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 (Field Programmable Gate Arrays)

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. Vishwas, the client, would like to be able to reprogram and repurpose for different purposes. Because there are multiple chip designs—and new designs are created regularly—the system would need to be flexible so that it can be reprogrammed and optimized to quickly detect flaws in different chip designs.	*FPGA is specific enough to be re-programmed to handle each specific purpose. It is field re-programmable, that is can be reprogrammed for another specific purpose after being shipped.*
The company has enough budget and wants a solution with up to 5 to 10 years life span.	FPGAs that use devices from IoT Group have a 10 years availability guarantee from the start of production and it's expensive, which won't be a problem since the company has enough budget.
The company want to detect chip flaws without slowing down the packaging process, they want the system to be able to run inference on the video stream very quickly	*FPGAs are fast for a very significant improvement, like processing images 5 times of already existing 30-35 FPS.*

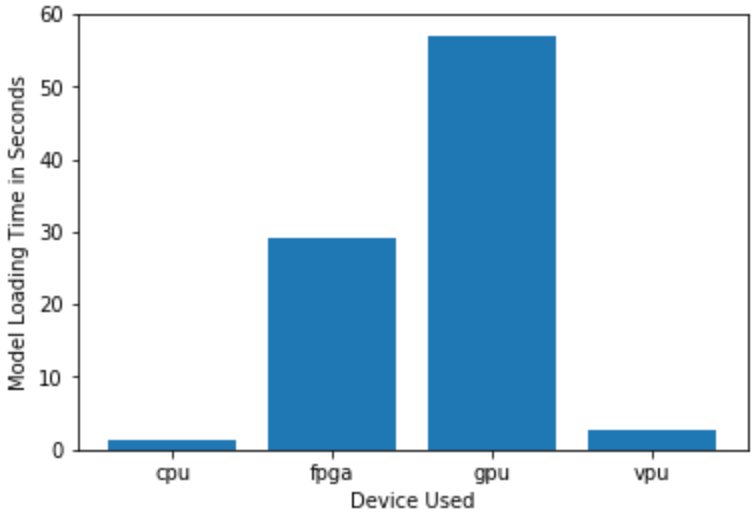
Queue Monitoring Requirements

Maximum number of people in the queue	*2*

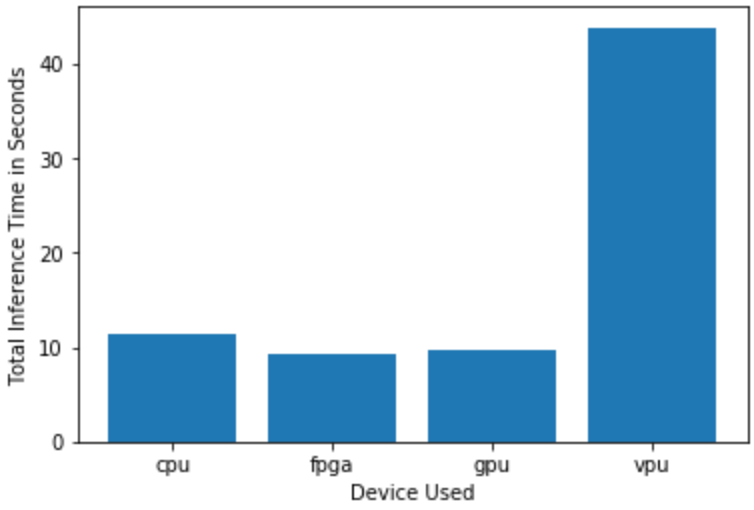
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

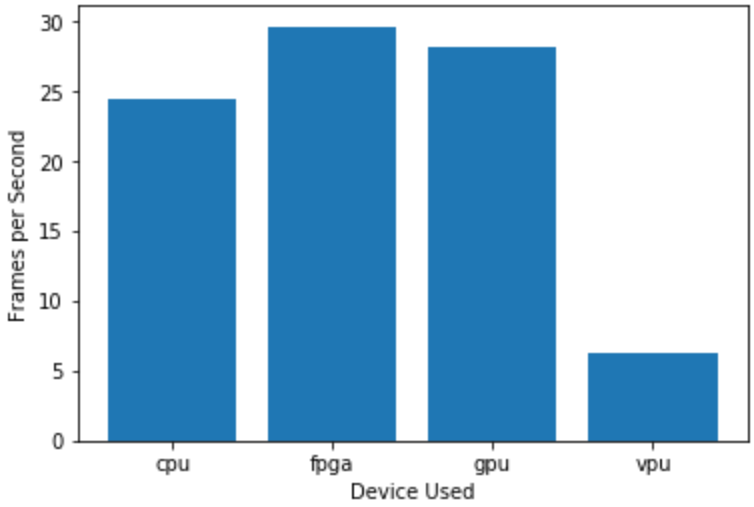
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
Being that the client wants a solution that can be repurpose and reprogrammed to solve different problems, designs are created regularly, and also want a solution with a 5 to 10 years life span, FPGA has those specifications. Also from the test results on Intel DevCloud, FPGA remains the recommended hardware for the client. It has the highest FPS and lowest inference time, which are the basic needs of the client.

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 (Central Processing Unit)

Requirement Observed (Include at least two.)	**How does the chosen hardware meet this requirement?**

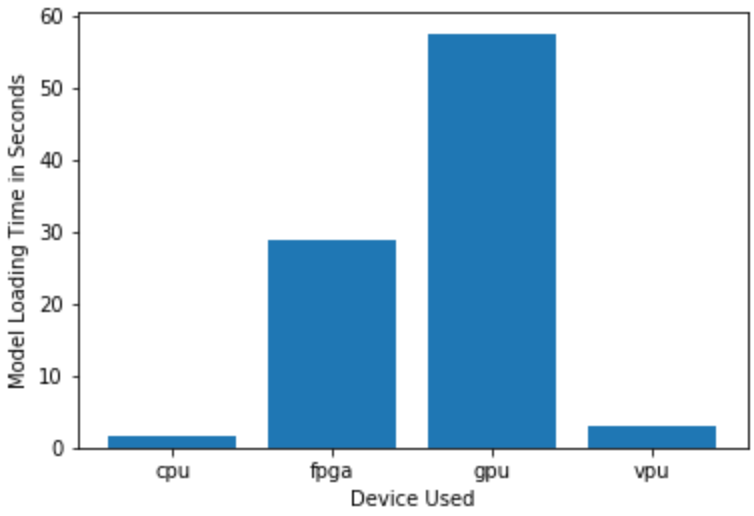
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
The company already has a modern computer at most of the store's checkout counters, each has an Intel i7 core processor and these processors are only used to carry out minimal tasks that are not computationally expensive.	Intel i7 core processors have multiple cores and integrated GPU, enough to run inference on images in addition to minimal tasks it's already performing.
Mr. Lin does not have much money to invest in additional hardware, the company wants to save for other expenses like paying employees and electric bills.	*Due to lack of budget, the exiting Intel i7 core processor is already enough for running the required inference.*
[TODO: Type your answer here]	*[TODO: Type your answer here]*

Queue Monitoring Requirements

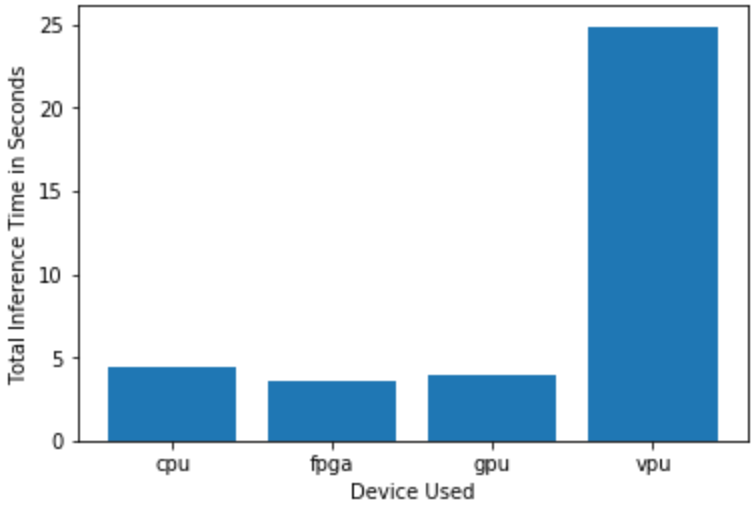
Maximum number of people in the queue	*2*
Model precision chosen (FP32, FP16, or Int8)	*FP32*

Test Results

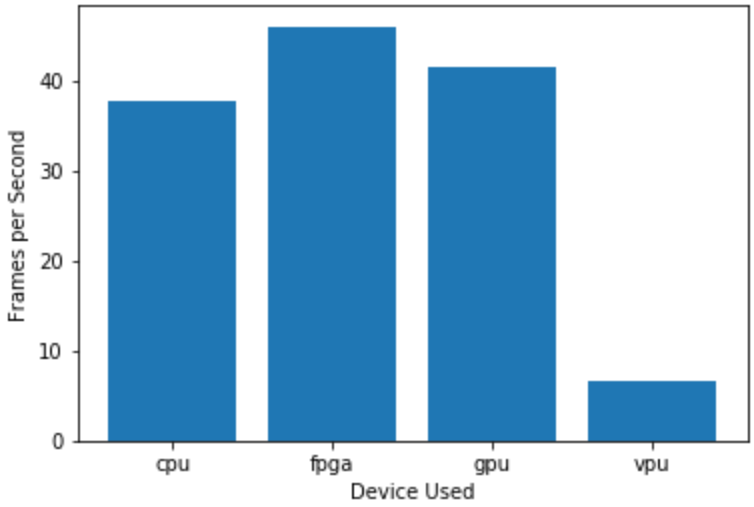
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
The client already has modern computers at most of the store's checkout counters with Intel i7 core processor each which are used for minimal inexpensive computational tasks and don't have money to invest in additional hardware. The Intel i7 core with multiple cores and an integrated GPU is the recommended hardware, given the requirement, budget and already available hardware. And from the test results, the CPU in combination with iGPU have a good and low inference time and high frames per second. Intel i7 core processor can perfectly do the job without any additional hardware for inference.

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 (Vision Processing Units)

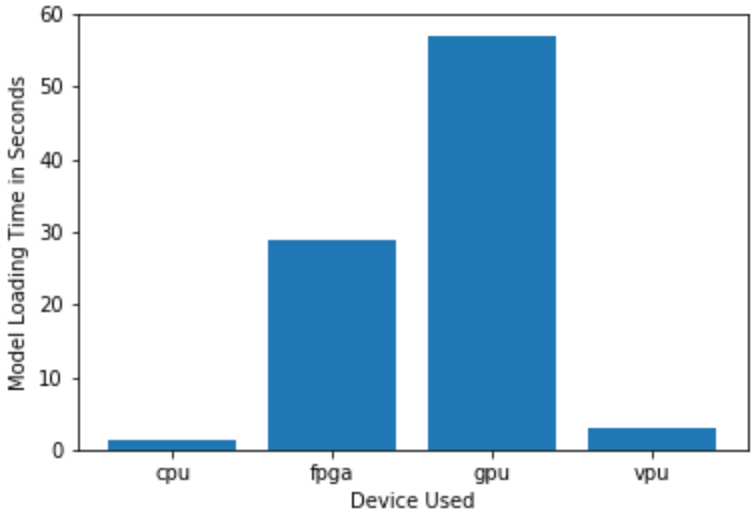
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
There are already all-in-one PCs available for surveillance monitoring that need additional computing power to run inference.	*Additional processing power is needed to add performance to a pre-existing system. VPUs or NCS2 as accelerators that specialized for AI tasks are fit to run the additional inference on the surveillance videos.*
The company's maximum is $300 per machine.	*NCS2 is a good option, costing around $70 to $100 each.*
They want to save for future power requirements.	*NCS2 is a low-power device, can be powered through USB plug and play interface*

Queue Monitoring Requirements

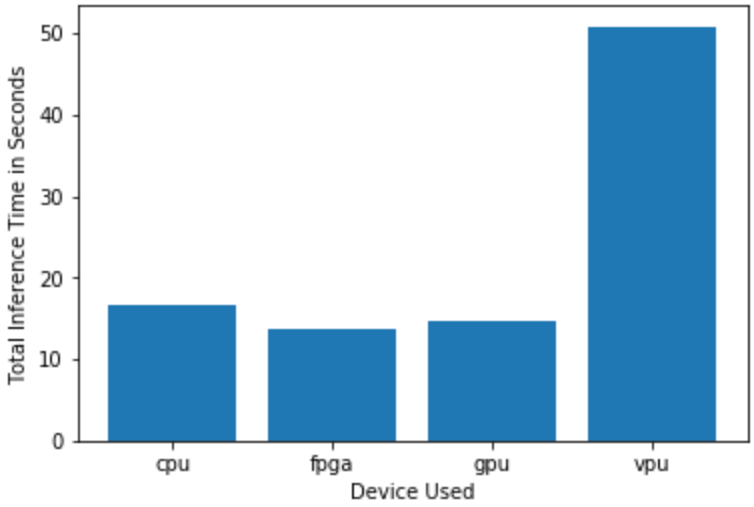
Maximum number of people in the queue	*2*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

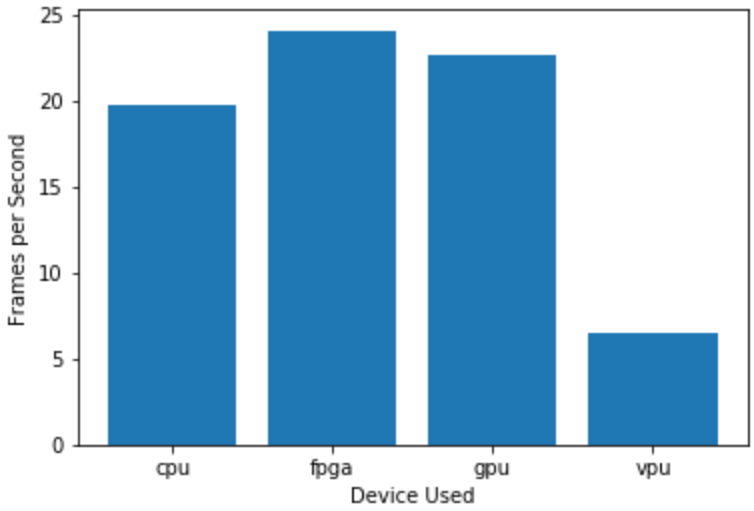
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
Given available hardware that needs additional computing power to run inference, the company maximum budget of $300 and the need to save power for future requirements. VPU, cost around $70 to $100 each and low power consumption, remain the recommended hardware that fits the company's needs. From the test results, even though it performed poorly in inference time and frames per second, it still does a good job of implementing a smart queuing system given the clients requirements. The client here is not seeking speed to increase sales, she mostly wants a smart queue that redirects people in the right manner.