

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)
The most appropriate hardware for this scenario is FPGA

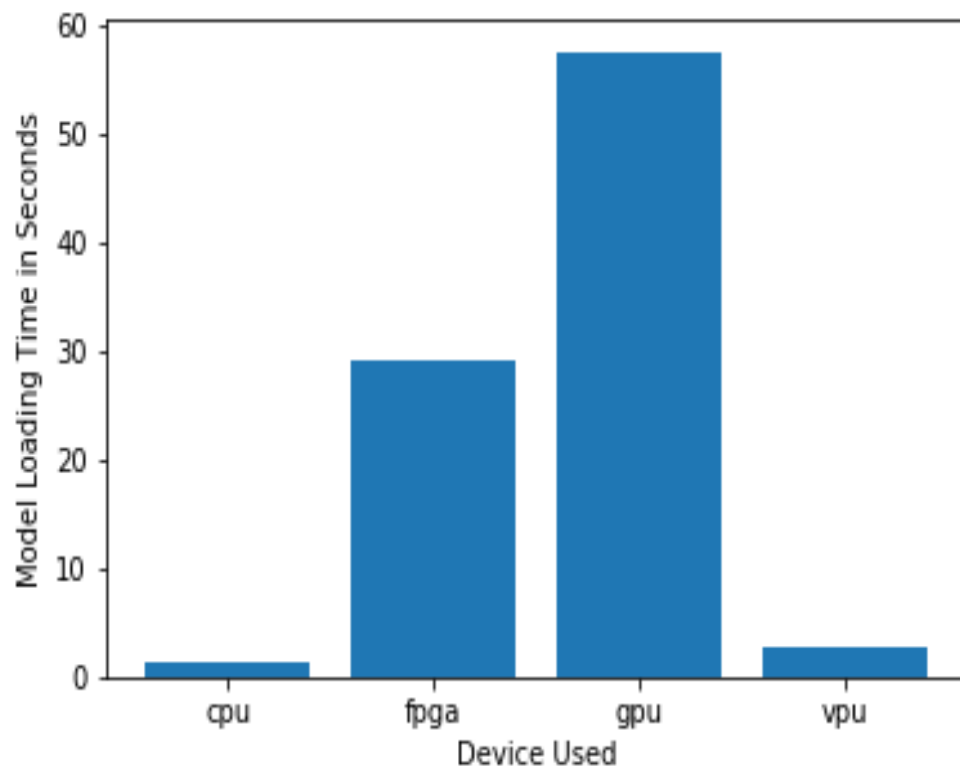
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device.	<i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
The company want the image processing task to be completed five times faster than the existing system.	FPGA have High performance, low latency
The company need a flexible system that can be reprogrammed and optimized to quickly detect flaws in different chip designs	FPGA are field-programmable i.e they can be reprogrammed to adapt to new, evolving, and custom networks
The company has enough budget to install high quality system that can last for 5 to 10 years	FPGAs from Intel's Internet of Things Group have a guaranteed availability of 10 years, from start of production and an uptime of 24/7 in 365 days.

Queue Monitoring Requirements

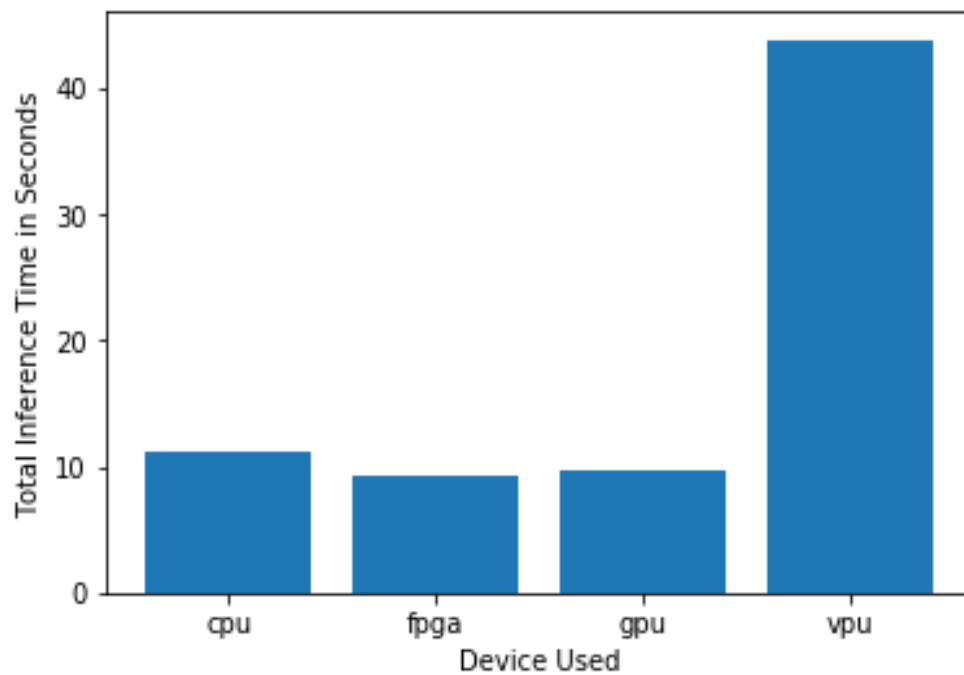
Maximum number of people in the queue	6
Model precision chosen (FP32, FP16, or Int8)	FP32 for CPU, and FP16 for GPU, VPU and FPGA

Test Results

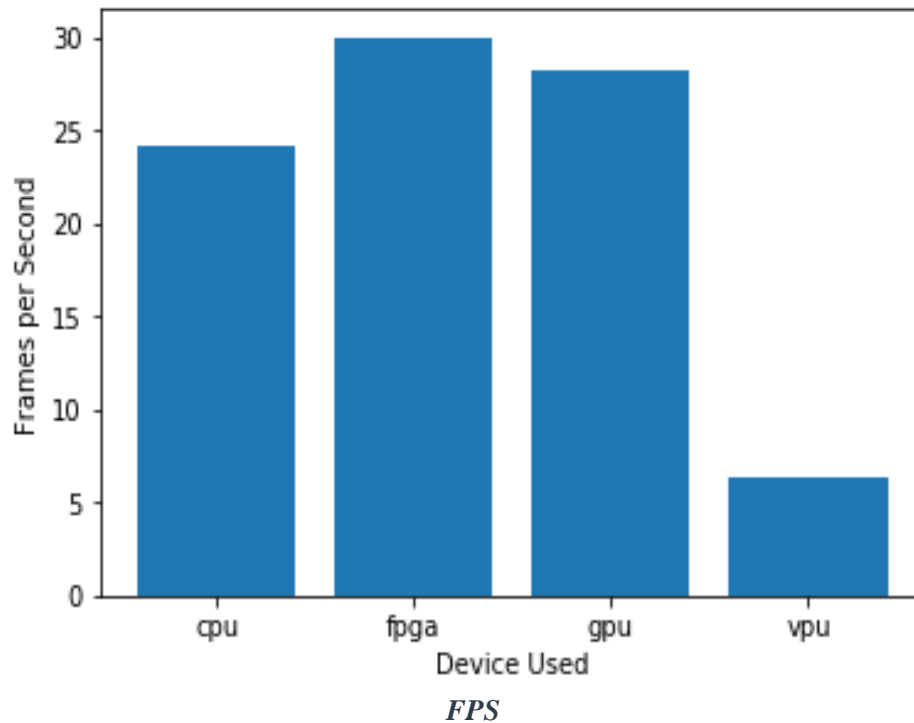
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

Naomi Semiconductors want a robust system that can perform image processing task five times faster than the existing system i.e run inference on the video stream very quickly, the company also need a flexible system that can be reprogrammed and optimized to quickly detect flaws in different chip designs, and finally, they company have enough budget to install high quality system that can last for 5 to 10 years

The final recommended hardware is FPGA, since the client's most important requirements is a system with high performance, flexibility and a long lifespan. Comparing the FPGA's performance with other hardware, the FGPA has a slow model loading time. Considering that the FPGA outperforms other hardware's in terms of inference time and frames per second, the slow loading time (which is a one-time event) is a good trade-off for inference speed, FPS, long lasting and quality hardware provided by the FPGA.

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)
The most appropriate hardware for this scenario is CPU

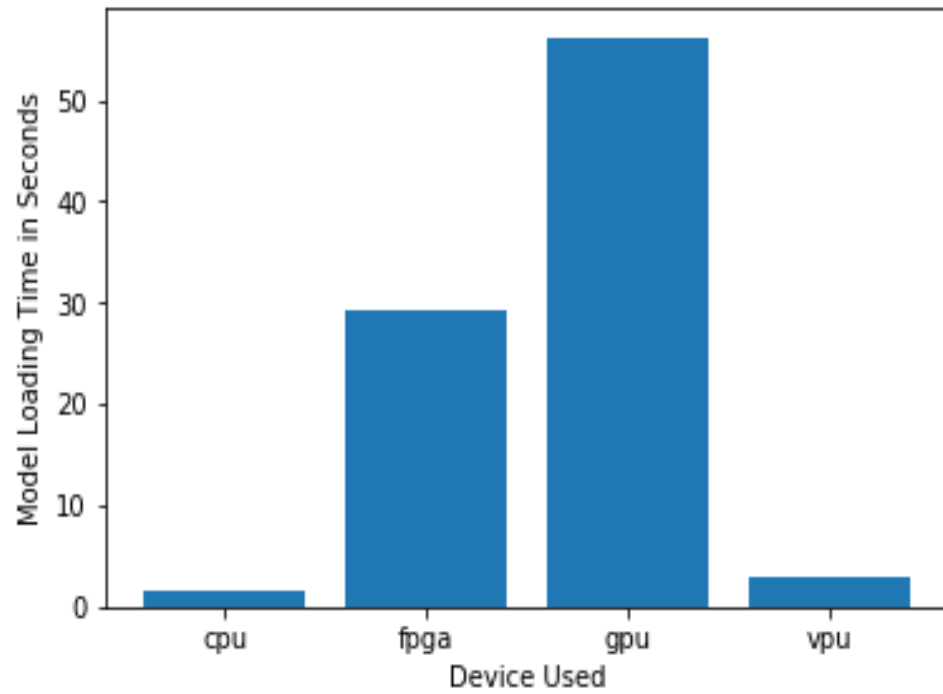
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device.	<i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
The client don't want to incur additional electric bill charges	Using the existing CPUs in place will not add additional electrical bills
The client does not have much money to invest in additional hardware	We can fully utilize the existing hardware. (CPU)

Queue Monitoring Requirements

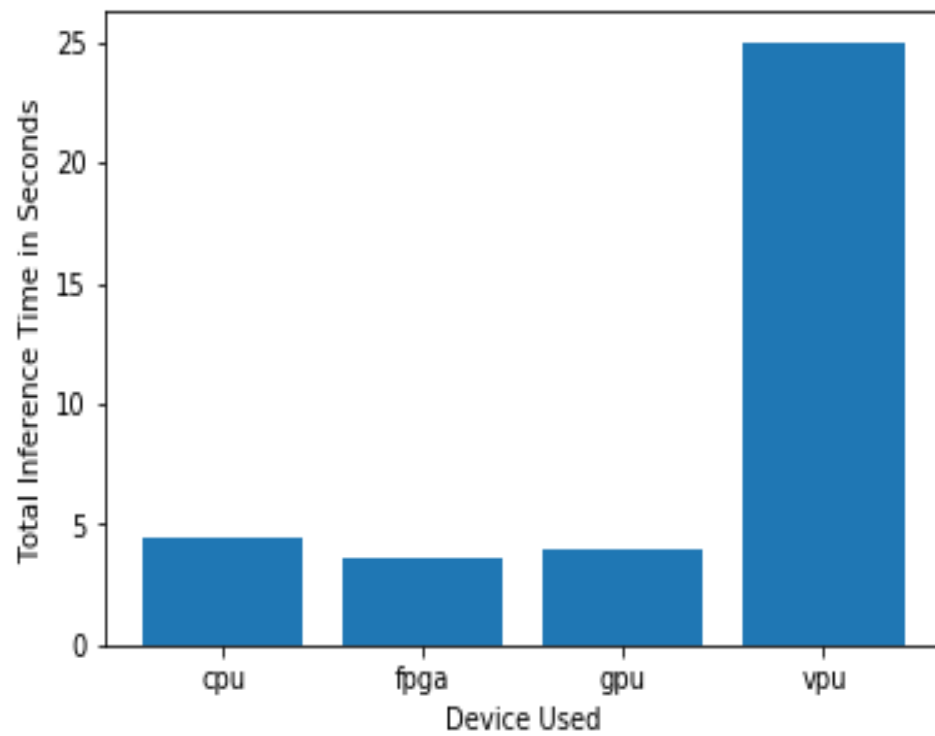
Maximum number of people in the queue	5
Model precision chosen (FP32, FP16, or Int8)	FP32 for CPU, and FP16 for GPU, VPU and FPGA

Test Results

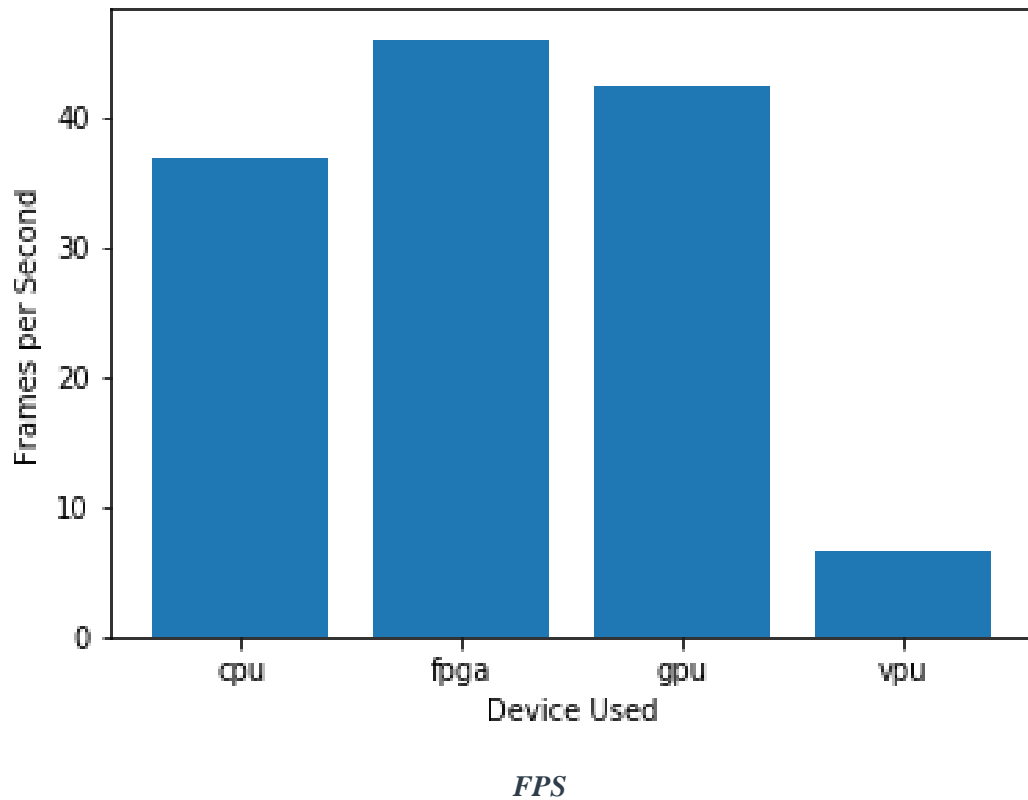
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

Mr. Lin does not have enough money to invest in buying additional hardware, and he is not willing to incur additional expenses on his electric bill. 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.

Since Mr. Lin has computationally under-utilized CPUs at his facility and is concerned about electrical bills, making use of the existing CPUs will not incur additional expenses. Even though FPGA and VPU outperform the CPU in terms of inference time and speed, they are going to incur more capital and electrical cost on the company.

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)
The most appropriate hardware for this scenario is VPU

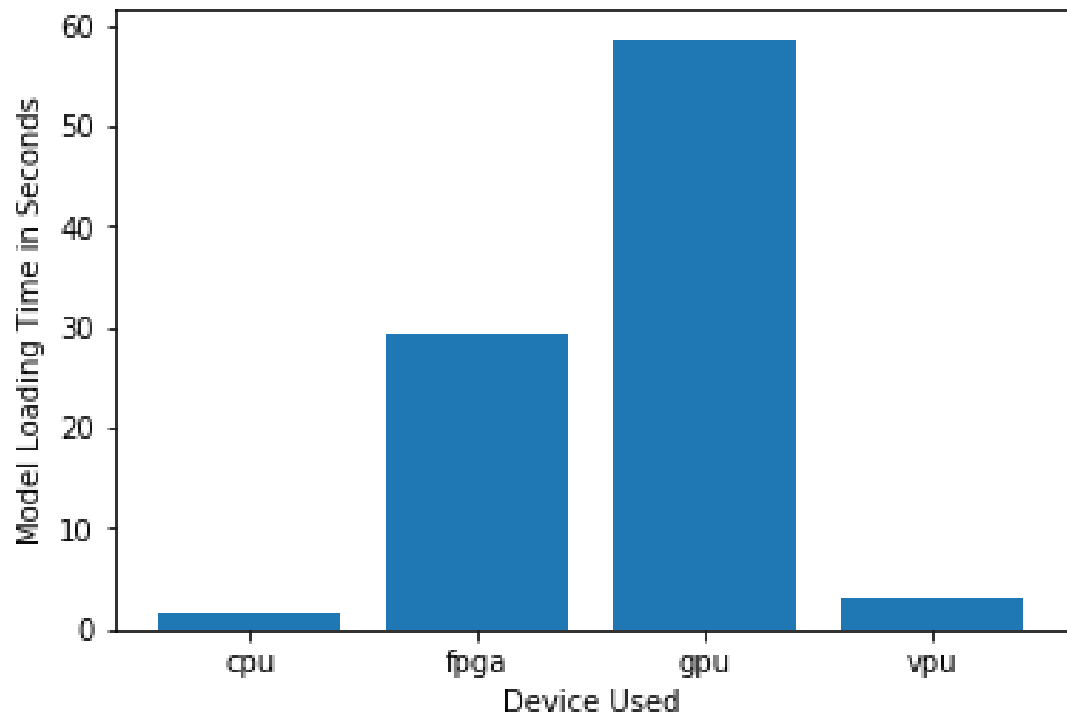
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device.	<i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
The client want an Edge AI system that would monitor the queues in real-time and quickly direct the crowd in the right manner.	VPU is a low cost device that can be used to accelerate the performance of an existing system.
In this scenario, no additional processing power is available.	NCS2 is required ~1W for its power consumption and it is good fit in this scenario.
The budget of the client is fixed and very low.	NCS2 price is approximate \$100 and would fit in the price range.

Queue Monitoring Requirements

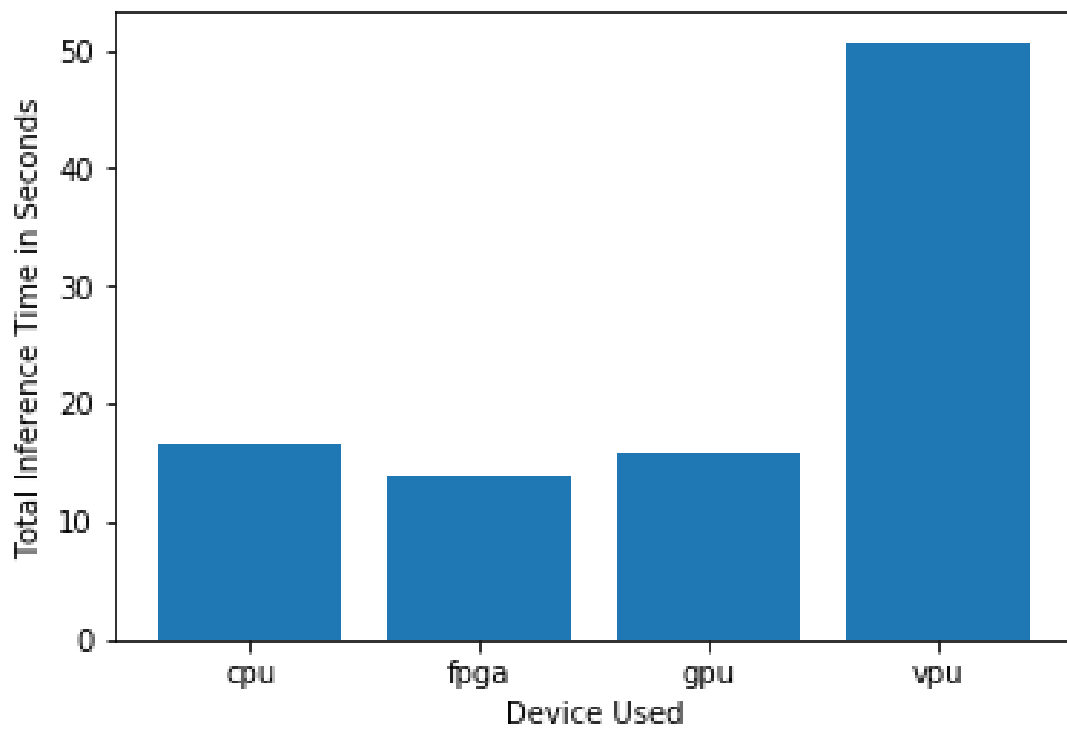
Maximum number of people in the queue	15
Model precision chosen (FP32, FP16, or Int8)	FP32 for CPU, and FP16 for GPU, VPU and FPGA

Test Results

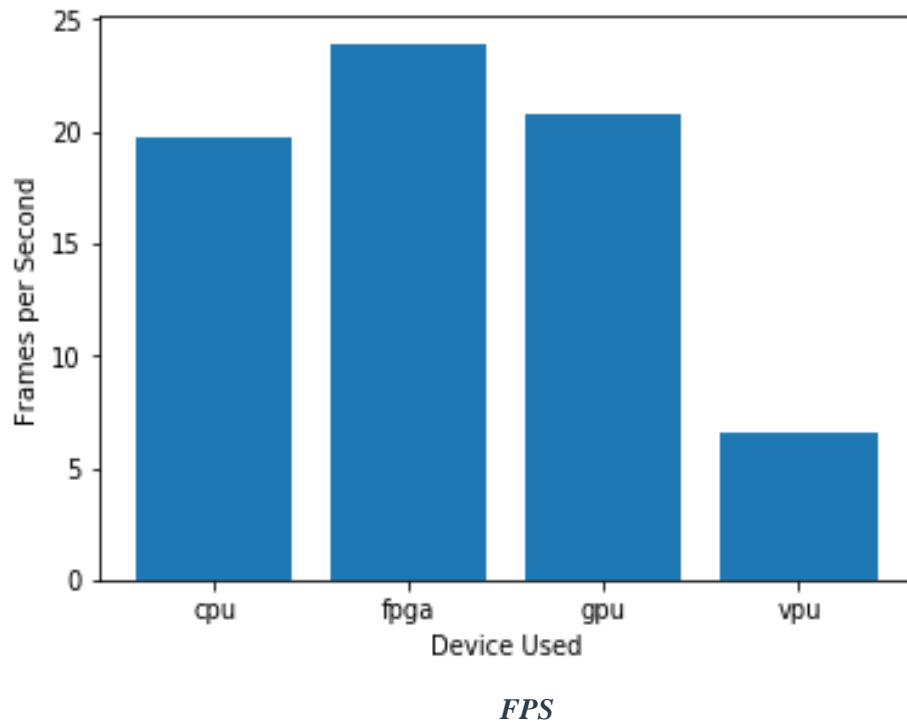
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

The client requires an Edge AI that will monitor the queues in real-time and quickly direct the crowd in the right manner with budget allows for a maximum of \$300 per machine. While the client has PCs on site, the computational resources on the PCs are scarce. Hence, the need for a low power and affordable device for inference. VPU is the recommended hardware for the client. Even though VPU lags other hardware in frames per second and inference time metric, it meets the client's core requirement on budget and energy consumption.