Choose the Right Hardware

Proposal Template

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

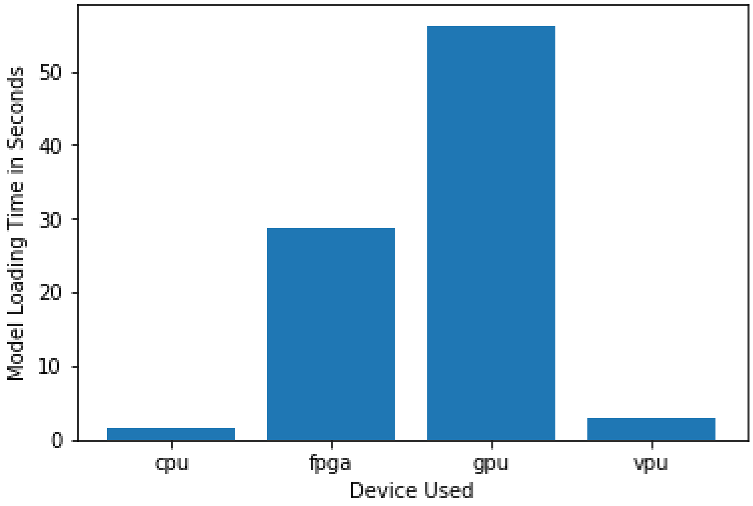
Requirement Observed **(Include at least two.)**	**How does the chosen hardware meet this requirement?**
Client requires a vision camera with 30-35 FPS to be connected to the system 5 times per second image processing speed for people counter application.	The high performance comes from the ability to run many sections of the FPGA in parallel.When running a neural network, we run the whole thing on the FPGA so the FPGAs don't go off-chip for the memory. This is faster than sending the output back to the CPU over the PCIe bus.
Client requires minimum possible inference time and flexible,reprogrammable FPGA for flaws detection in chips.	**Field-Programmable Gate Arrays (FPGAs)** are chips designed with maximum flexibility, so that they can be reprogrammed as needed in the field (i.e., after manufacturing and deployment).
Cost and power is not constrained.	*FPGAs are comparatively expensive to other accelerators but company have no constraint on budget and power.*
Good durability for 5-10 years.	*FPGAs are robust and have a long life span.*

Queue Monitoring Requirements

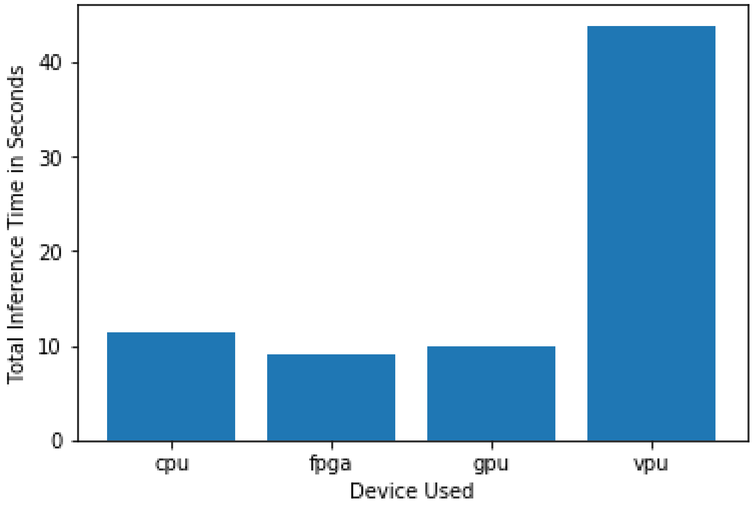
Maximum number of people in the queue	*2*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

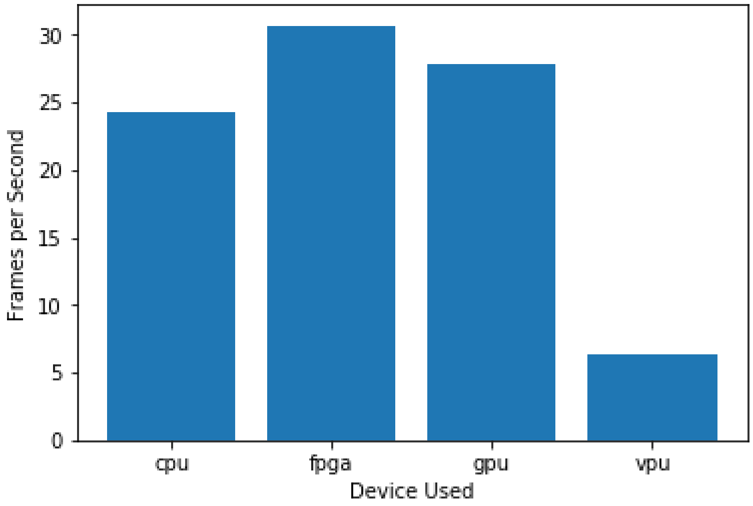
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
FPGA is most suitable , frames per second rate is highest for FPGA also inference time is lowest. Model loading time is greater in comparison to CPU and VPU but this trade off doesn't affect requirements. I would recommend 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?**

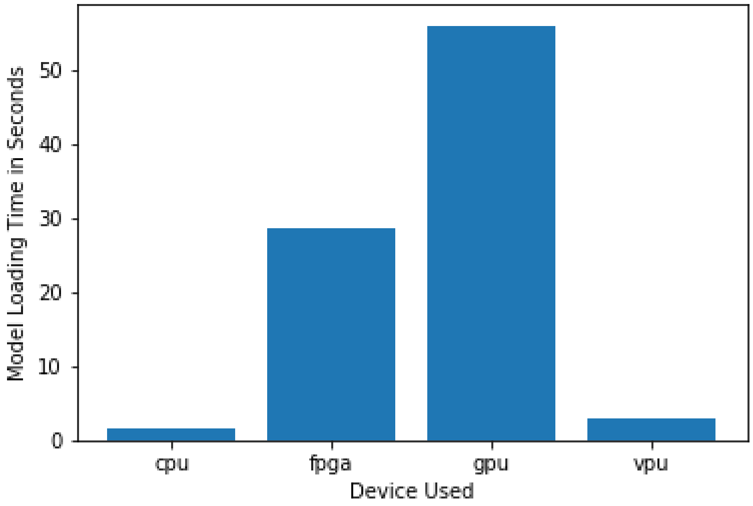
Already have a modern computer, each of which has an Intel i7 core processor, and don't have the budget to invest in hardware purchasing.	CPU is suitable for client desktop/PC use case.
Need only during weekend rush hours.	Computational complexity is not high ,so the core i7 processor can handle it.
Budget is limited.	Intel i7 core processor is enough and present in client’s desktops.No need to buy hardware.
Client wants to save money as much as possible on electric bills.	CPU TDP is low and suitable.It consumes less power in comparison to other hardware accelerators.

Queue Monitoring Requirements

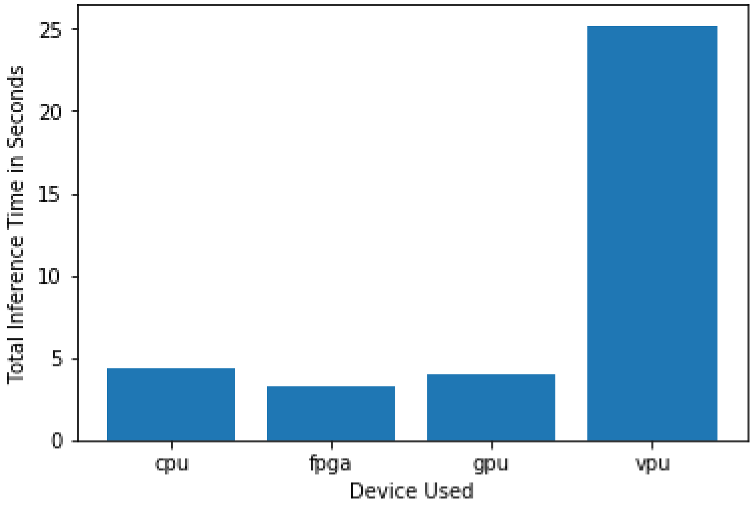
Maximum number of people in the queue	*2-5*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

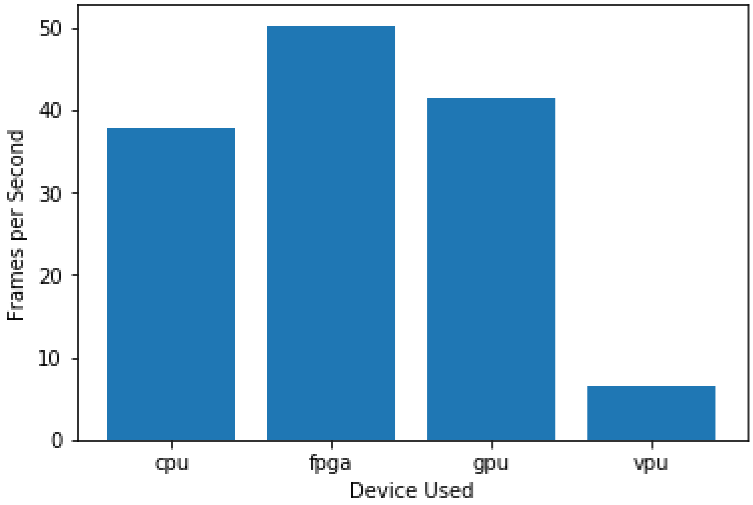
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
CPU is best for retail requirements. Inference time is comparable to FPGA and GPU. Model loading time is lowest. More than 30 FPS is good and suitable for requirements.

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

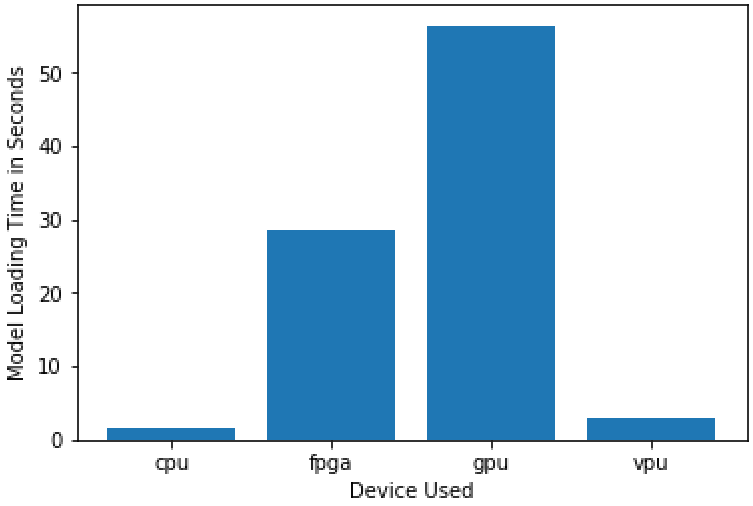
Requirement Observed (Include at least two.)	**How does the chosen hardware meet this requirement?**
Client needs significant additional processing power to run inference on machines.	VPUs are small, low-cost, low-power devices that can dramatically improve the performance of a system without the need to upgrade the other hardware.
Client's budget allows for a maximum of $300 per machine and has 7 machines.	NCS-2 cost around $ 70-100,it is under budget.The Neural Compute Stick 2 (NCS2) is a USB3.1 plug and play removable VPU for AI inferencing
Client wants to save as much as possible both on hardware and future power requirements.	Myriad X processors consume less than 2 watts power.

Queue Monitoring Requirements

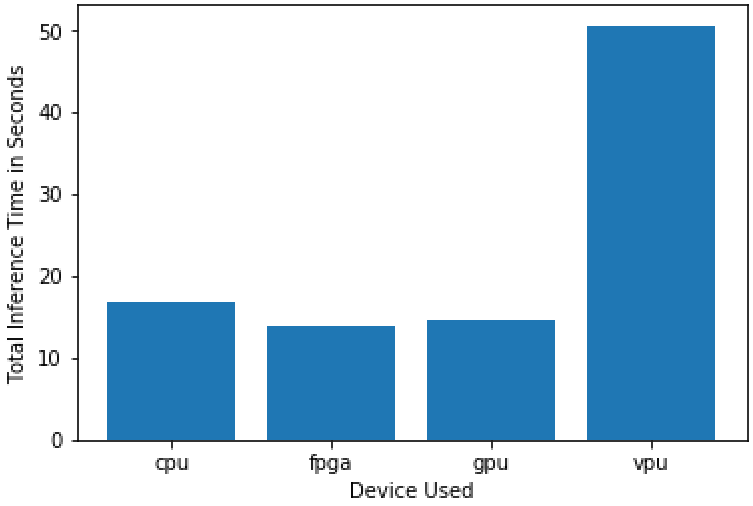
Maximum number of people in the queue	*7-15*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

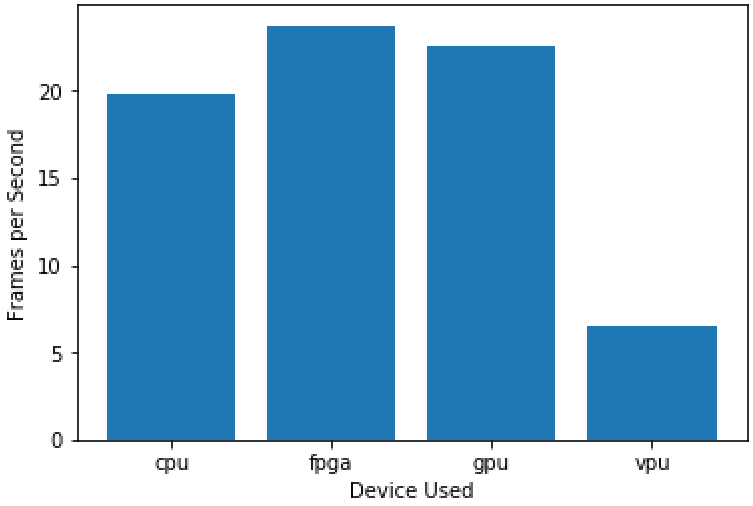
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



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
I would recommend VPU. Model loading time is minimum and comparable to CPU.Inference time is highest and FPS is lowest.VPU as accelerator will increase performance of CPU.