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?
The factory has a vision camera installed at every belt. Each camera records video at 30-35 FPS (Frames Per Second). The client would like the image processing task to be completed five times per second.	This aspect shows that latency is a major concern in this case. Once an FPGA is programmed with the bitstream required for this application can run the model with very high performance and give a very low latency. As an FPGA can run many sections of the chip in parallel and the ability of FPGA to not go off-chip for performing inference from the model would particularly be very useful for this kind of scenario. An FPGA also does not send the output back to the CPU using PCIe bus making the inference a lot faster.
The second issue the client has encountered is that a significant percentage of the semiconductor chips being packaged for shipping have flaws. These are not detected until the chips are used by clients. If these flaws could be detected prior to packaging, this would save money and improve the company's reputation.	To solve this issue the edge system would require to be able to deliver high performance. FPGAs would be a perfect fit for doing so as they provide a very high performance. Another aspect of FPGA which might help in this scenario would be its ability to be reprogrammed on the field, which can help improve inferences.
To be able to detect chip flaws without slowing down the packaging process, the system would need to be able to run inference on the video stream very quickly.	FPGAs can also be used as hardware accelerators speeding up the inference. THe parallel processing and no need to go off chip further provide an added advantage in performing the inference faster.

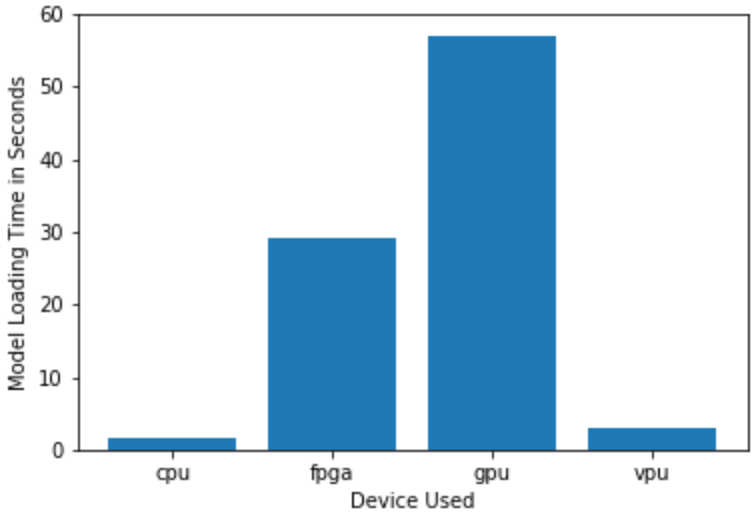
Additionally, because there are multiple chip designs and new designs are created regularly the system would also need to be flexible so that it can be reprogrammed and optimized to quickly detect flaws in different chip designs.	This need of the client makes FPGA align to the required hardware. FPGAs are highly flexible, they are field programmable and can be reprogrammed as needed.
The client would ideally like it to last for at least 5-10 years.	FPGAs have a very long lifespan generally 10 years and thus could be used by the client.

Queue Monitoring Requirements

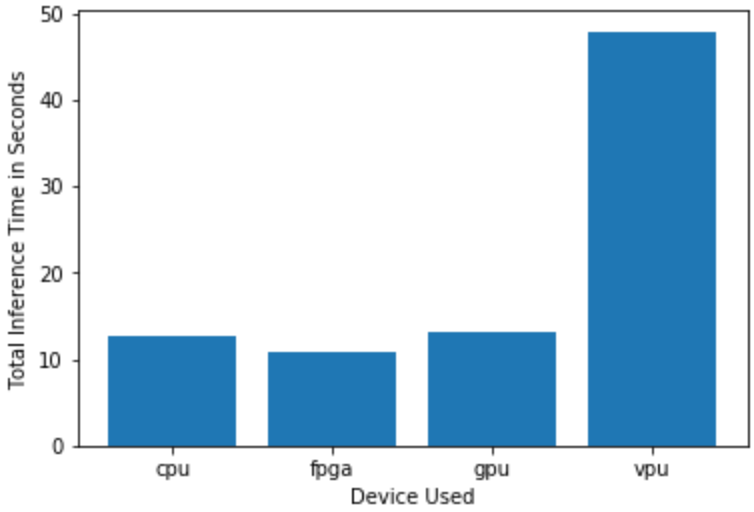
Maximum number of people in the queue	2
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

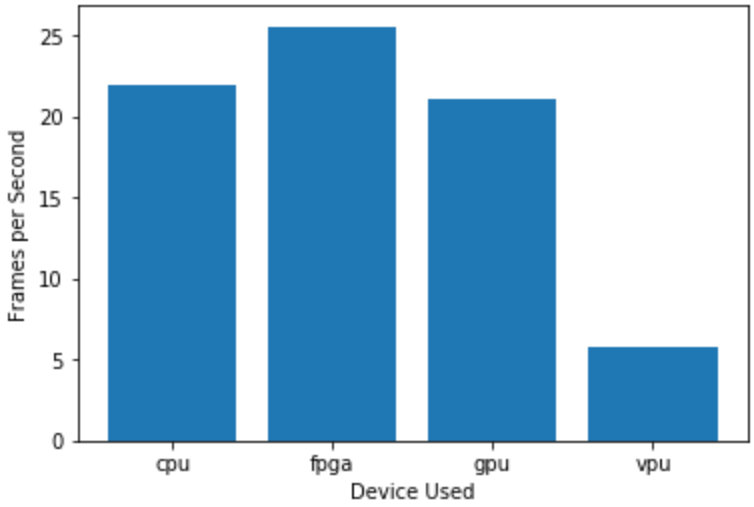
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
As shown in the inference time graph FPGAs take the least amount of time to perform inferences. The client needs inference to be performed fast so this FPGA can do a commendable job. We can see that FPGA also gives the highest FPS. The client requires 25 - 30 FPS which can be addressed by an FPGA. FUrther they are also field programmable and have a high life which is requested by the client. Thus, **FPGA would be a good choice** for this use case.

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

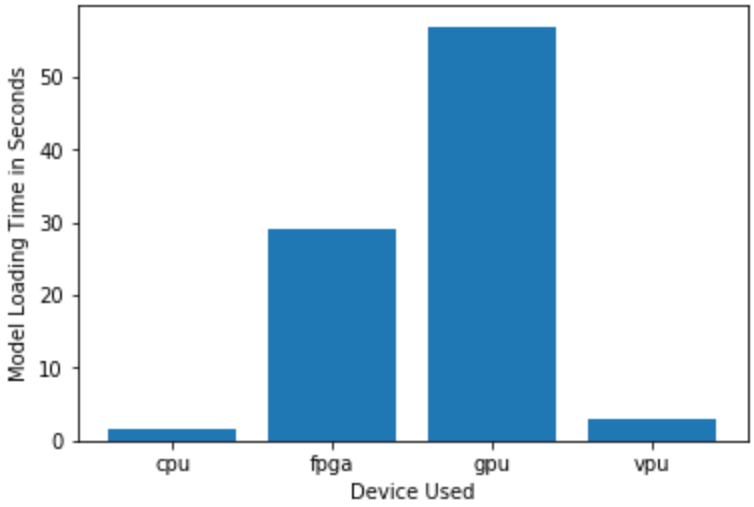
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Most of the store's checkout counters already have a modern computer, each of which has an Intel i7 core processor. Currently these processors are only used to carry out some minimal tasks that are not computationally expensive.	Since, the client already has a lot of CPUs which are not able to carry some computationally expensive tasks. THese CPUs with some new ones for the expensive tasks could be used.
The client does not have much money to invest in additional hardware.	Existing CPUs with some new ones can be used by the client can be made use of to reduce the costs.
The client would like to save as much as possible on his electric bill.	CPUs could meet the hardware requirements and also help save on the client's electric bill.

Queue Monitoring Requirements

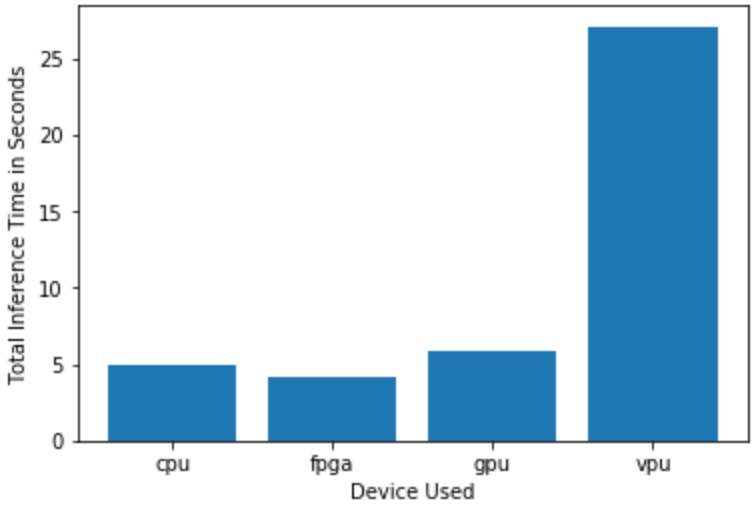
Maximum number of people in the queue	2 - 5
Model precision chosen (FP32, FP16, or Int8)	FP32

Test Results

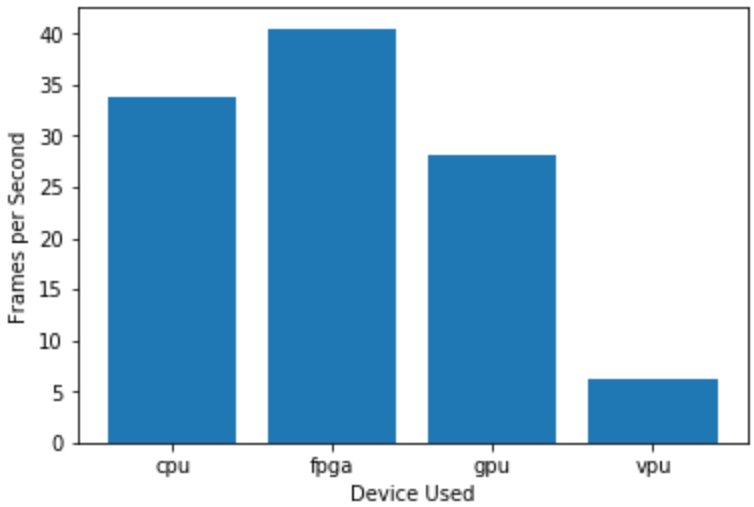
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



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 has a comparatively lower inference time than a GPU and VPU which is crucial for this application. A CPU can also make inferences on a good enough FPS for a retail store. It would also help save costs and electricity bill as requested by the client. Thus, **CPU would be a good choice** for this use case.

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

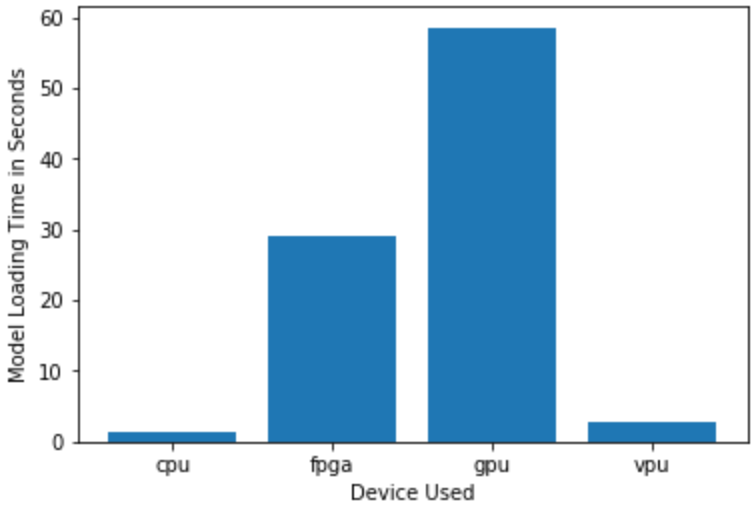
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
The CPUs in these machines are currently being used to process and view CCTV footage for security purposes and no significant additional processing power is available to run inference.	A VPU or NCS 2 can be plugged in a USB port and has a convenient plug and play kind of performance. This particularly is favourable as no more additional processing power is available to run inference.
The client's budget allows for a maximum of $300 per machine.	*A VPU or a NCS 2 stick costs almost $100 opposed to the comparatively higher costs of FPGA, CPU or GPU.*
The client would like to save as much as possible both on hardware and future power requirements.	A VPU or NCS 2 stick is designed to run on very low power. THe NCS 2 can run on just 1 W of power adhering to the clients needs.

Queue Monitoring Requirements

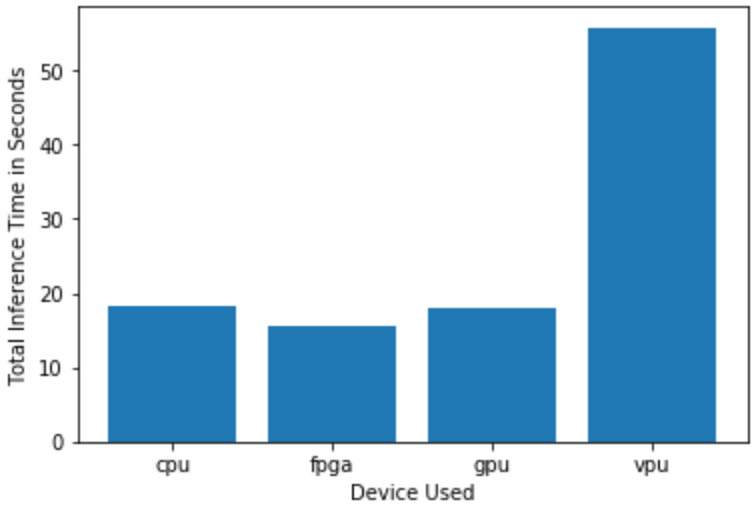
Maximum number of people in the queue	*7 - 15*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

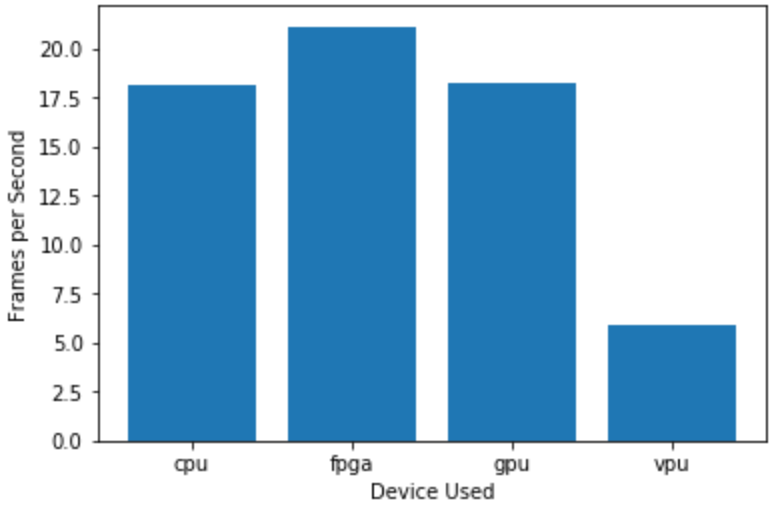
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
VPU provides a very high inference time and inference on very low FPS. A CPU or GPU would ideally be good for this if the budget and power scenario would be ignored. Since the client requires these aspects **VPU would be a good choice** for this scenario.