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

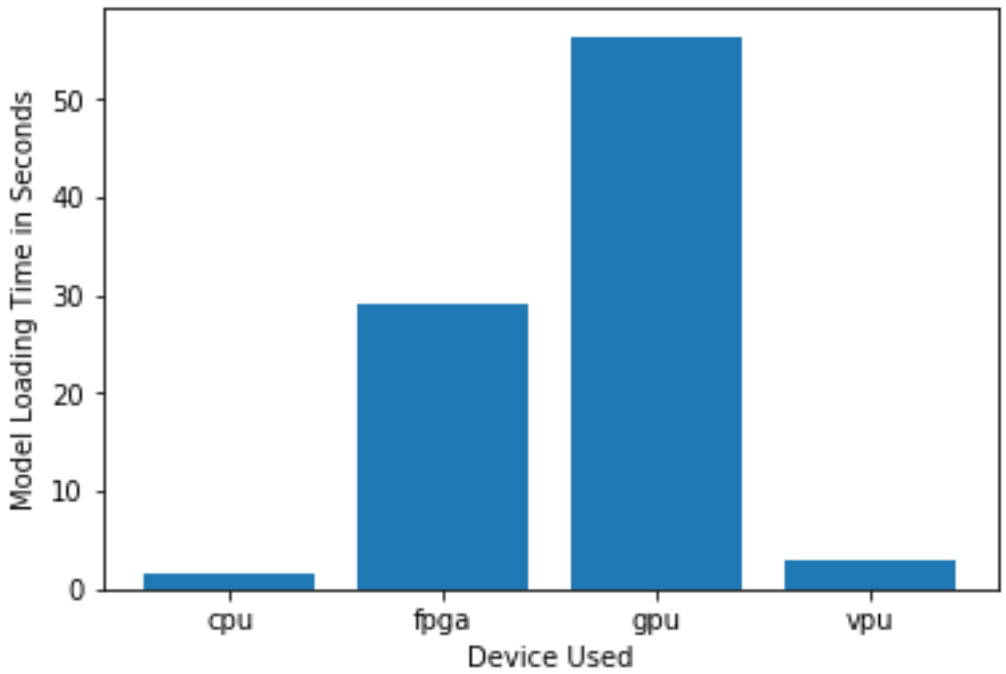
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Revenue is good so they have money to spare on the devices	*Good revenue so client can make significant Investment.*
To detect chip flaws without slowing down the packaging process, the system would need to be able to run inference on the video stream very quickly so the system would also need to be flexible so that it can be reprogrammed and optimized to quickly detect flaws in different chip designs. The client would like the image processing task to be completed five times per second.	*FGPAs are flexible. They are field-programmable; they can be reprogrammed to adapt to new, evolving, and custom networks and are faster.*
Significant investment and they would ideally like it to last for at least 5-10 years	*For long term FPGA are advisable for 5-10 years with significant Investment.*
To help understand and address these issues, the client wants a system to monitor the number of people in the factory line. The factory has a vision camera installed at every belt. Each camera records video at 30-35 FPS (Frames Per Second) and this video stream can be used to monitor the number of people in the factory line. The client would like the image processing task to be completed five times per second.	*Need inference fast which FPGA can provide*

Queue Monitoring Requirements

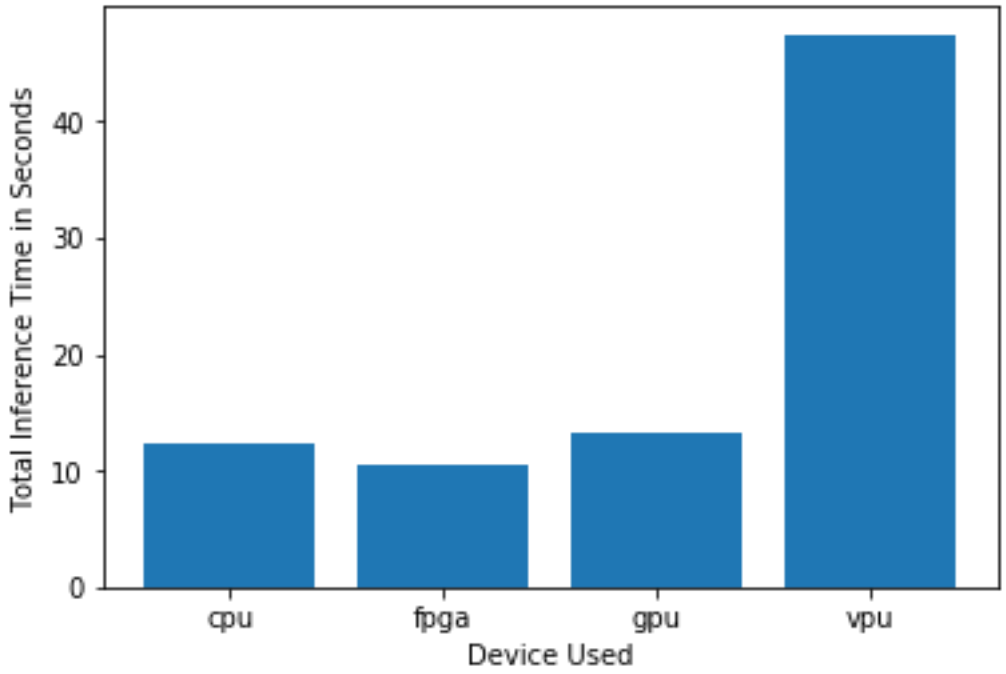
Maximum number of people in the queue	*2 to Many*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

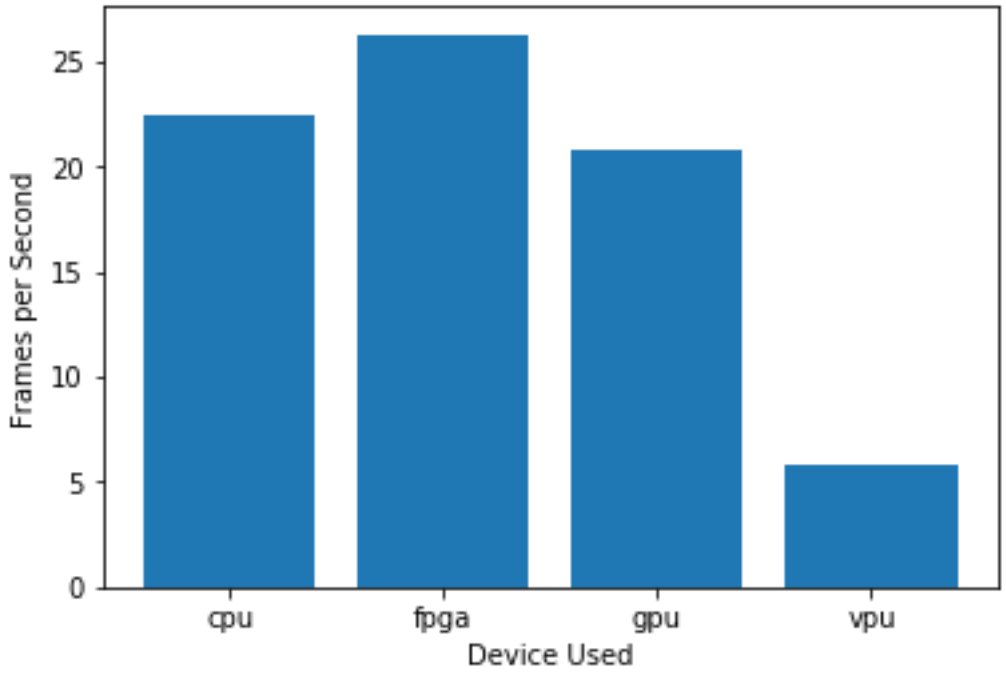
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
My final recommendation remains an FPGA: ** They have a long life-span* ** meets speed requirements (mostly thanks to the CPU)* ** reprogrammable* *According to our test results, FPGA can allow the image processing task to be completed well above the five times per second requirement, it can be reprogrammed and optimized to quickly detect flaws in different chip designs, and it has a very long lifespan of at least 5 to 10 years, exactly as our client has specified.*

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?
Doesn't need to be fast - Average wait time of 230-400 secs (2-5 per queue), <20 seconds	*Doesn't need to run several fps for improvement*

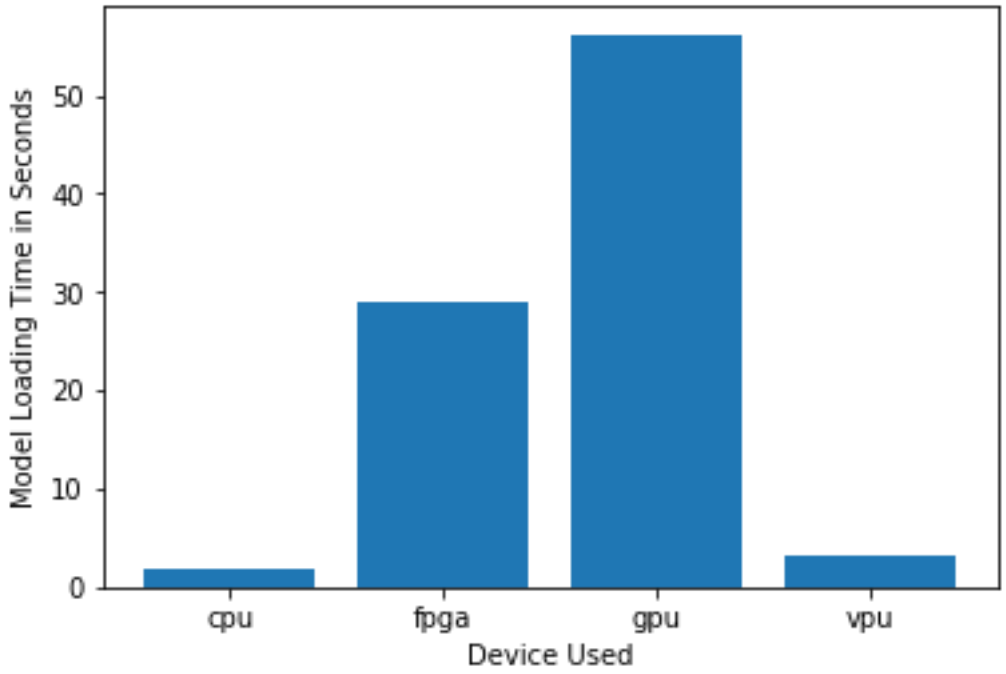
No money to spare much, minimal requirements	*As the client has specifically indicated that they do not have much money to invest in additional hardware, and that they already have one CPU installed at each checkout counter, we could use these CPUs for developing our smart queuing system to direct people to less-congested queues in the store.*
Less power consumption (save on electric bill)	*Already powered up for cash registers*
Cheap – avoid additional hardware, has i7 CPUs already	*The client already has some modern computers at the store's checkout counters. These processors are currently only used to perform some minimal tasks that are not computationally expensive, such that we can use these existing CPUs to meet the hardware requirement.*

Queue Monitoring Requirements

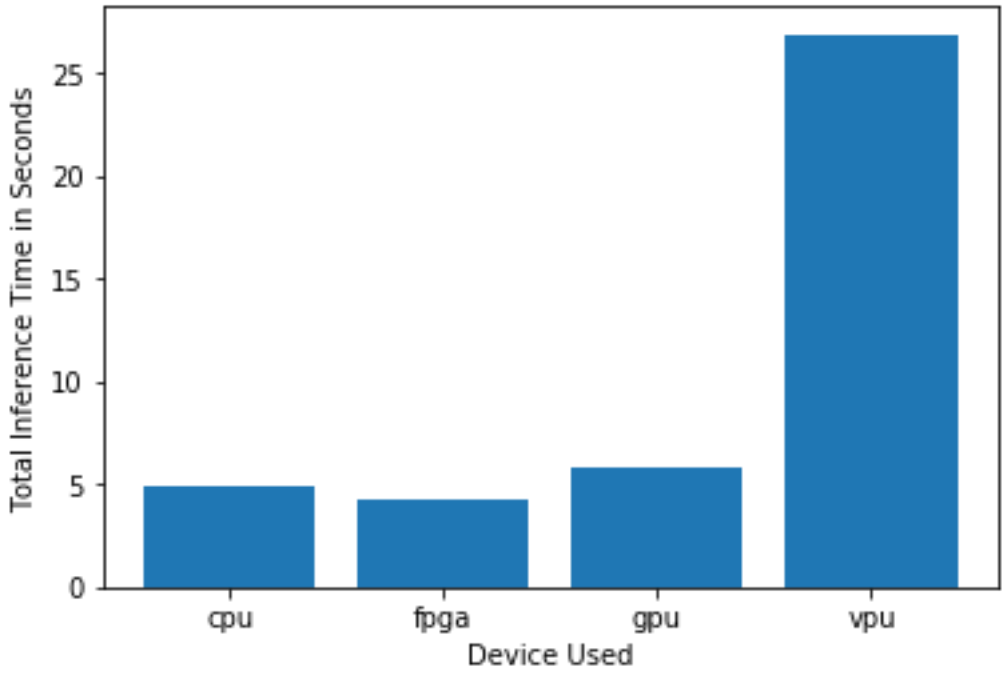
Maximum number of people in the queue	5 Average of 2 per queue (during normal daily hours) to 5 per queue (during rush hours)
Model precision chosen (FP32, FP16, or Int8)	*FP32*

Test Results

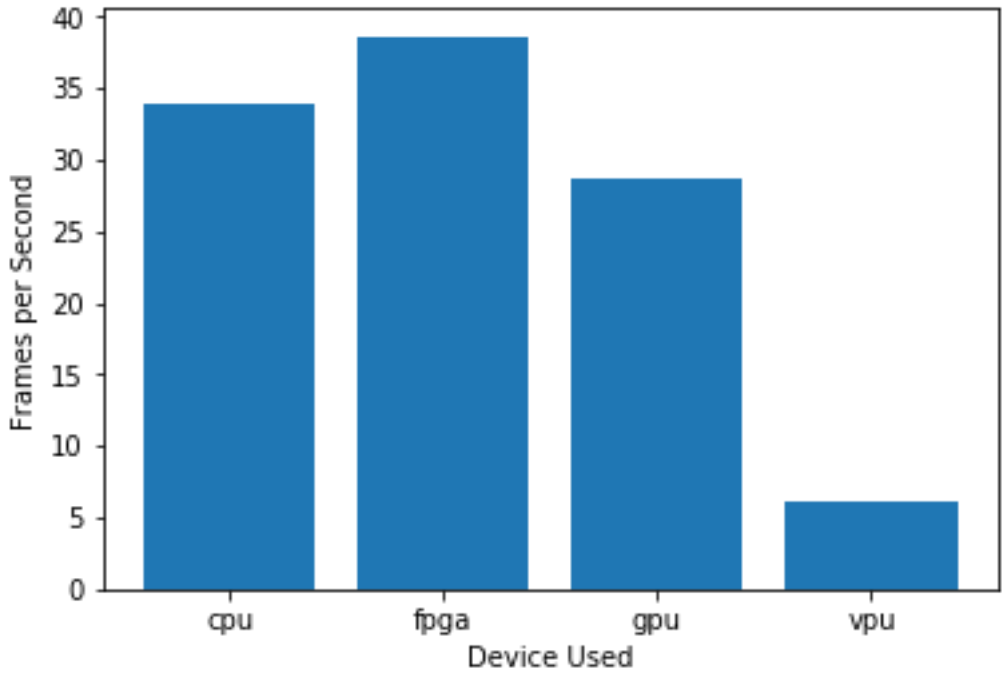
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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 is the final hardware recommendation for the retail scenario ** Doesn't need to buy new hardware* ** Doesn't need to be crazy fast, CPU is plenty fast in my tests* ** Low additional power - power is already being consumed by CPUs* *It is more than sufficient for the smart-queuing task based on the model loading time, total inference time, and number of frames per second results from our performance assessment with Dev Cloud. Moreover, the most important specification from the client is using the existing Intel i7 core processor in each modern computer at each checkout counter. Hence the CPU is the natural choice.*

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 or NCS2

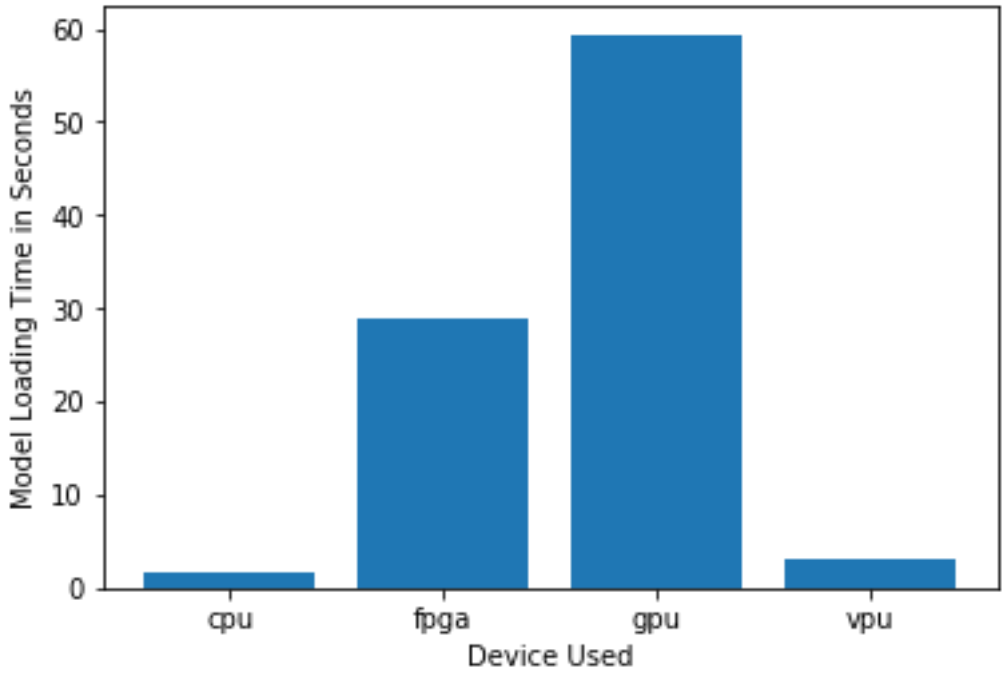
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Save on hardware and future power requirements	*VPUs are low power devices*
The client's budget allows for a maximum of $300 per machine, and she would like to save as much as possible both on hardware and future power requirements.	*A VPU or NCS2 costs less than $100 each, and the NCS2 is extremely low power.*
CPU has no additional processing power	*Pairs well with current setting as we can't use CPU, so VPU will act as AI accelerators.*
They monitor the entire situation with 7 CCTV cameras on the platform. These are connected to closed All-In-One PCs that are located in a nearby security booth. 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.	*VPU or NCS2 might be suitable as each of these has a convenient USB3.1 plug and play interface. Note that the NCS2 can also be used on systems with only a USB2 port, but the inference will run slower due to I/O throttling.*

Queue Monitoring Requirements

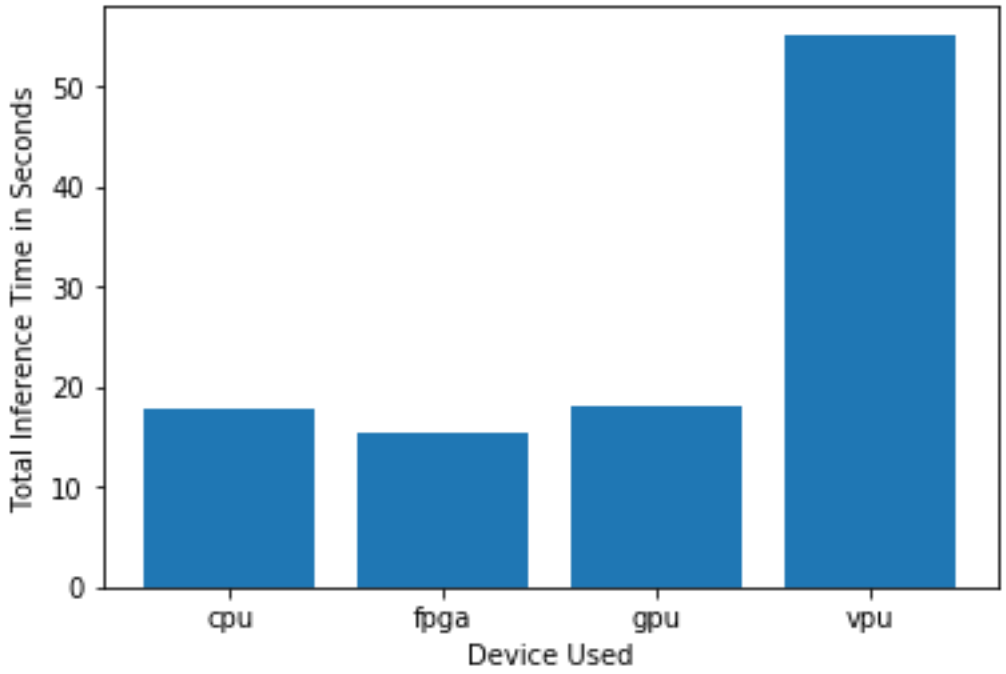
Maximum number of people in the queue	*Average 7-15 people in single queue*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

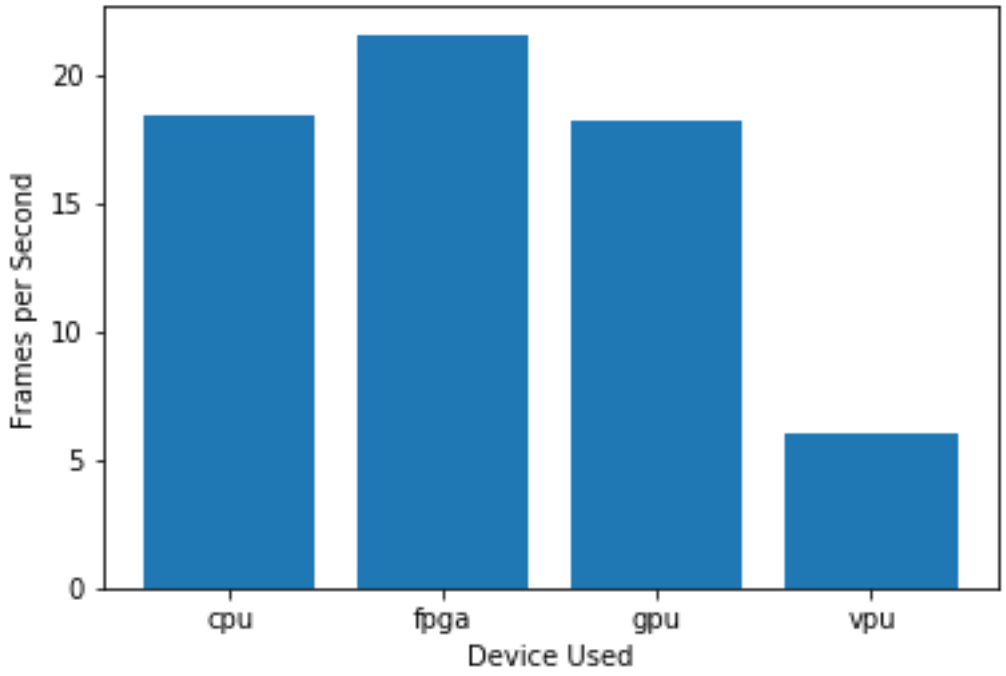
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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 or NCS2 is the final recommendation for the transportation scenario. This is because despite its relatively poor performance in the inference time and the FPS, it does the job for implementing a smart queuing system adequately while satisfying the cost requirement of the client. ** it is low power* ** inexpensive (less than $300)* ** Doesn't use current CPU* ** fast enough for use case*