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?**
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.
Requirement of 5 FPS	*FPGAs meet the requirements.*
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. 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.	*FPGAs are reprogrammable.*
While Naomi Semiconductors has plenty of revenue to install a quality system, this is still a significant investment	*FPGA has a long lifespan of more than 10 years.*

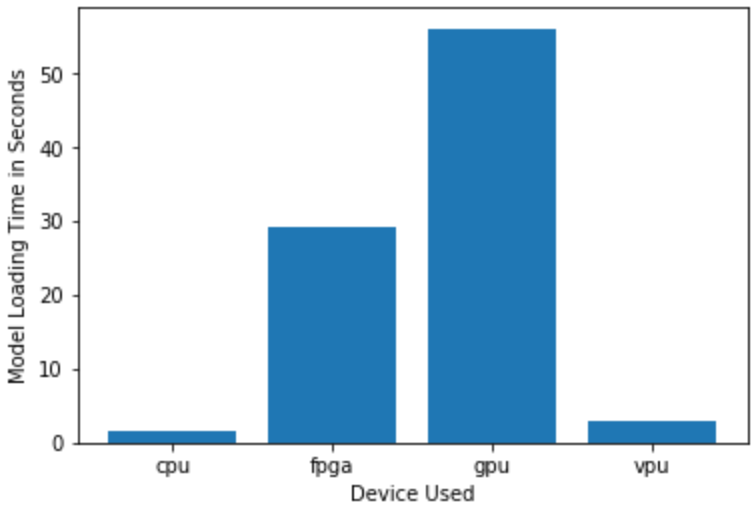
and they would ideally like it to last for at least 5-10 years.	

Queue Monitoring Requirements

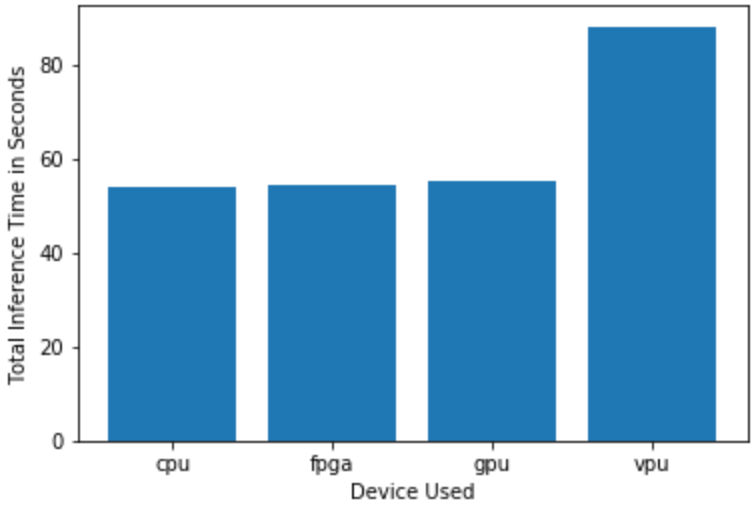
Maximum number of people in the queue	*Number of people in queue = {1: 1, 2: 1}* *Total = 2*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

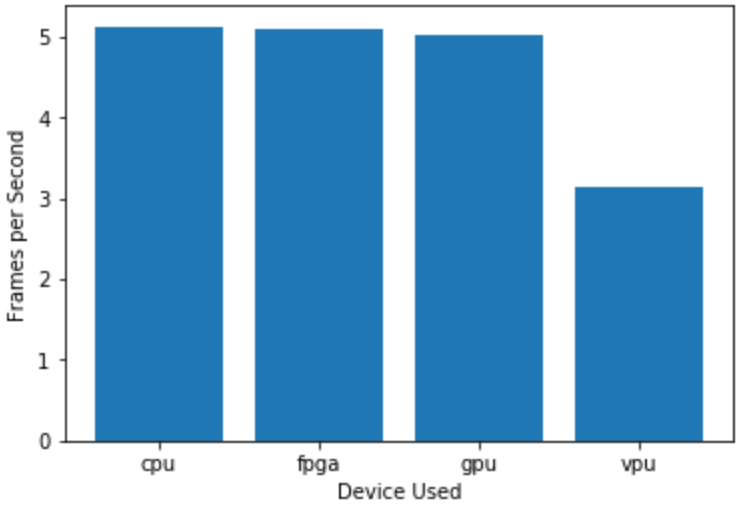
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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 reprogrammable and flexible and has high FPS and low inference and loading time which makes it a suitable requirement for MR VIshwas's business. It also has a long lifespan which will suit his needs.

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

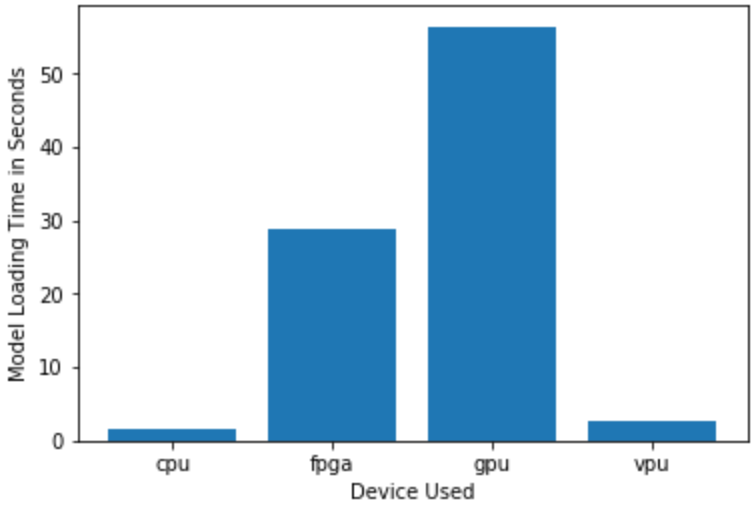
Requirement Observed (Include at least two.)	**How does the chosen hardware meet this requirement?**
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
Mr. Lin does not have much money to invest in additional hardware	Wouldnt have to spend money by using CPU
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.	*I7 core processor are good for running tasks during checkout.*
Mr Lin would also like to save as much as possible on his electric bill.	*Using a CPU would be helpful here.*

Queue Monitoring Requirements

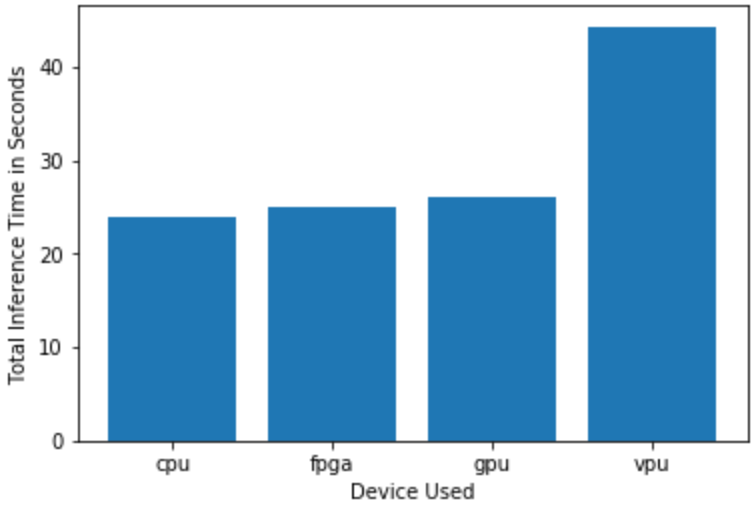
Maximum number of people in the queue	*2 to 5.*
Model precision chosen (FP32, FP16, or Int8)	*FP16, FP32*

Test Results

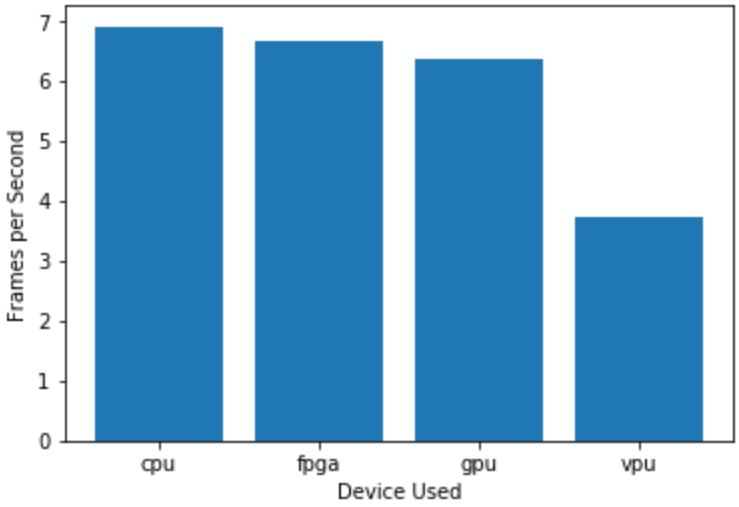
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
A CPU has low model loading time, low inference time and high FPS. So I think we can use CPU as and make some hardware changes if required. This would fulfill the need of saving money and also CPU has the necessary features in terms of performance.

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

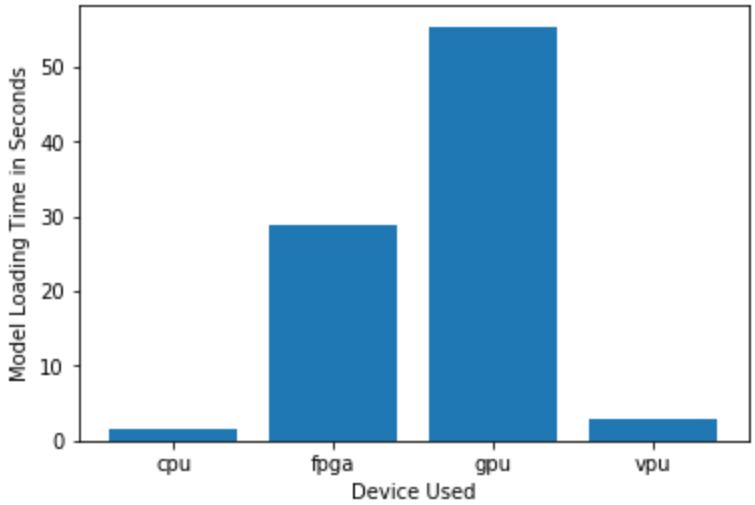
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
Ms. Leah's budget allows for a maximum of $300 per machine.	VPU's cost less than $300 per machine. Intel Movidius Neural Compute Stick 2 is affordable.
She would like to save as much as possible both on hardware and future power requirements.	VPUs do not consume a lot of power.

Queue Monitoring Requirements

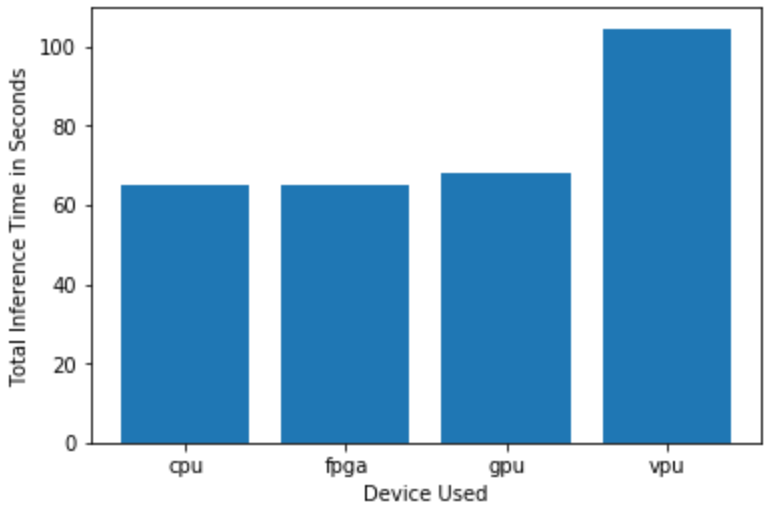
Maximum number of people in the queue	In peak hours they currently have over 15 people on average in a single queue outside every door in the Metro Rail. But during non-peak hours, the number of people reduces to 7 people in a single queue.
Model precision chosen (FP32, FP16, or Int8)	FP32, FP16

Test Results

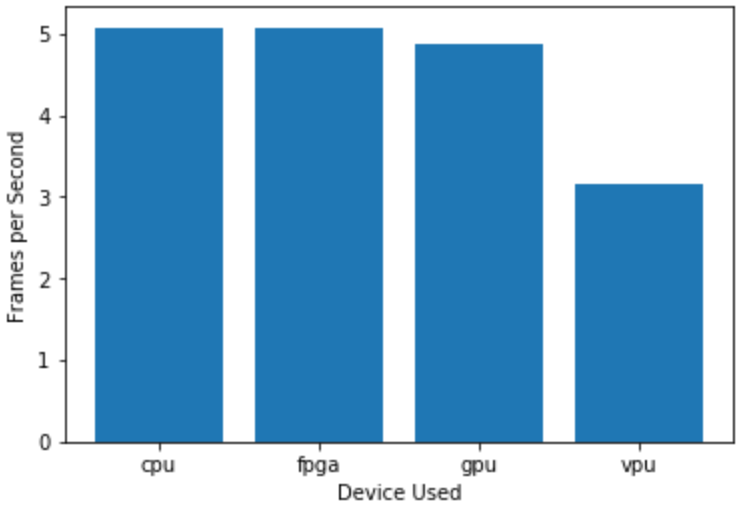
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
Taking the money and power aspect into account, I think the client should use a VPU as it doesn't cost much compared to other alternatives even though the inference time is high and FPS is low.