Choose the Right Hardware

Completed 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?
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.
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.	Once programmed with a suitable bitstream, FPGAs can execute neural networks with high performance and very little latency. The high performance comes from the ability to run many sections of the FPGA in parallel. FPGAs also flow the data from one layer to the next, while keeping the data from one output to the next input layer on the same FPGA device. When running a neural network, we run the whole thing on the FPGA so the FPGAs don't go off-chip for the memory. This is faster than sending the output back to the CPU over the PCIe bus.
Once the productivity problem has been addressed, the client would like to be able to repurpose the system to address a second issue. 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.	FGPAs are flexible. They are field-programmable; they can be reprogrammed to adapt to new, evolving, and custom networks.

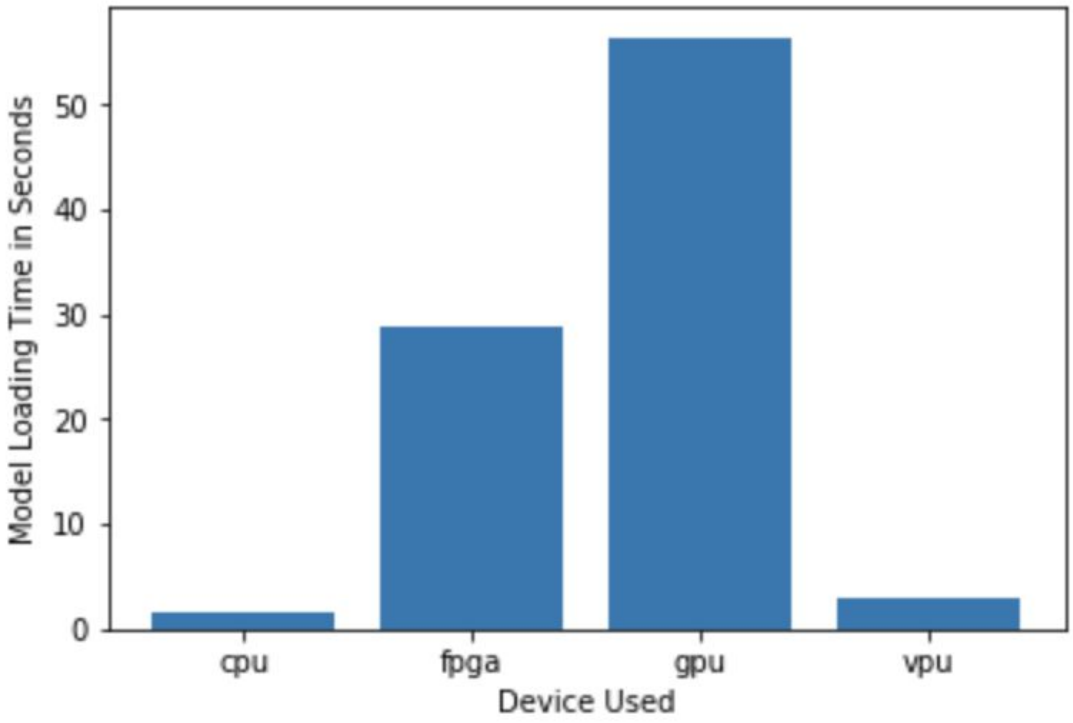
While the company has plenty of revenue to install a quality system, this is still a significant investment and they would ideally like it to last for at least 5-10 years.	FPGAs have a long lifespan. For example, FPGAs that use devices from Intel's Internet of Things Group have a guaranteed availability of 10 years, from start of production.

Queue Monitoring Requirements

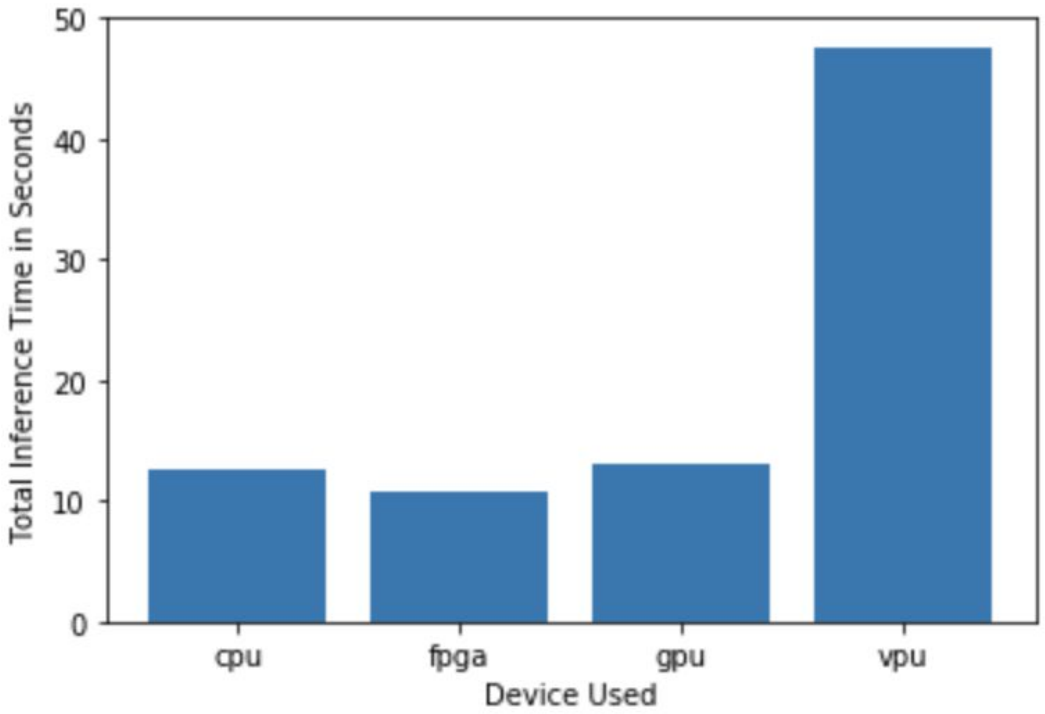
Maximum number of people in the queue	2
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

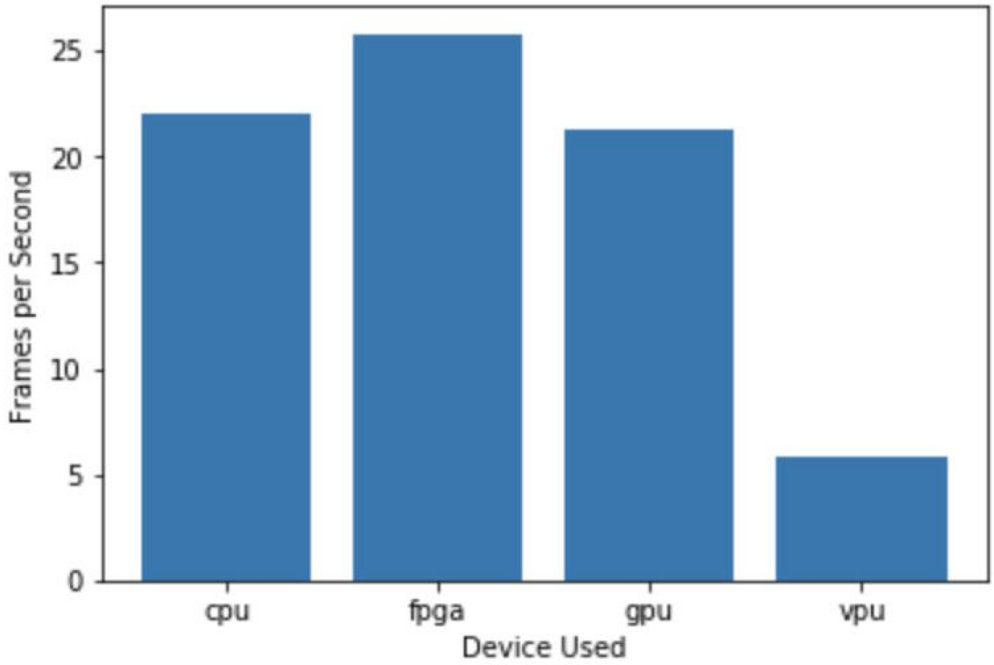
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
FPGA is the final recommendation for the manufacturing scenario. This is because this solution meets all requirements. 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

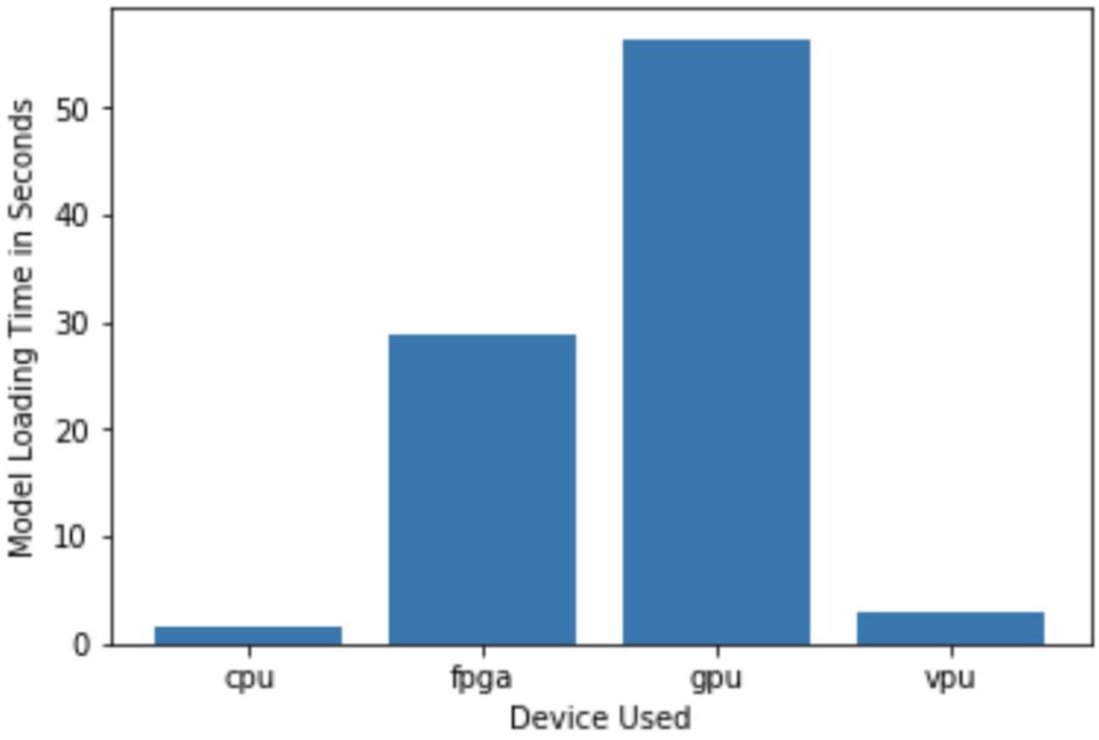
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
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.	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.
The client does not have much money to invest in additional hardware, and also would like to save as much as possible on his electric bill.	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.

Queue Monitoring Requirements

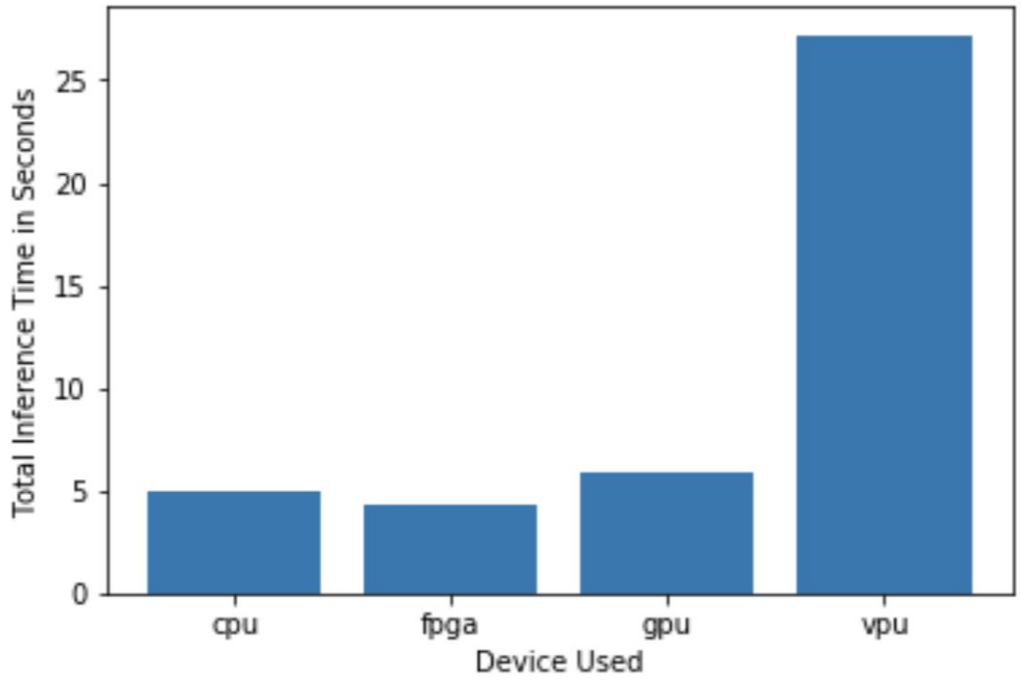
Maximum number of people in the queue	5. The total number of people in the checkout queue ranges from an average of 2 per queue (during normal daily hours) to 5 per queue (during rush hours).
Model precision chosen (FP32, FP16, or Int8)	FP32

Test Results

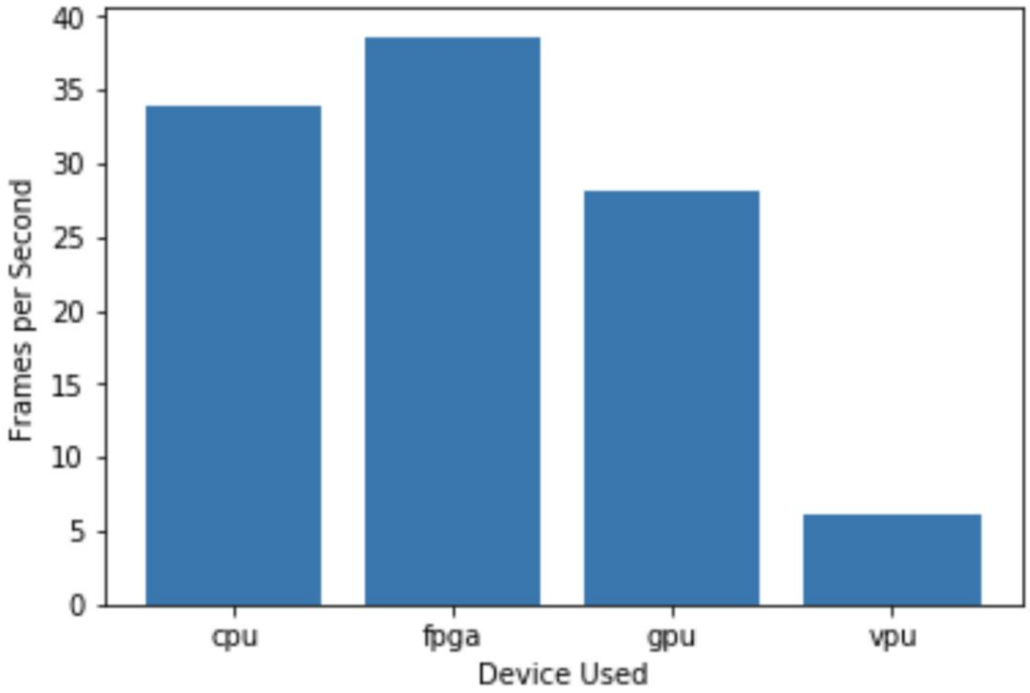
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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. This is because 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 DevCloud. 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

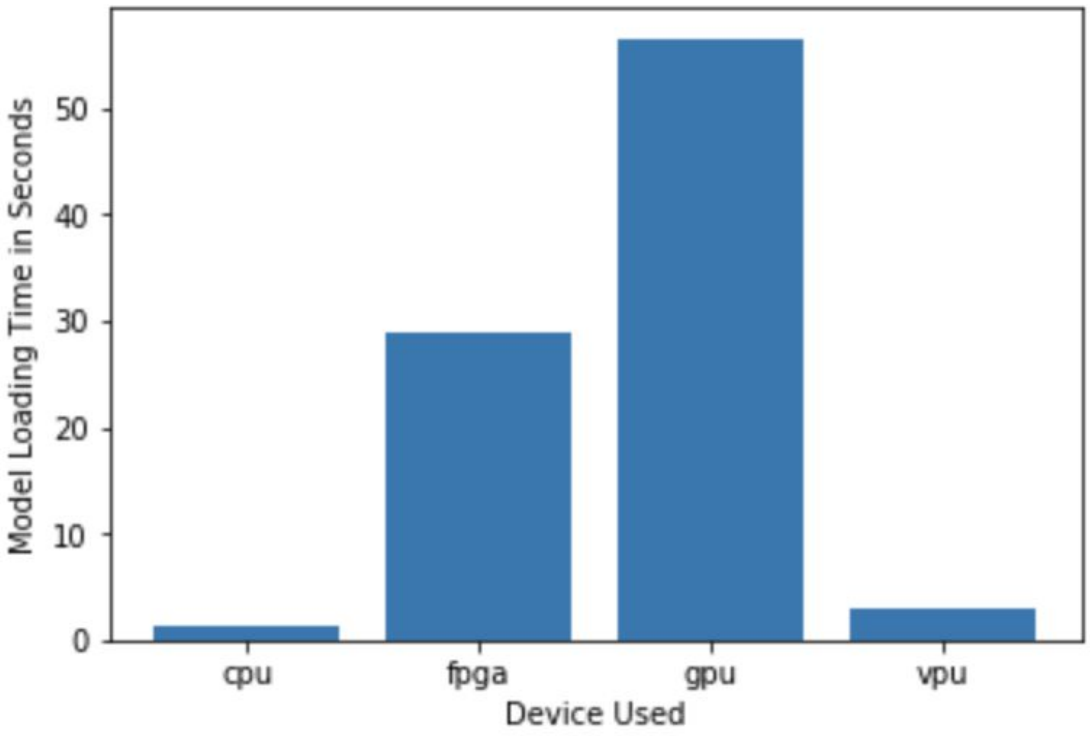
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
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.
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.

Queue Monitoring Requirements

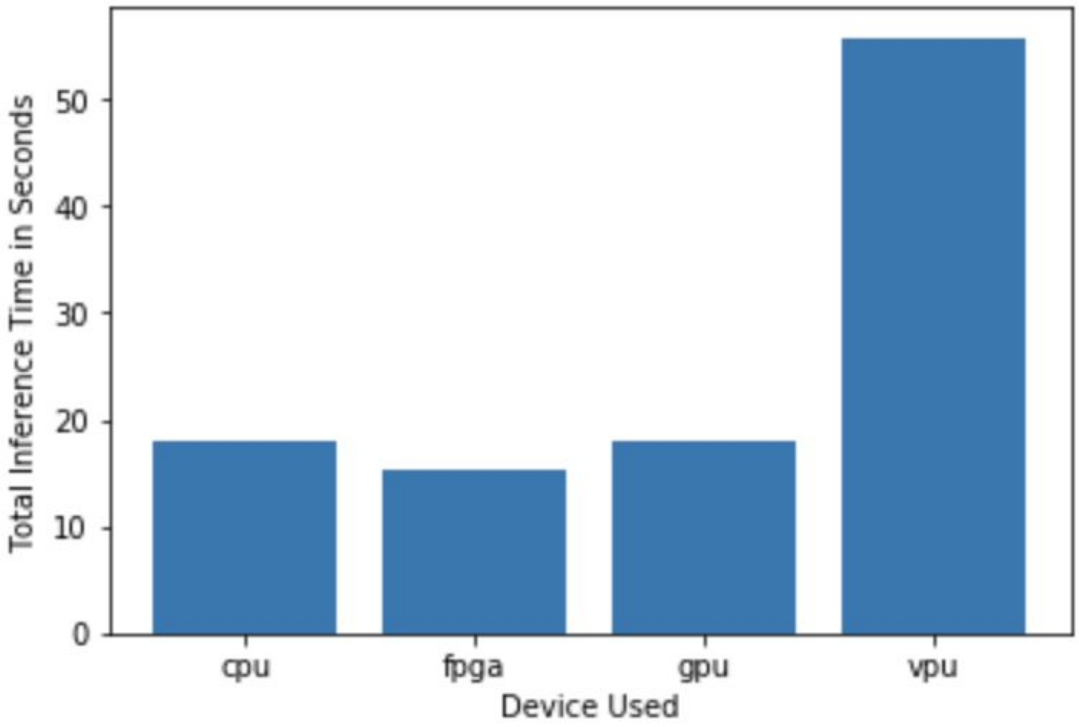
Maximum number of people in the queue	Since there is one queue on each side of the door, the maximum number of people in each queue is about 3. In peak hours they currently have over 15 people on average in a single queue outside every door in the Metro Rail. But during non-peak hours, the number of people reduces to 7 people in a single queue.
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

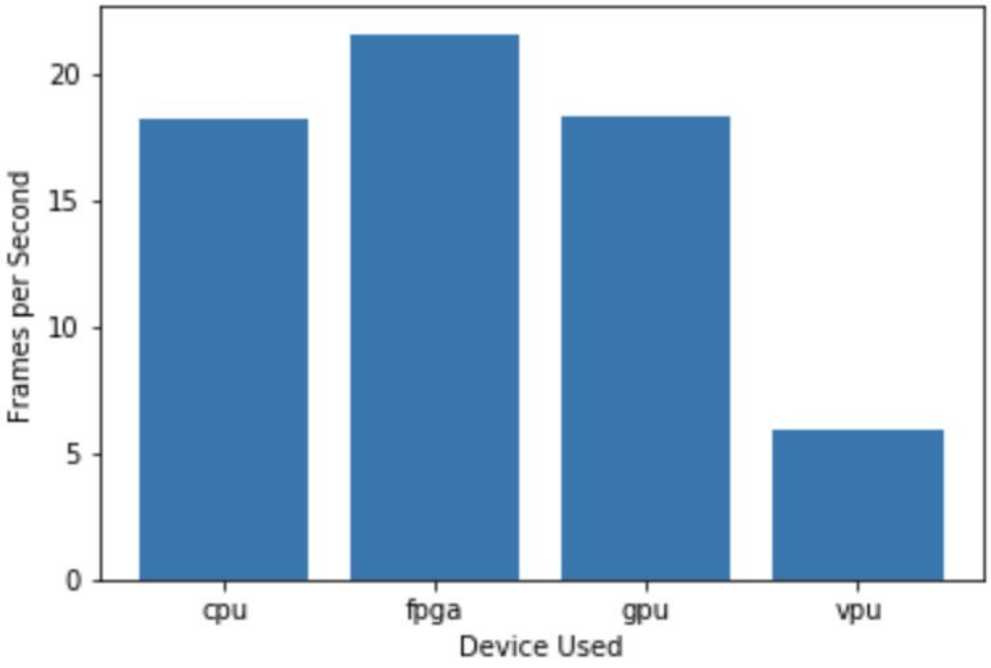
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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.