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

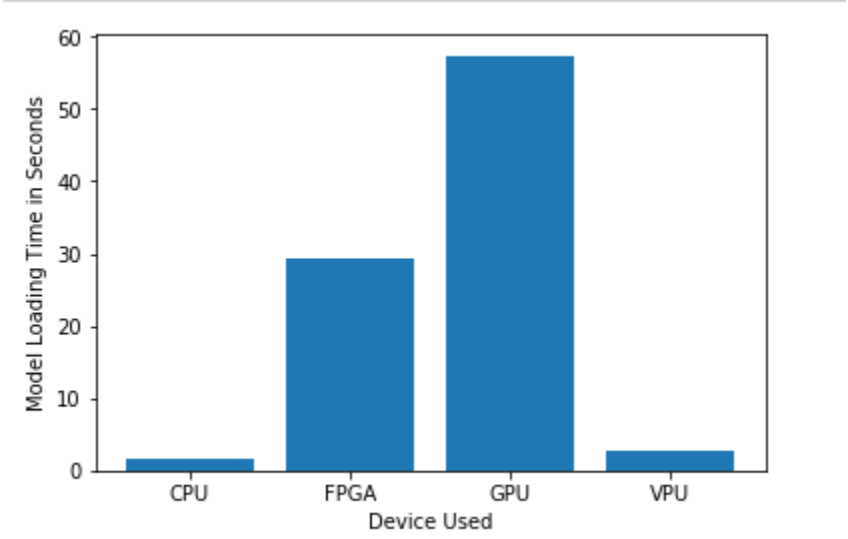
Requirement Observed (Include at least two.)	**How does the chosen hardware meet this requirement?**
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
The client requires a device that can perform inference of about 5 times the input frames.	Having video streams of 30-35FPS, a system that can perform inference of 5 times the infeed frames is required. FPGA can handle inference of around 125 Frames per second which is slightly lower that the customer's requirement.
The client also desires to have a flexible system that can be easily reprogrammed. The system should perform inferences quickly on the video streams and also detect flaws in the chips manufactured.	FPGA are flexible in different ways. They can be reprogrammed to fit the changes and development in an industry. They would fit the customer's flexibility requirement.
Lastly, the client desires to have a quality system that has a life span of about 5-10 years.	FPGAs have a long lifespan of about 10 years making the suitable for the customer's long system lifespan.

Queue Monitoring Requirements

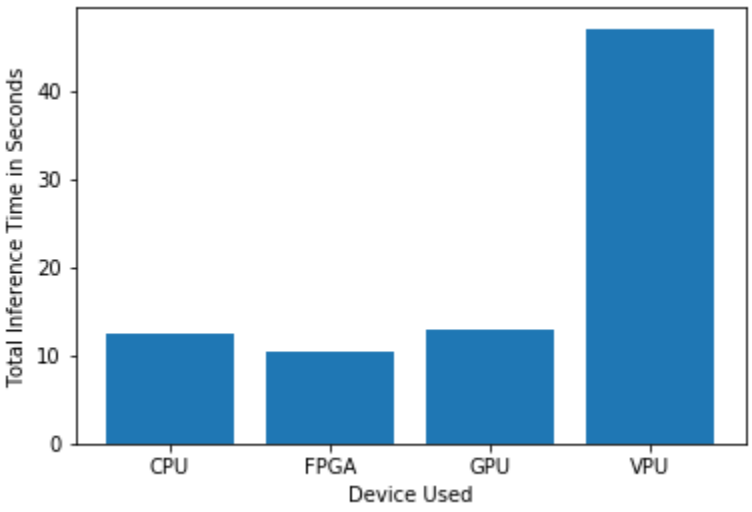
Maximum number of people in the queue	*2*
Model precision chosen (FP32, FP16, or Int8)	FP32

Test Results

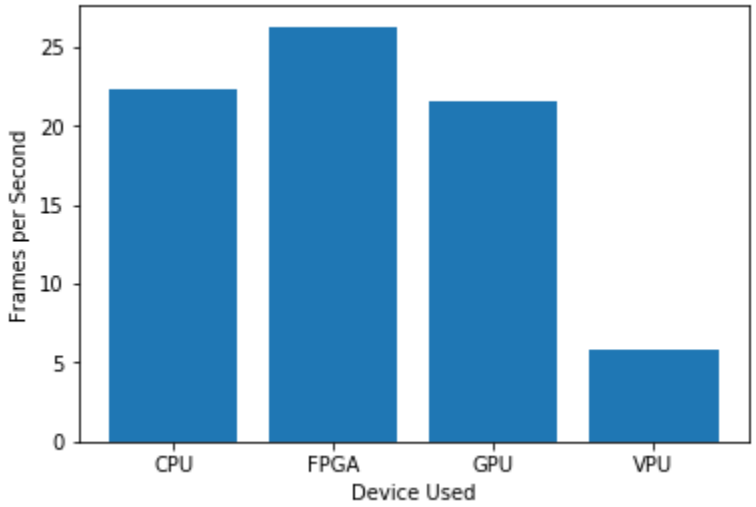
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
In this case scenario, the FPGA qualifies to be the best hardware to fit the customer's needs. This is because to start with, the customer wants a device whose life span ranges from 5-10 years. In addition to that, the customer also wants a hardware device that can perform inference 5 times faster than the infeed stream. From the graphical results, the FPGA out performs all the other hardware devices in terms of number of frames inferenced in one second and has the lowest inference time matching the customer's needs. The Field Programmable Gate Array is the only device which is flexible allowing the client to modify the program to enable him to detect flaws in different types of chips. The only downside of using the FPGA in this scenario is the large model load time compared to the CPU and VPU.

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

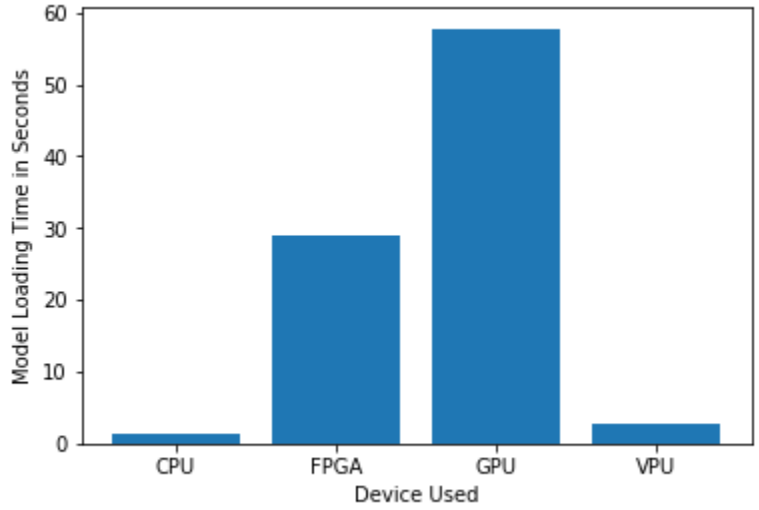
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
Low budget	Mr. Lin has core i7 processors at each counter that can be used as the edge device hence not spending on purchasing an edge device
Low power consumption	Power consumption won't be increased that much because he already has modern computers with core i7 processors already at his shop
No additional hardware	There are CPU at the checkout counters that will be used as the edge devices for this scenario

Queue Monitoring Requirements

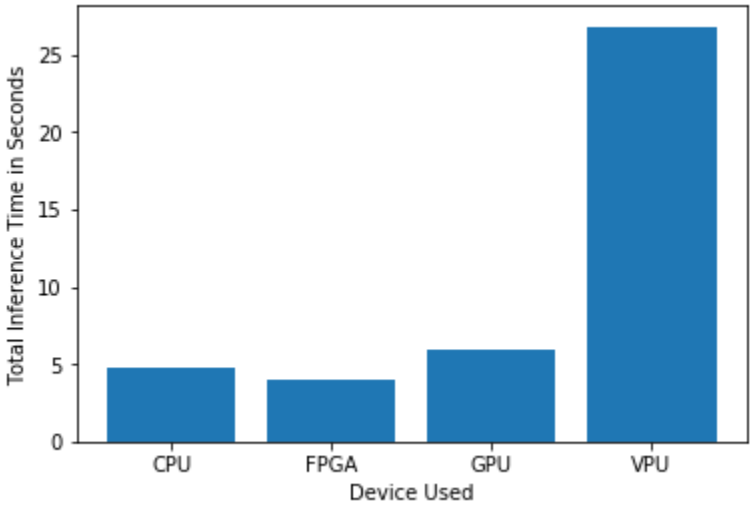
Maximum number of people in the queue	2
Model precision chosen (FP32, FP16, or Int8)	FP32

Test Results

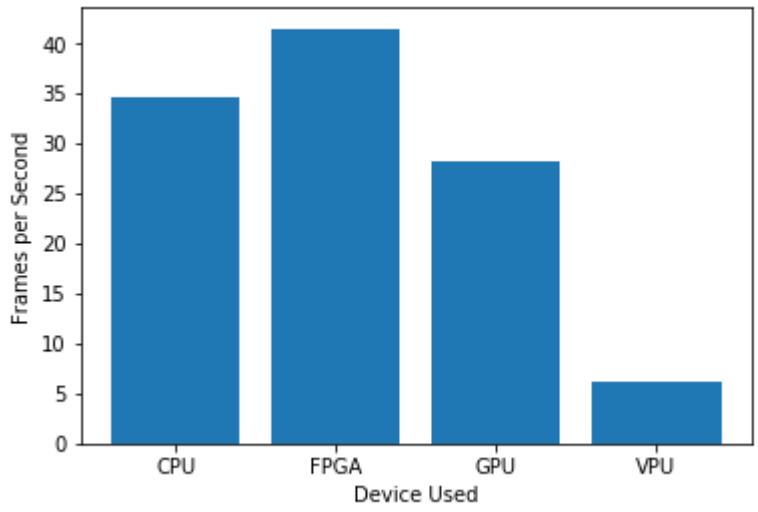
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
CPU would be the best hardware to run inference for the smart queue system in this scenario. This is as a result of the small model load time as shown in the graphical results obtained from running inference on different devices. The customer does not desire to have addition hardware to perform inference and would like to utilize the computational power CPU's at the counter to perform inference of the video footage. Power consumption would not be very much affected. This is because the computers were already there before installation of the smart queuing system. The only difference is that much of their computational power will be utilized to perform inference on the video streams.

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)
IGPU

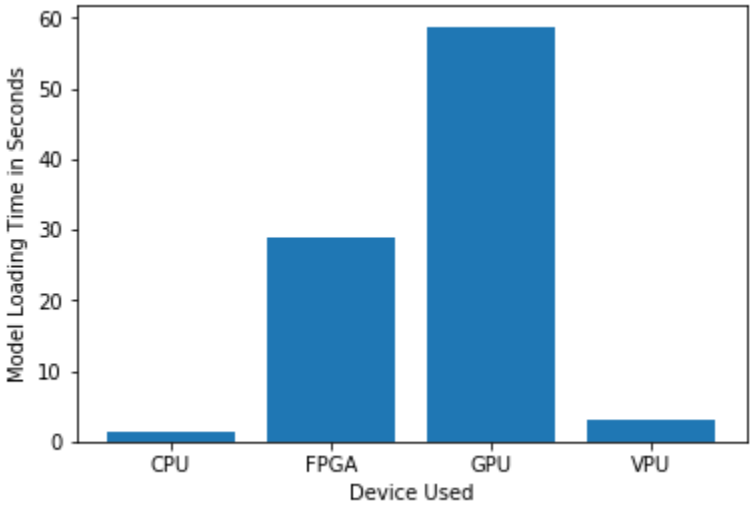
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
Ms. Leah has a budget of $300 on each device	An Intel Xeon Processor with an Integrated GPU has a customer recommended price of $362 which is slightly above the customer's budget.
She wants to save on hardware	Having an IGPU will ensure that inference of all the video frames is done simultaneously using one hardware.
Save on future power consumptions	Power consumption will remain the same compared with the time of installation of the Edge system and also in future.

Queue Monitoring Requirements

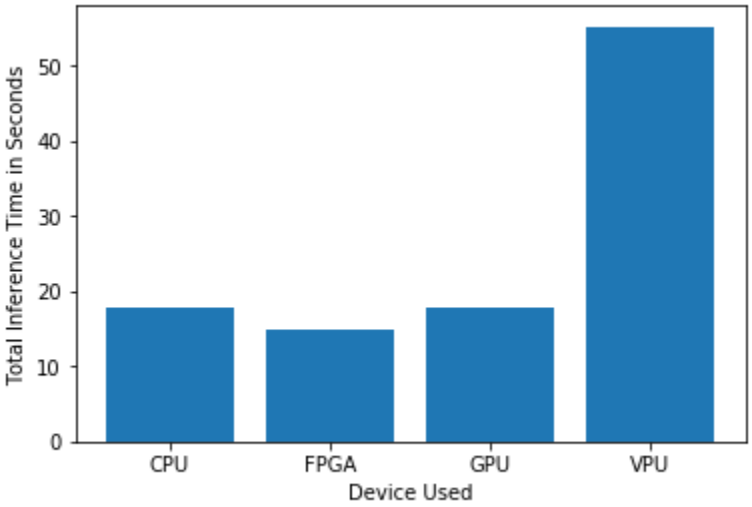
Maximum number of people in the queue	4
Model precision chosen (FP32, FP16, or Int8)	FP32

Test Results

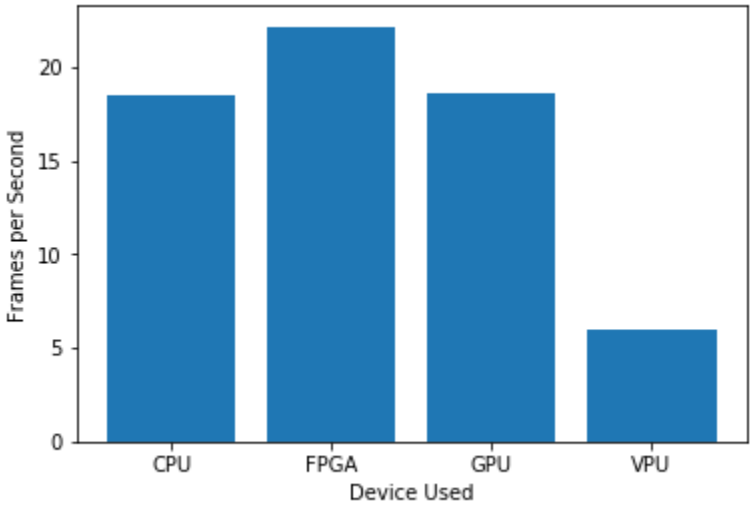
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
In this case scenario, the customer has a budget of $300, she also desires to save on hardware and future power requirements. She will need to perform inference from 7 CCTV cameras. From previous exercises, it is advisable to perform multiple inferences on an Integrated GPU rather than on the other hardware devices because the inferences will be performed faster. Having an integrated GPU will ensure that the customer will save on hardware because all inferences will be done from one hardware device and will also be faster.