crossing the client's budgetary limitations.