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

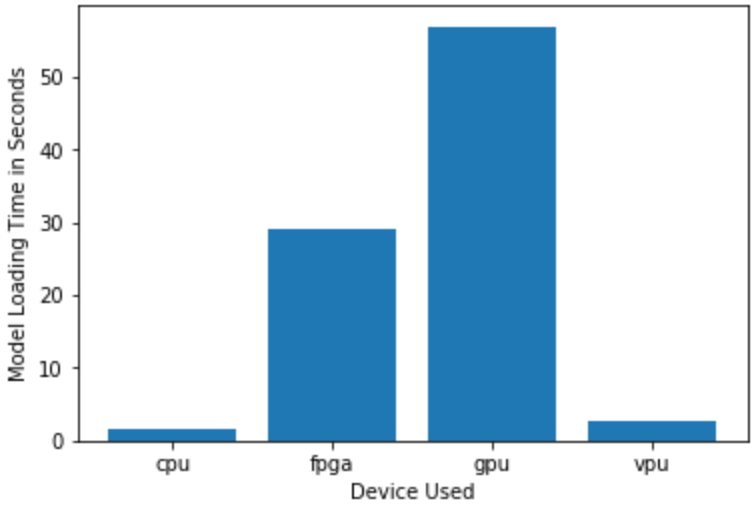
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
The client has enough revenue to buy an expensive device to run inferences on the frames.	*FPGAs are too expensive devices but are suitable for this scenario*
The client, later on, wants to use the same device to address another issue that is regarding the defects in the product itself and this would to reprogramming the FPGA according to the newer model.	*FPGA uses bitstream according to the model that can be reconfigured and this will help to accordingly test the defects products in the manufacturing line after addressing the workers issue.*

Queue Monitoring Requirements

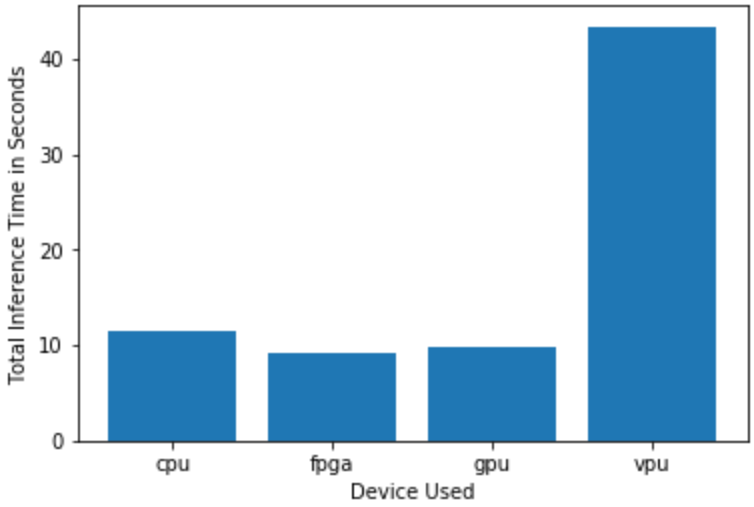
Maximum number of people in the queue	*4*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

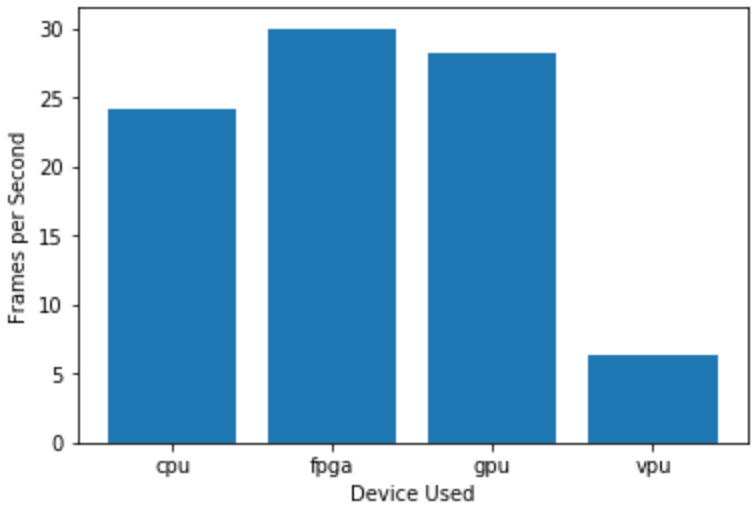
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



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
- *Final Recommendation:* ***FPGA*** - *The hardware chosen by me before running inference is the same final Hardware that I would like to recommend for the given scenario for the reasons mentioned below.* - *After running the model on FPGA it shows the least* ***Inference Time in seconds*** *as it is configured according to the model used and even if the price of the hardware is expensive it is going to be used for the same task and the client has the budget for purchasing the hardware.* - *Even the* ***frames per second*** *parameter are high which means the throughput from the hardware is the best among the rest of the hardware.*

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

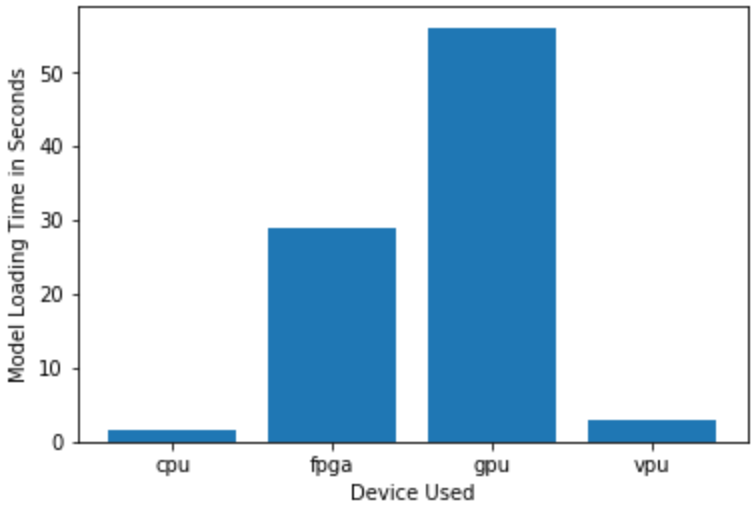
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
The client doesn't have enough budget to invest in new hardware.	*The store's checkout counters already have i7 core processor which is a good CPU to run inference.*
Even though the average number of customers that visit the store is more the queue lengths are relatively less.	*This is the reason a CPU inference will be more than sufficient and if less processing time is needed then an IGPU can be used to reduce inference time.*

Queue Monitoring Requirements

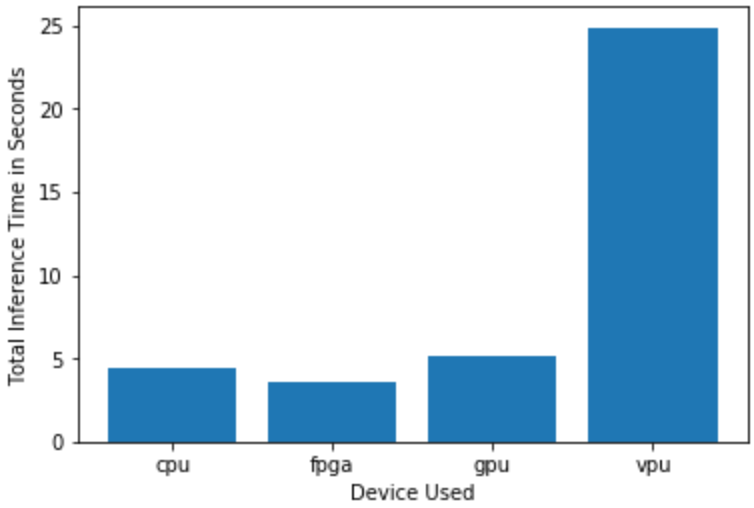
Maximum number of people in the queue	*2*
Model precision chosen (FP32, FP16, or Int8)	*FP32*

Test Results

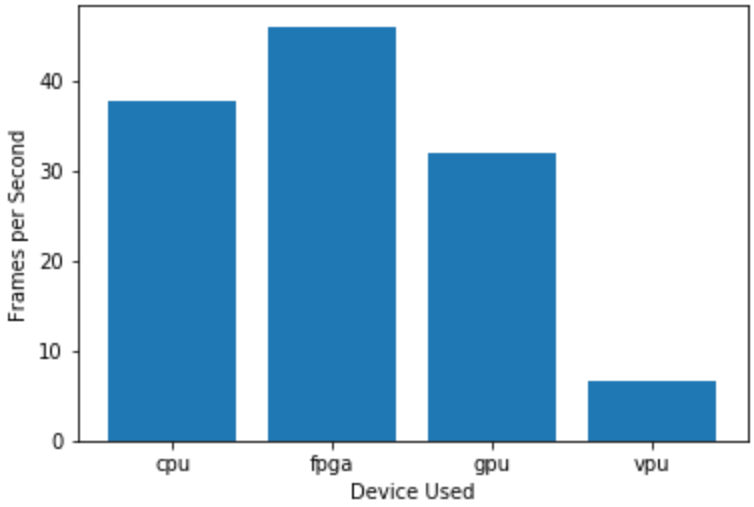
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



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
- *Final Recommendation:* ***CPU*** - *As the counter computers have i7 processor which is a high-end processor and also aren't being used for any high computational tasks CPU is the best option as the client doesn't have any more amount to spend on buying hardware.* - *Through graphs, we can also see that high FPS is achieved by the CPU 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

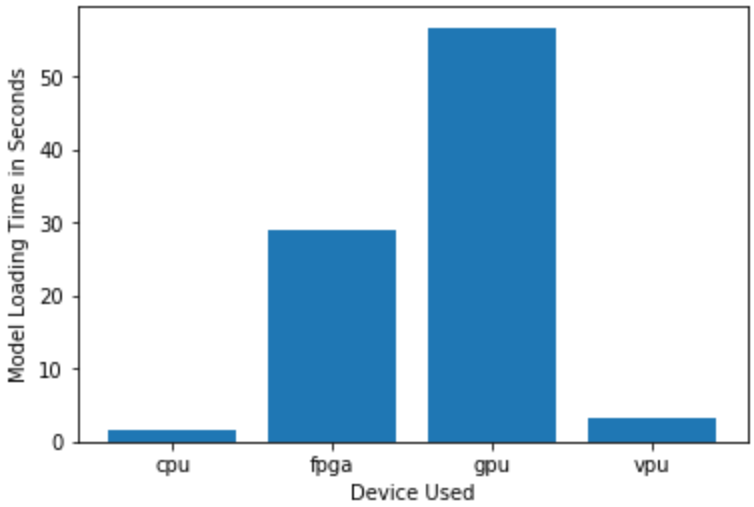
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
The client's budget allows to spend $300 per machine.	*The Intel's NCS2 is within the price range of client as it costs around $100.*
Currently the processors of the All in One PCs is used for gathering the information without any additional hardware to run inference.	*As there is multiple frames of input coming to the PC and processing it fairly on CPU won't be optimal so multiple MYRIAD processors i.e NCS2 Sticks can be used as per machine minimum 2-3 sticks can be used to run inference.*
2-3 NCS2 Sticks	*MULTI Plugin of the OpenVINO can be used to run Inference on frames coming from 7 CCTV cameras as relying on one won't be optimal.*

Queue Monitoring Requirements

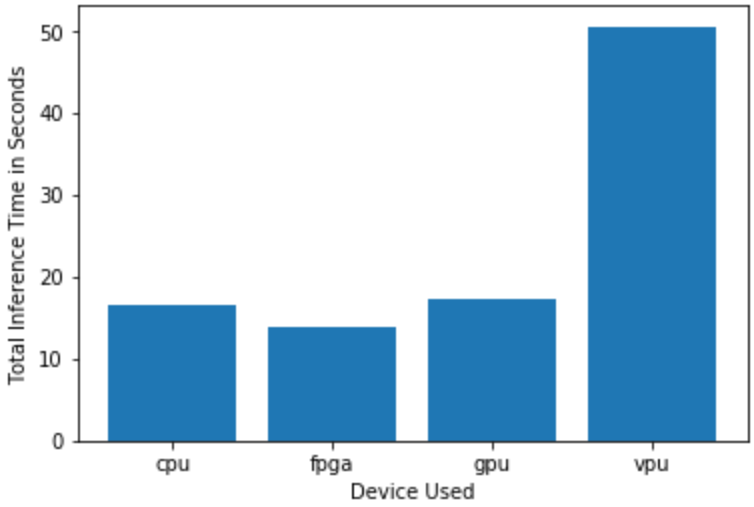
Maximum number of people in the queue	10
Model precision chosen (FP32, FP16, or Int8)	*FP32*

Test Results

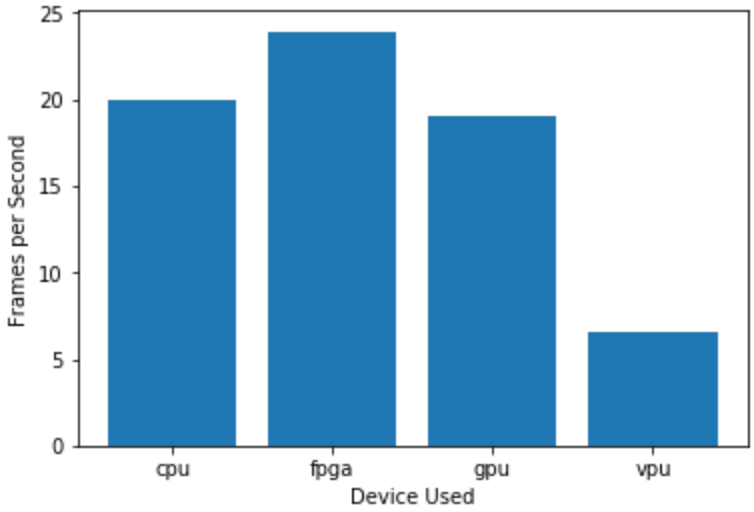
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



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
- *Final Recommendation:* ***VPU*** - *Even though the throughput considered of VPU alone isn't as good as other hardware in terms of inference time and frames per second but when multiple VPUs are used to run inference this overhead will get reduced and it falls under the budget that the client is willing to spend on each machine.*