

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

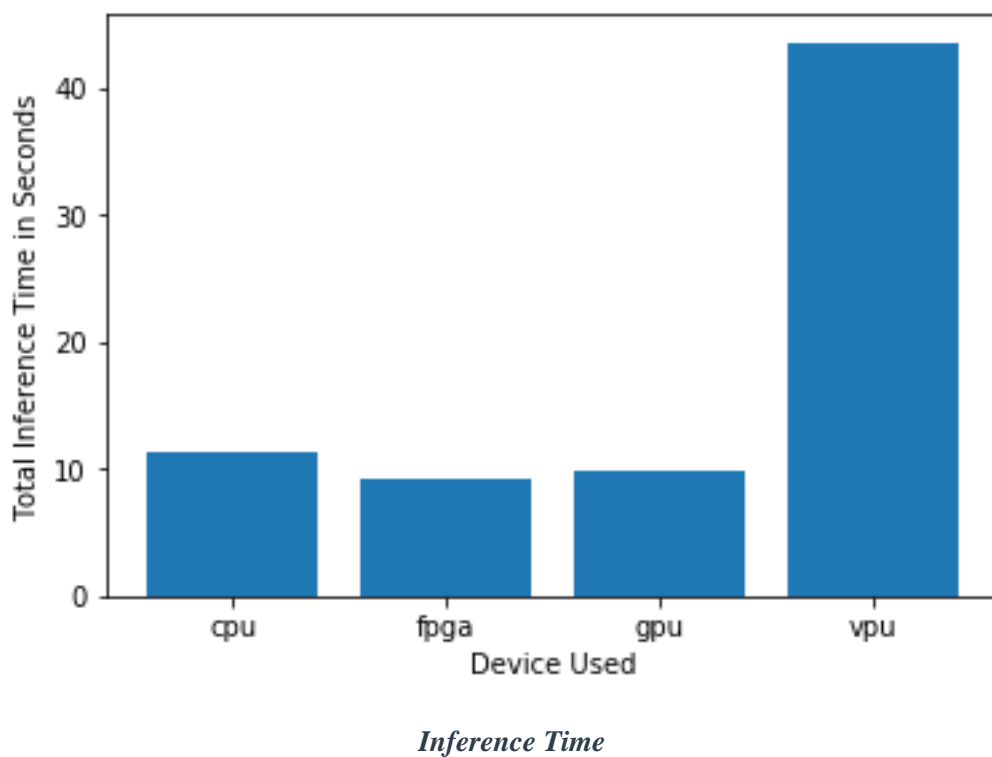
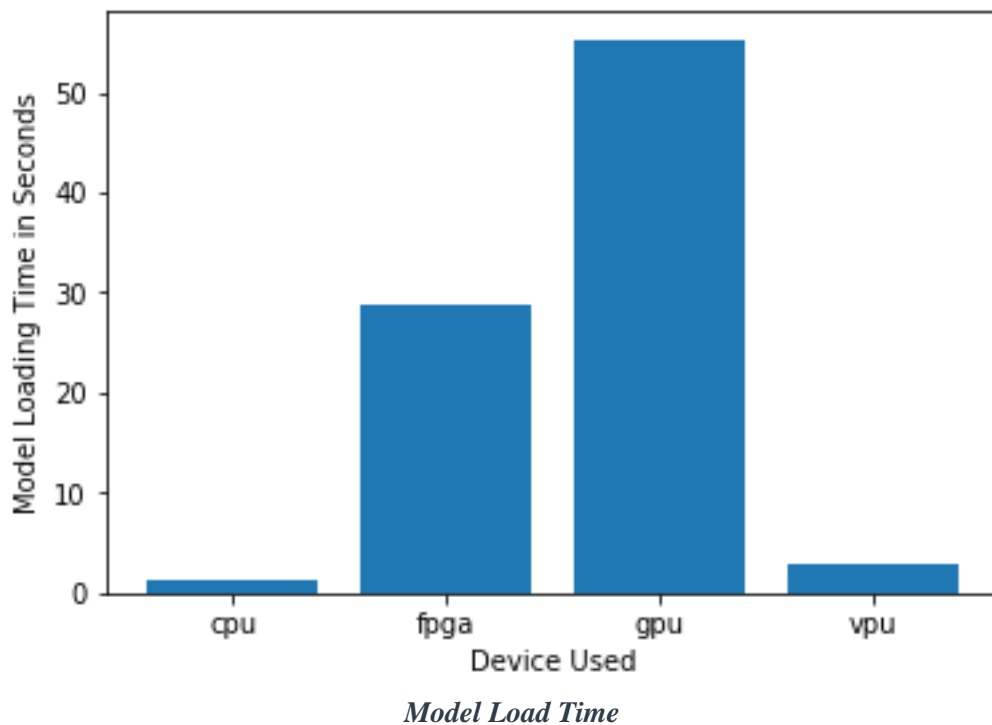
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device.	<i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
<i>The client has plenty of revenue and wants to install a quality system on the factory floor that lasts 5 to 10 years.</i>	<i>FPGAs are robust and have a lifespan of ~10 years. They are also able to function over a wide range of temperatures, from 0° C to 60° C. This means that FPGAs can be deployed in harsh environments like factory floors and still perform optimally.</i>
<i>The client wants to develop another solution later to address a second problem that requires reprogramming of systems every time a new chip design is being manufactured.</i>	<i>Field programmable gate array (FPGA) is a flexible device that can be reprogrammed to meet the needs of the client. The bitstreams can be updated without replacing the hardware every time.</i>
<i>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.</i>	<i>The various precision options (FP16, 11 and 9 bit) are supported—allowing developers a balance between speed and accuracy. FPGAs are designed to have 100% on-time performance, meaning they can be continuously running 24 hours a day, 7 days a week, 365 days a year.</i>

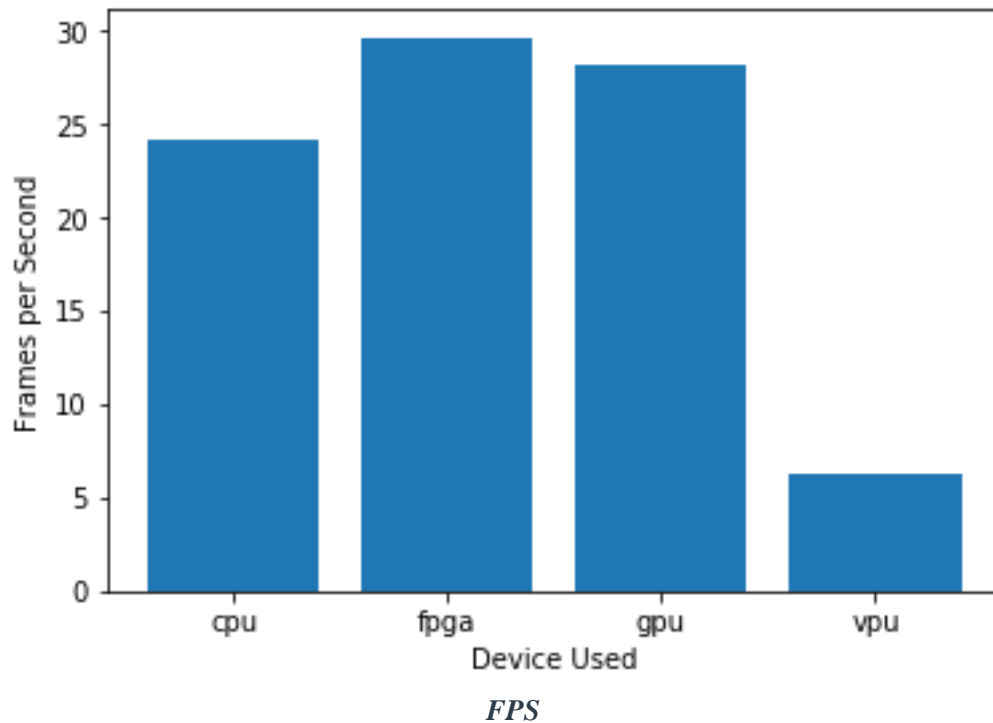
Queue Monitoring Requirements

Maximum number of people in the queue	5 people
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).





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]

- The client wants a reliable and robust system that can last around 5 to 10 years, FPGA has a proven life of ~10 years.
- This hardware type is flexible, can be reprogrammed according to the client's needs and there will be no hardware replacement required later.
- The client wants analysis on 30-35 frames per second video and FPGA can read 30 FPS also the inference time of FPGA is lowest however the model loading time of FPGA is a lot more than the CPU and VPU.
- FPGA is highly recommended hardware for the client's requirements and we can compromise on the model loading time.

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)
IGPU

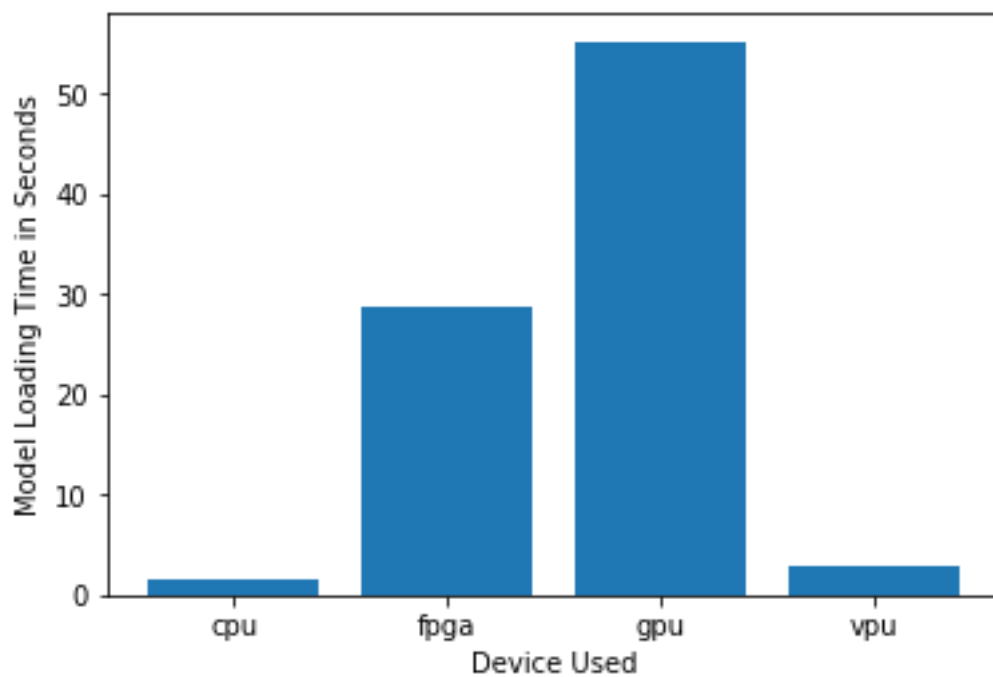
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device.	<i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
<i>The client doesn't have much money to invest in new hardware.</i>	<i>The store's checkout counter already have a modern computer, each of which has an Intel i7 core processor, therefore, we can utilize the integrated GPU for the client's requirements and there is no need to get new hardware</i>
<i>The client want to save on the electricity bill</i>	<i>As there is no additional hardware is going to add and the existing systems can be power down during weekdays to save electricity.</i>

Queue Monitoring Requirements

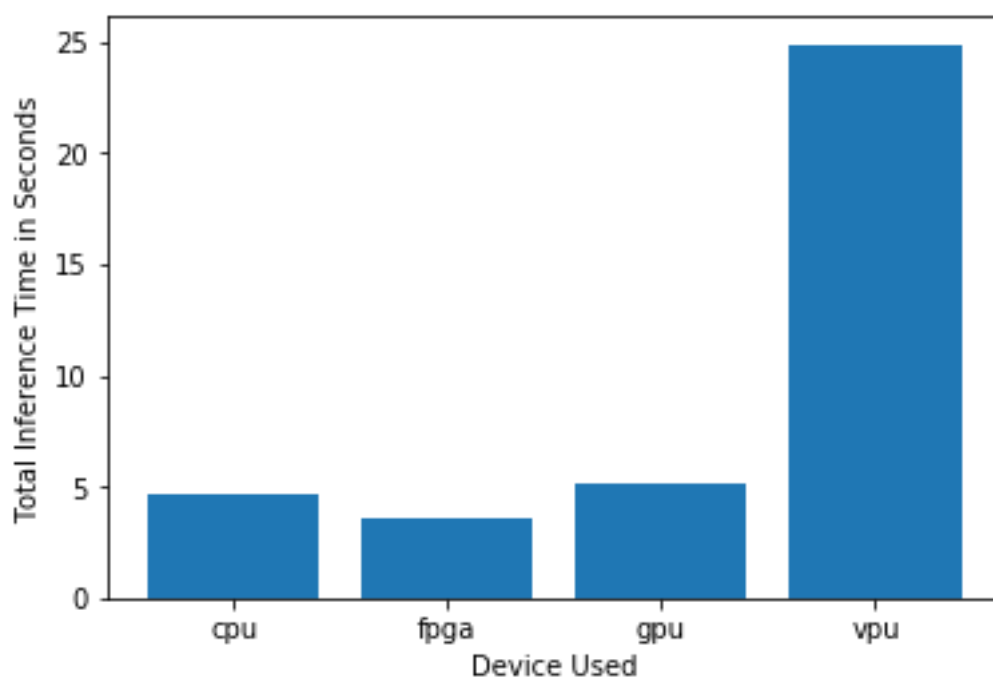
Maximum number of people in the queue	5 people
Model precision chosen (FP32, FP16, or Int8)	FP32

Test Results

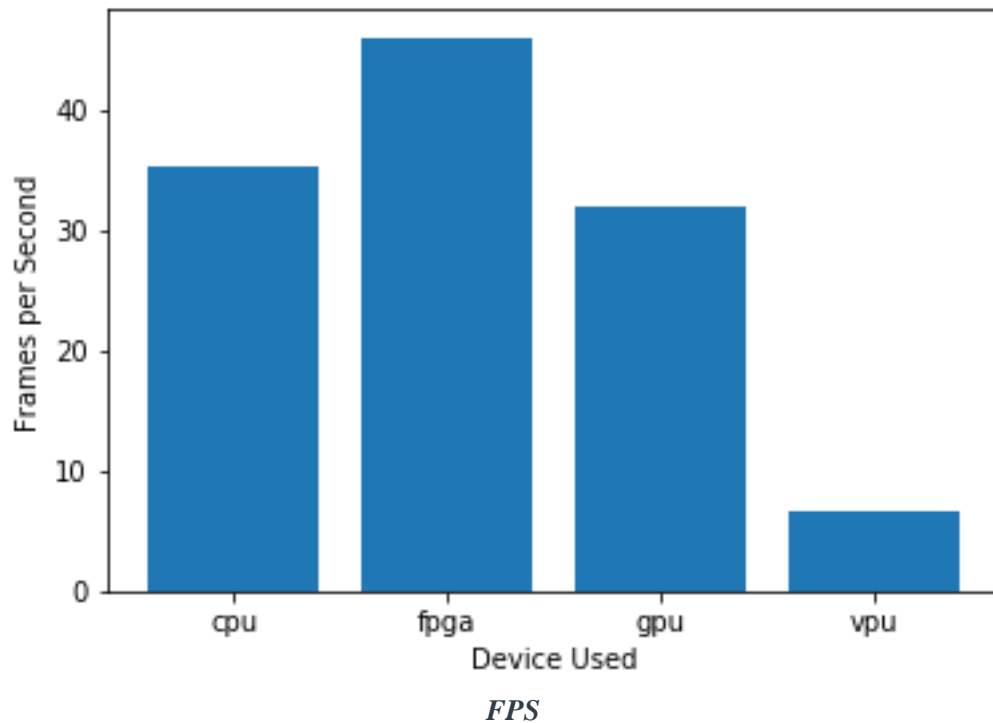
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



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

[IGPU]

- As the client doesn't have enough money to upgrade the hardware and the already available systems are equipped with Intel's i7 processor that is not performing any computational task. Therefore we can utilize the integrated GPU to fulfill the client's requirements.
- From the hardware analysis it is pretty much clear that the model loading time of the IGPU is a lot more than the other hardware devices that is more than the 50s, the frames per second are around 30 and the inference time is around 5s which is more than the FPGA. We can compromise on model loading time as it is the onetime task and the other parameters are fulfilling the requirements therefore the final recommendation is IGPU.

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

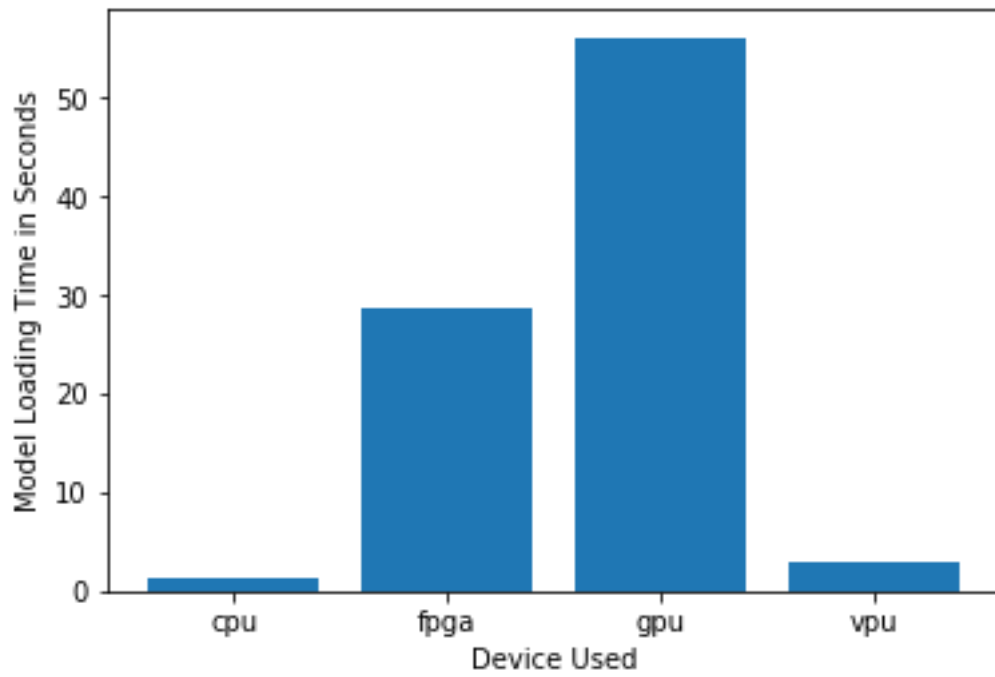
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device.	<i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
<i>The client only have \$300 budget.</i>	<i>VPU or NCS2 is a low budget device and cost less than \$100 each.</i>
<i>The available system is being used to process and view the video stream of 7 CCTV cameras and not enough processing power available to run inference in real time.</i>	<i>VPU's have low latency and can be used for real time processing of video streams. These are also low powered devices and fairly enhance computational speed.</i>

Queue Monitoring Requirements

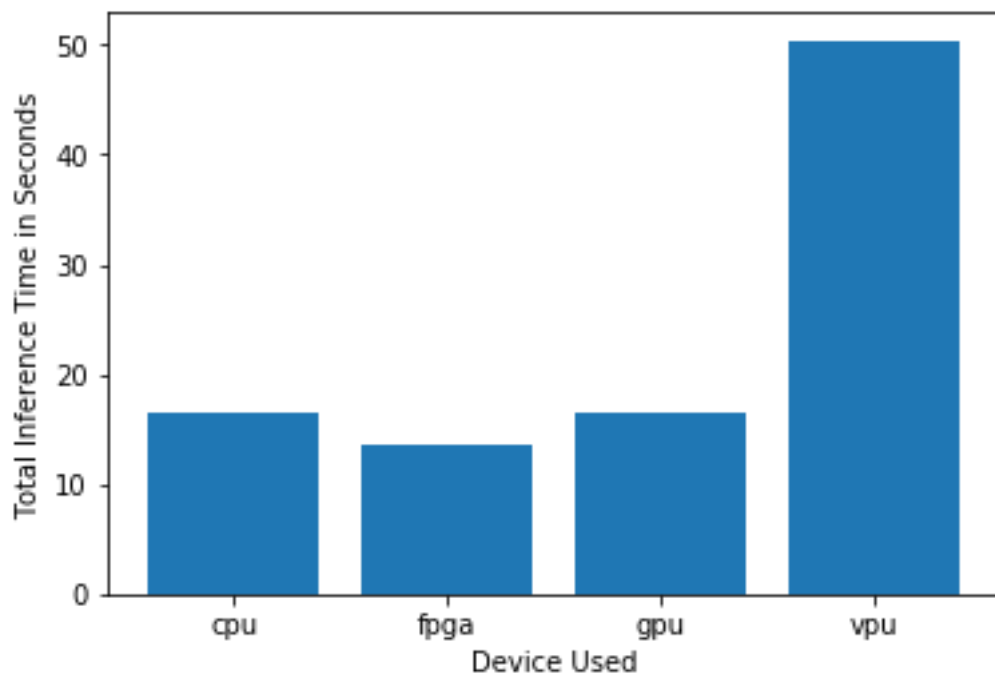
Maximum number of people in the queue	10
Model precision chosen (FP32, FP16, or Int8)	FP32

Test Results

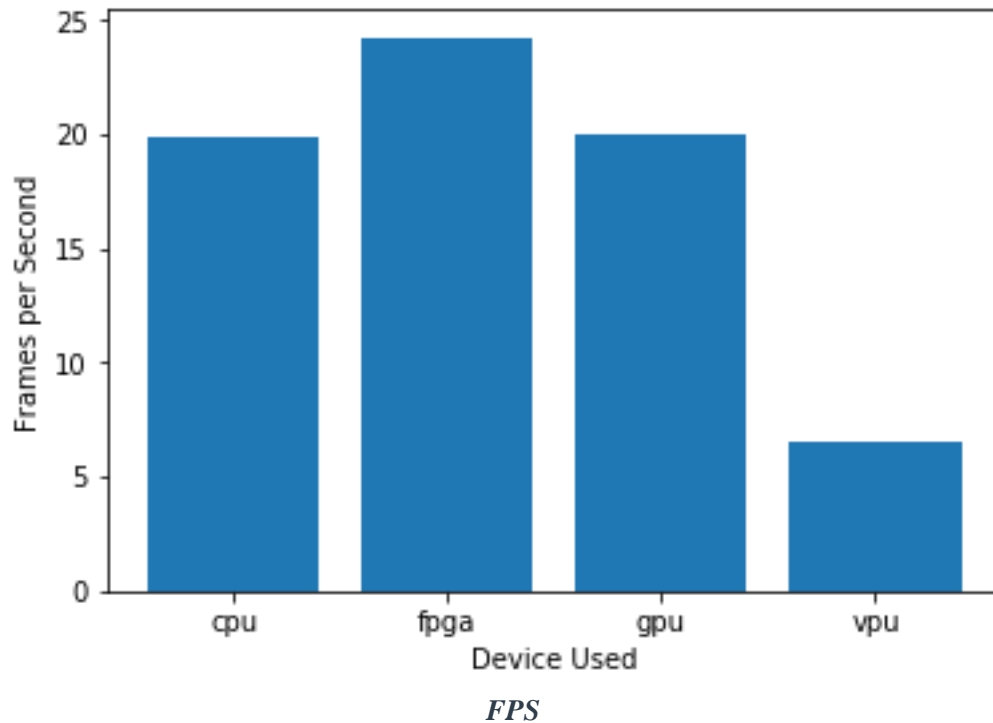
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



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]

- The model loading time in case of VPU is much less than the GPU and FPGA.
- The frames per second are lowest among all the hardware and the inference time is highest, we can't use the CPU power as it is already being used for CCTV video processing and viewing. The FPGA solution is very much costly for the client, therefore, VPU is the recommended solution to fulfill the requirements of low power solution within the budget.