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

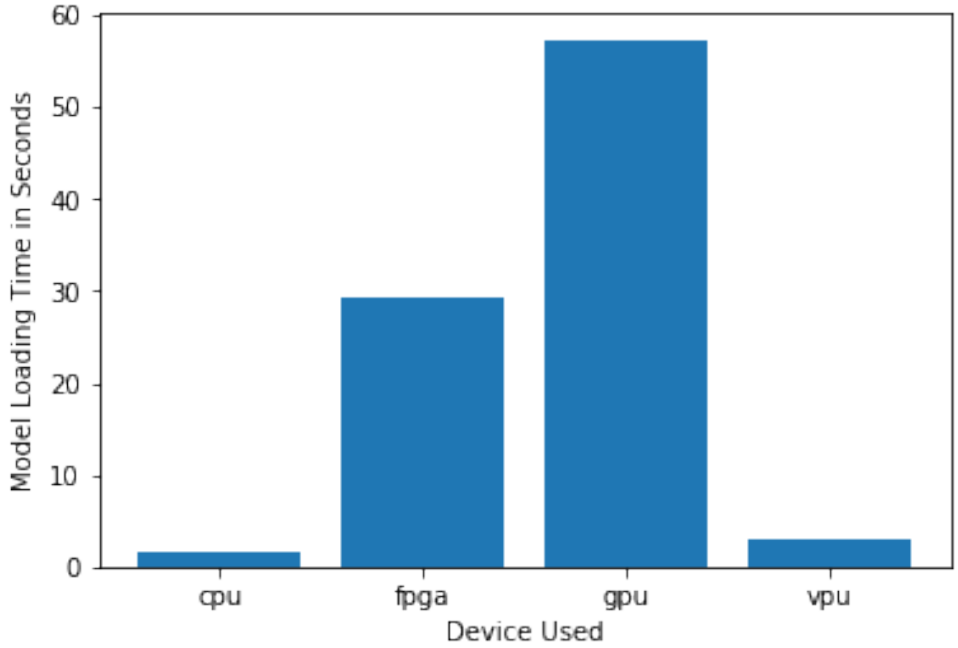
Requirement Observed (Include at least two.)	**How does the chosen hardware meet this requirement?**
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
Performance: the system would need to be able to run inference on the video stream very quickly	*we can program an FPGA to act as an AI accelerator so that it performs well when running inference.*
Need for flexibility: new designs are created regularly	*FPGAs are chips designed with maximum flexibility, they can be optimized and reprogrammed as needed in the field (i.e., after manufacturing and deployment). FPGAs can be reprogrammed to adapt to new, evolving, and custom networks.*
The client want to install a quality system that can last for atleast 5-10 years	*FPGAs are designed to have 100% on-time performance and they have a long lifespan, Guaranteed long lifespan of ~10 years*

Queue Monitoring Requirements

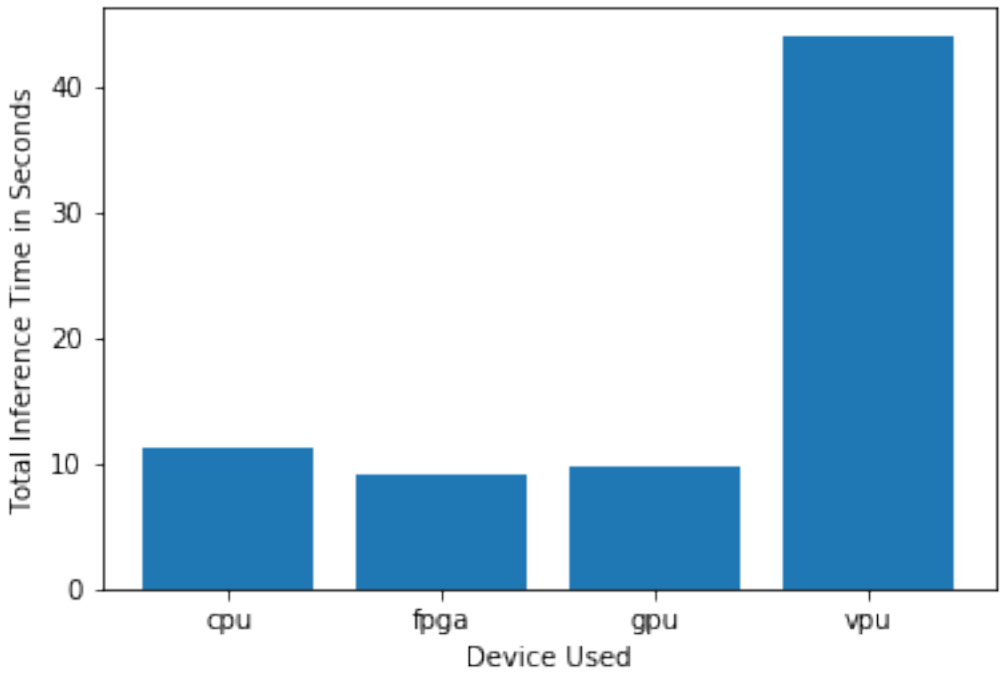
Maximum number of people in the queue	*5*
Model precision chosen (FP32, FP16, or Int8)	*FPGA+CPU: FP16*

Test Results

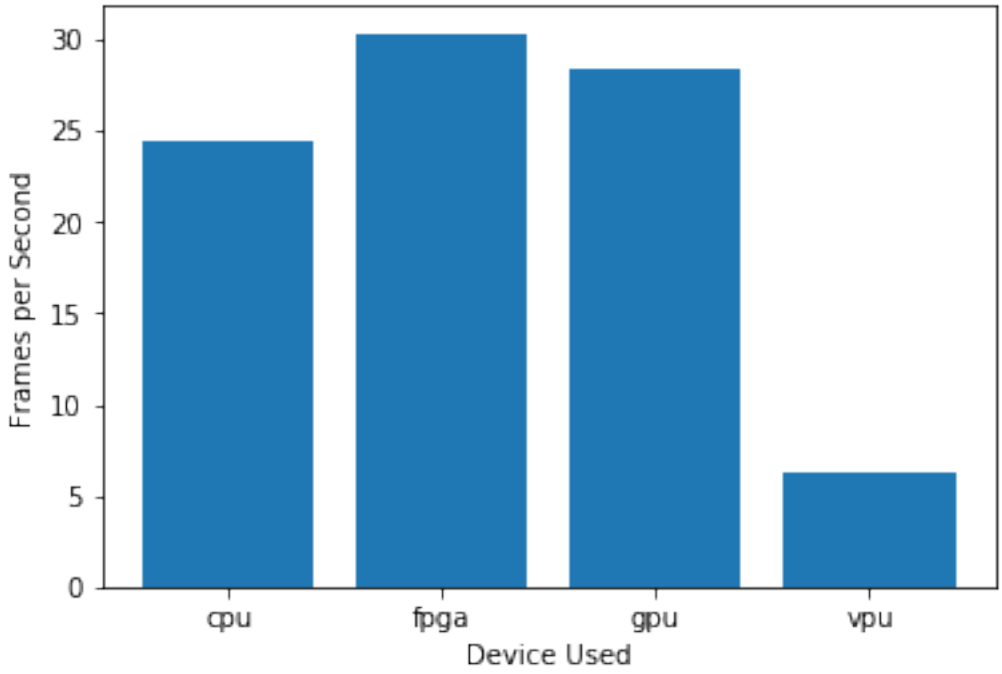
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
By considering the client requirements: performance, flexibility and long lifespan, we can see that the FPGA would be the best hardware for this project, if we consider flexibility and lifespan of this hardware, FPGA is the only hardware here that can be reprogrammed as needed and have a long durability. On the other hand FPGA cost alot more than other hardwares, this is still Ok as The client has plenty of revenue that can be used for this project and install a quality system. *comparing the result from the above figures, the model loading time of FPGA is higher than two other hardwares, we should not worry about this, as is just one time event, FPGA has low inference time and process frames faster than any other hardware.* *the FPGA would be the best hardware for 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/IGPU

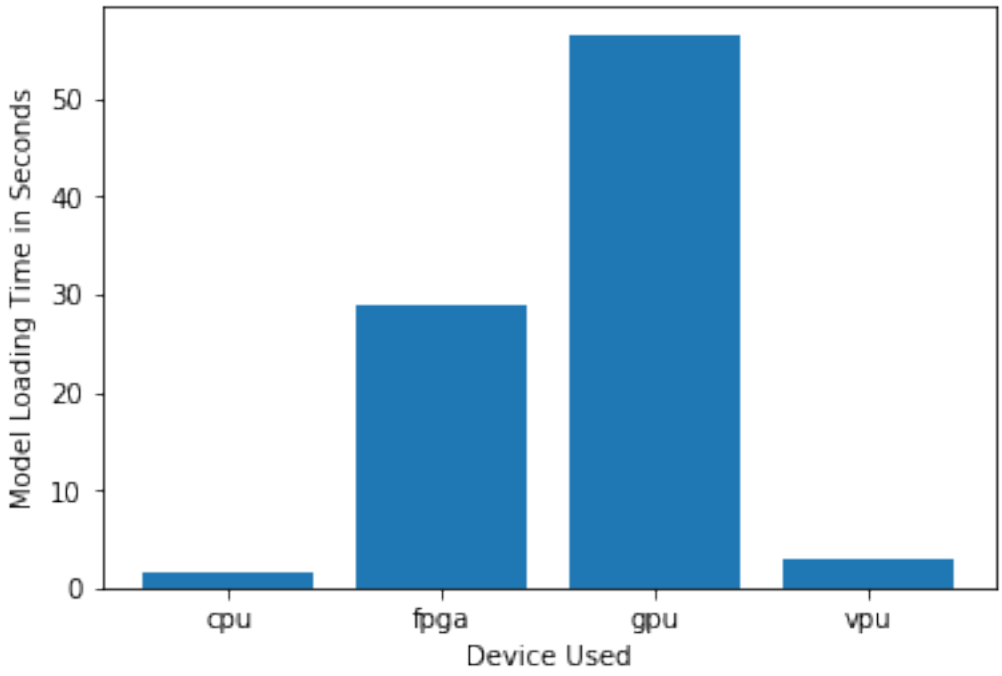
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
The client does not have much money to invest in additional hardware.	*Since there are CPUs Currently used only to carry out some minimal tasks that are not computationally expensive, these CPUs can be used to run inference as they are not fully utilized for other tasks. using the CPU or IGPU will not require additional hardware.*
The client like to save as much as possible on his electric bill.	*By Using the CPUs available on the client computers no need to expand the power of system. But using the their IGPUs will require more power than using only CPUs*

Queue Monitoring Requirements

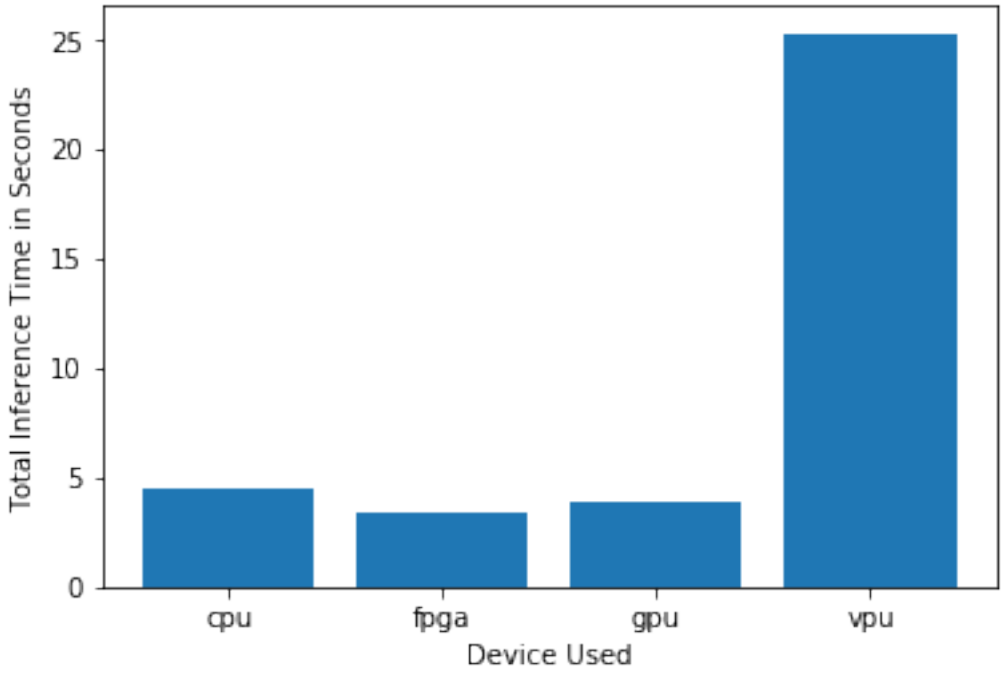
Maximum number of people in the queue	*2*
Model precision chosen (FP32, FP16, or Int8)	*CPU: FP32* *CPU+IGPU: FP16*

Test Results

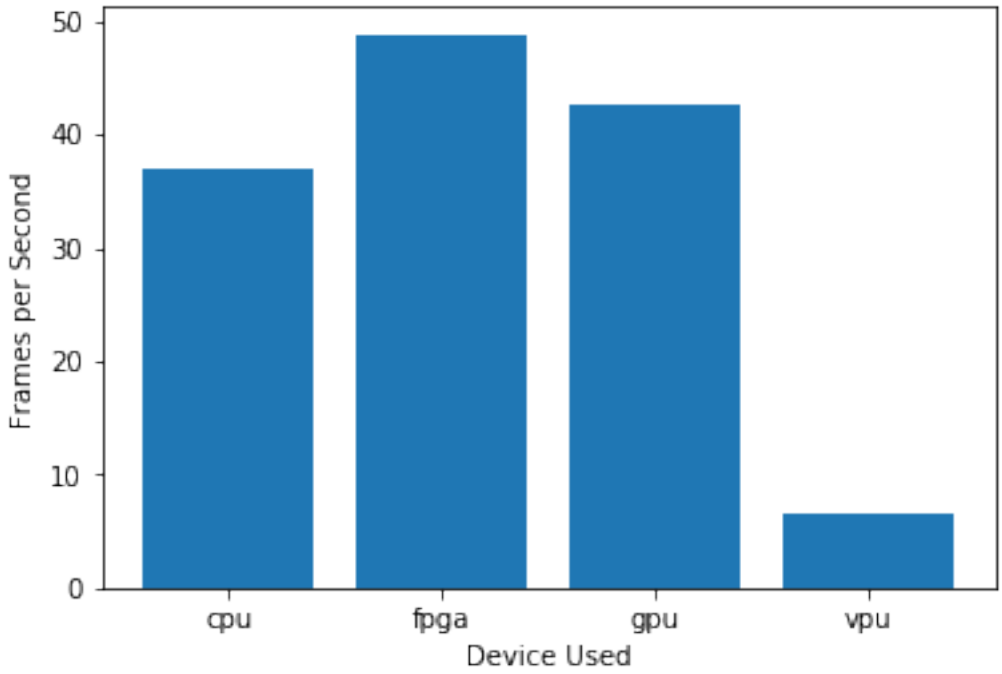
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
From the result in figures above, we can see that the CPU has lowest model loading time. for inference time and FPS process, since the client does not want to invest for additional hardware and he want to save his electic bill, the client should consider using these CPUs only as they would be the best option at the moment. later if the client is ready and have enough power he can use the IGPUs available on these processors to increase the reliability of his system.

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

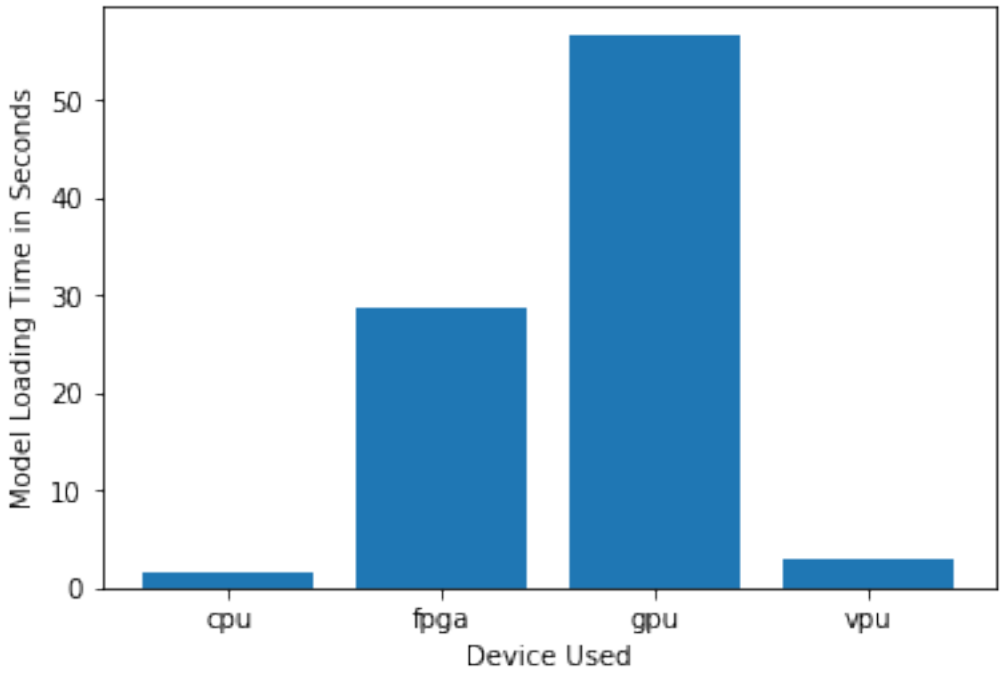
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
The clients requires to use the CPUs for other task and no significant additional processing power is available to run inference.	*AI Accelerators now provide a relatively inexpensive way to boost your performance. Intel NCS2 (with Mariad-X VPU) removable USB device can be used for AI inferencing.*
Cost constraints: The client budget $300 per machine	*the NCS2 is an inexpensive option, typically costing around $70 to $100.*
The client would like to save as much as possible both on hardware and future power requirements.	*The Intel® Neural Compute Stick 2 (NCS2) is a USB3.1 plug and play removable VPU for AI inferencing. This allows ease in the use of the hardware. the NCS2 is extremely low power device, The VPU, Myriad X processor present in NCS2 has a very low power consumption of only 1-2 watts.*

Queue Monitoring Requirements

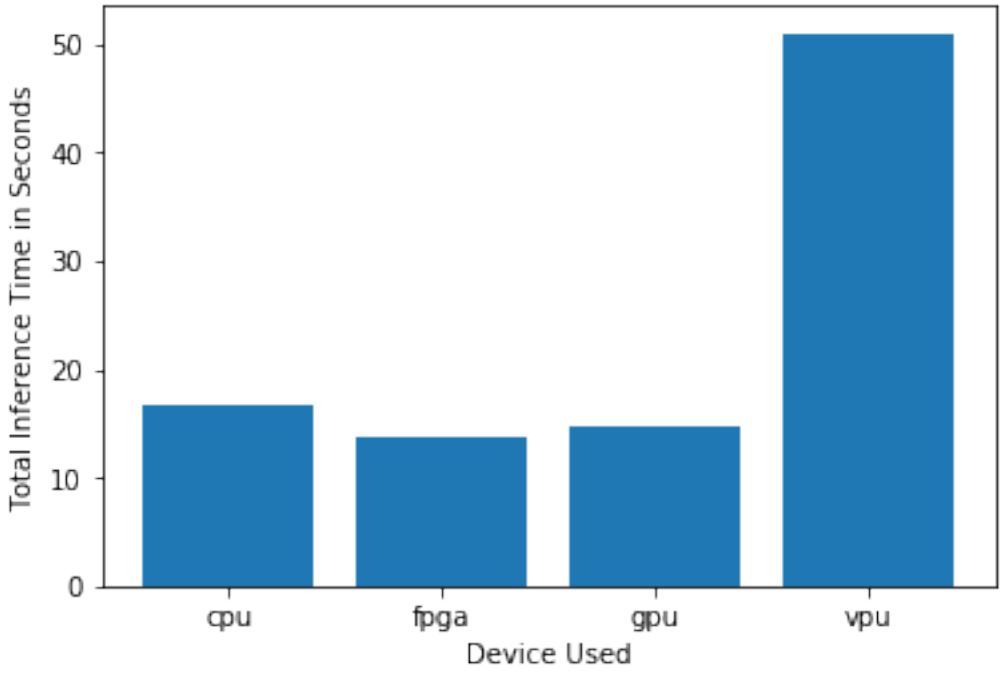
Maximum number of people in the queue	*5*
Model precision chosen (FP32, FP16, or Int8)	*VPU: FP16* *The Intel® Neural Compute Stick can only run FP16 models at this time.*

Test Results

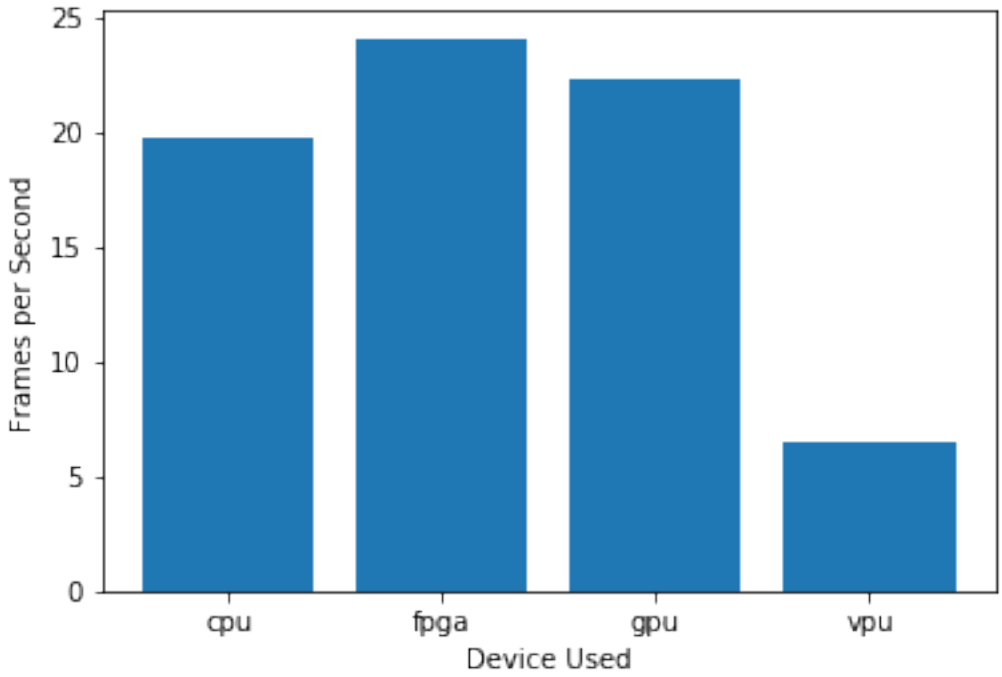
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
From the stats shown in the above figures we can see that VPU inference time is higher than others but is left behind in frames processing. the client can't afford FPGA because they cost alot more than their budget, also the CPU is used for other tasks and IGPU will require more power and may affect the performance of these CPU as they share the same memory. *Due to these reasons we can conclude that, the best option for this client is to use the intel NCS2, a device with Mariad-X VPU.*