Choose the Right Hardware

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

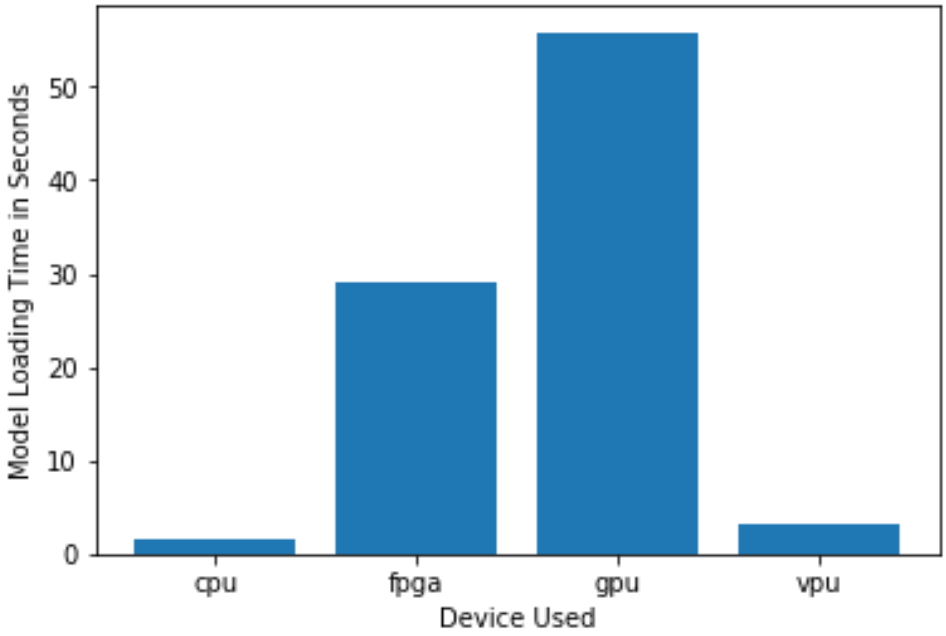
Requirement Observed (Include at least two.)	**How does the chosen hardware meet this requirement?**
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
The client wants to do faster image processing at factory floors	FPGAs have high performance and low latency and are robust which make them work in harsh temperatures like factory
The client requires the investment to last for 5-10 years	FPGAs have long life span upto 10 years
Because there are multiple chip designs and new designs are created regularly .The client requires a system that would be flexible so that it can be reprogrammed and optimized to quickly detect flaws in different chip designs.	FPGAs are flexible. They can be reprogrammed to adapt to new, evolving, and custom networks

Queue Monitoring Requirements

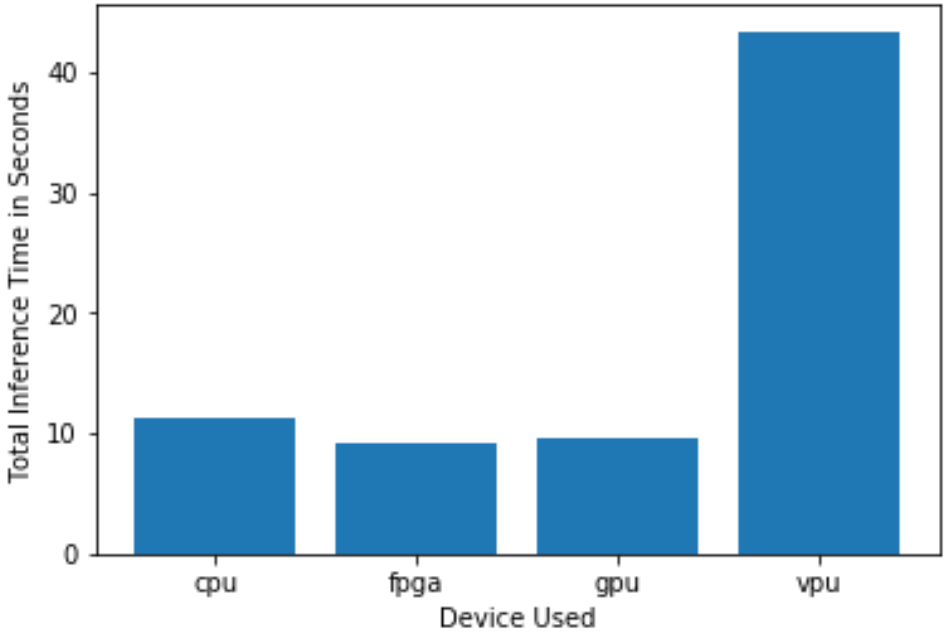
Maximum number of people in the queue	4
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

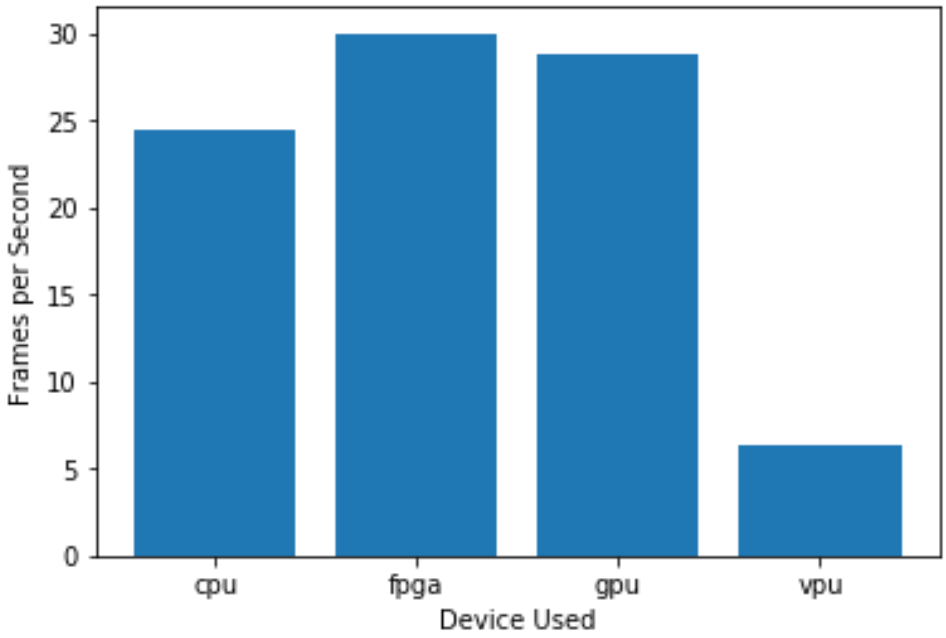
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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. It is seen that FPGA takes approximately less than 10seconds to load the model which is comparatively faster compared to other devices .The client already has a camera which records 30-35 Frames Per Second and from the above FPS graph seen that FPGA can process around 25-30 frames which is the higher compared to other devices. The load time of the model is around 25-30 seconds which is around 4 times more than CPU(less than 5 seconds) and VPU(less than 8 seconds). This one of the major drawback but the clients requirement was to have a quick inference time and has to adopt to new design requirements. Hence FPGA is the final hardware recommendation for this scenario

---

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

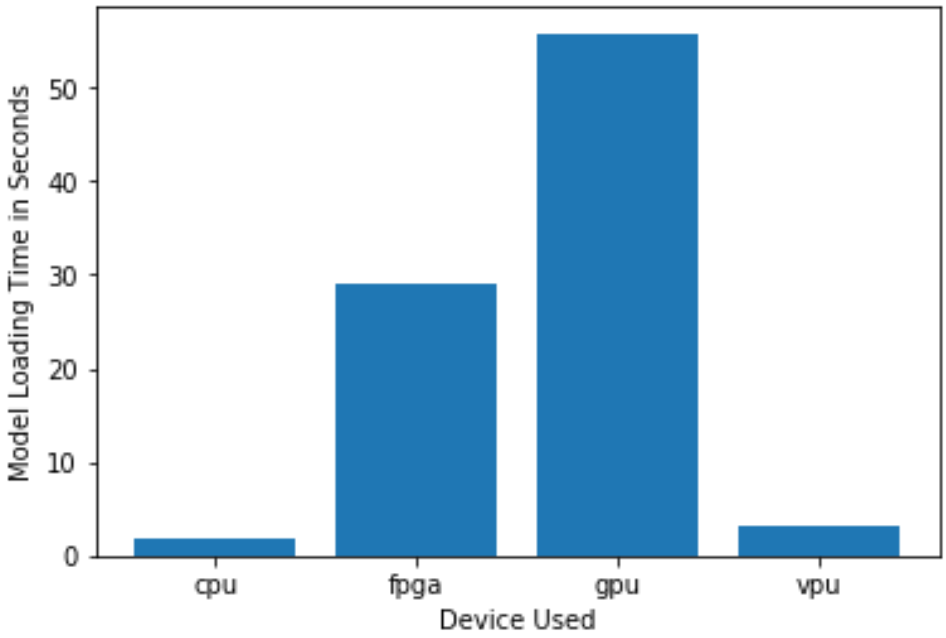
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
The client doesn't have much money to invest in additional hardware	So based on the current cost, the existing i7 CPU and IGPU could be used for the task
The client wants to implement a queue monitoring system to direct the consumers to less congested bill desk	The i7's IGPU processor is powerful enough to do fast computing and gives high inference speed
The client doesn't want to add additional overhead to electricity	Any additional hardware might lead to increase in power consumption. So it's better to use the existing IGPU since the IGPU provides configurable power consumption

Queue Monitoring Requirements

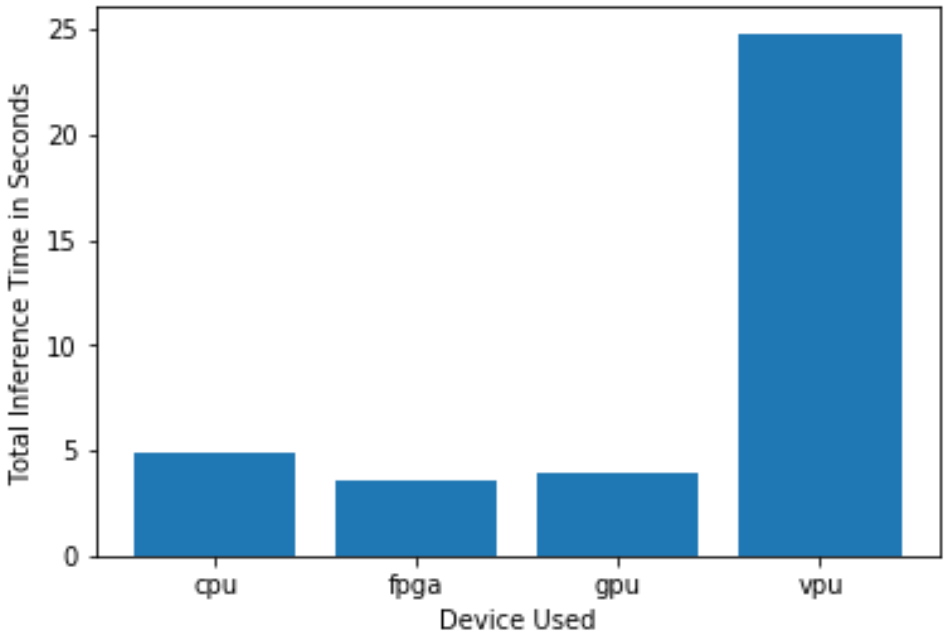
Maximum number of people in the queue	2
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

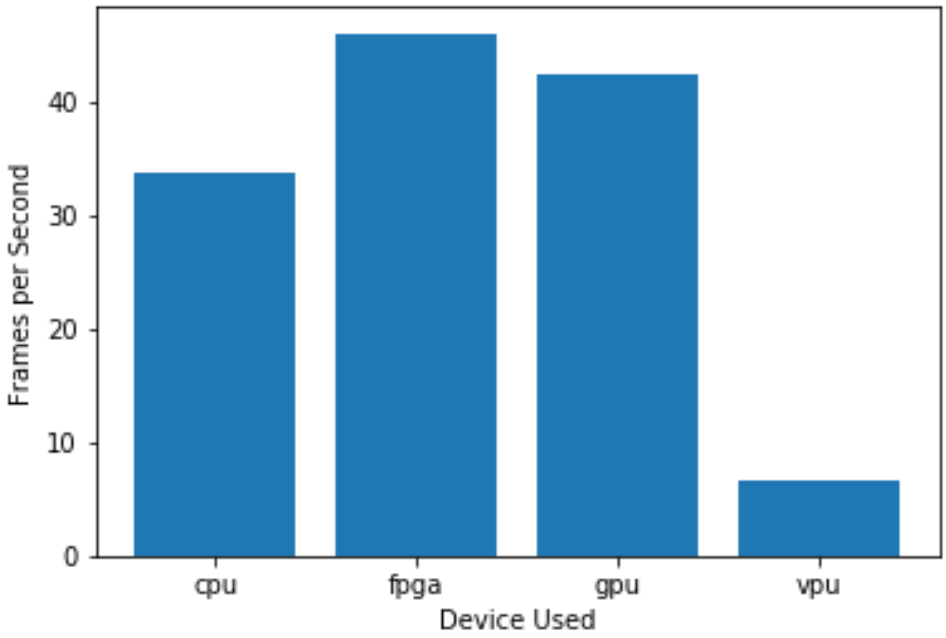
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time

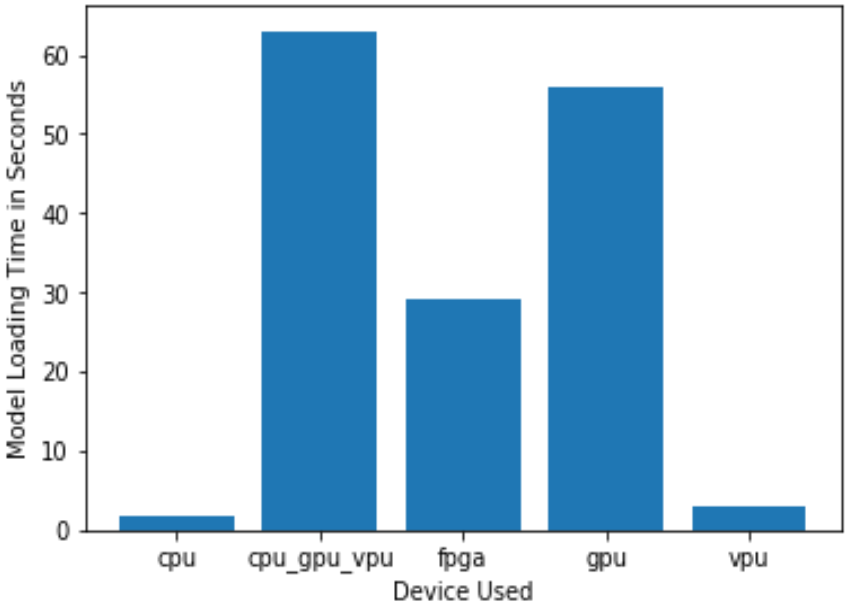


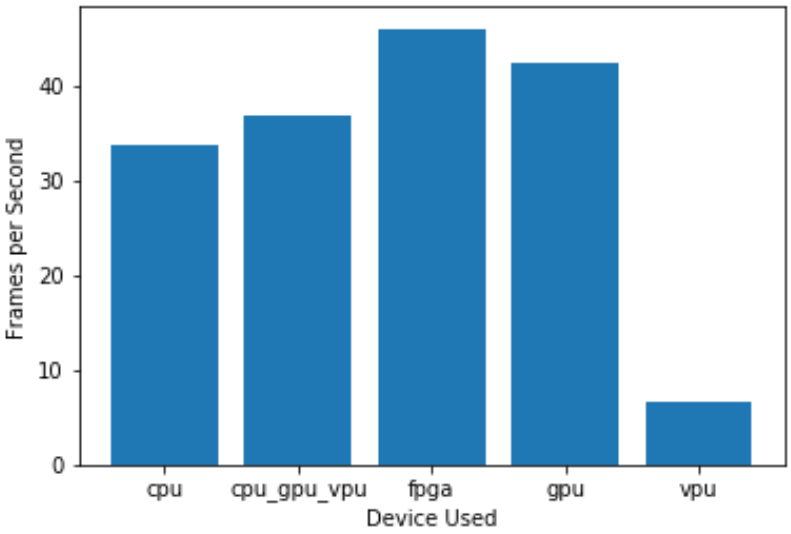
FPS

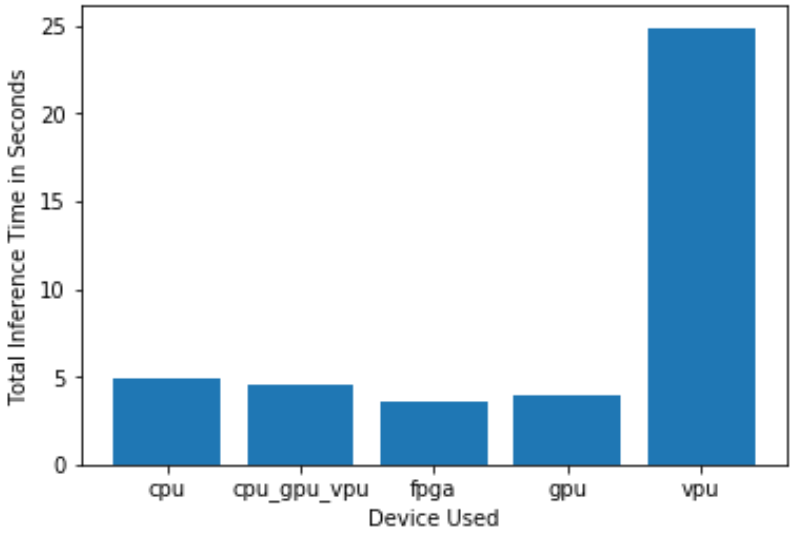
Final Hardware Recommendation

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
IGPU The client requirement is too have less- congested queues in the store to increase the sales by directing people to less-congested queues in the store. From the above graph it can be seen that IGPU has an inference time of less than five seconds but higher than FPGA. More over its seen that it has a good processing frames per second around 30-35FPS. But has a high model load time compared to other devices around 50seconds. Since model loading is a single time job it can be neglected here. FPGA can be considered as a good option here but the client doesn't want to invest in additional hardware the inbuilt i7's IGPU can be a good fit for this scenario as they are using the system for minimal requirement only . We can also the CPU but from the above graph its seen that GPU has less inference time compared to CPU and can process more number of frames per second than CPU


Model Loading Time in Seconds
0
10
20
30
40
50
60
cpu
cpu_gpu_vpu
fpga
gpu
vpu
Device Used


Frames per Second
0
10
20
30
40
cpu
cpu_gpu_vpu
fpga
gpu
vpu
Device Used



I have tried multi-device plugin since a vpu would cost less than $100 and wouldnot incur too much cost and the client can bare this.It's seen that using multi plugin using VPU doesn't provide much improvement compared to the GPU alone. Hence using the IGPU is best in this scenario

---

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.
Their budget is only about $300 for each device.	VPU or NCS2 would fit in the price range.
The clients existing CPU has no additional processing power available to run inference	VPU's have dedicated hardware accelerator which are optimized for deep learning applications

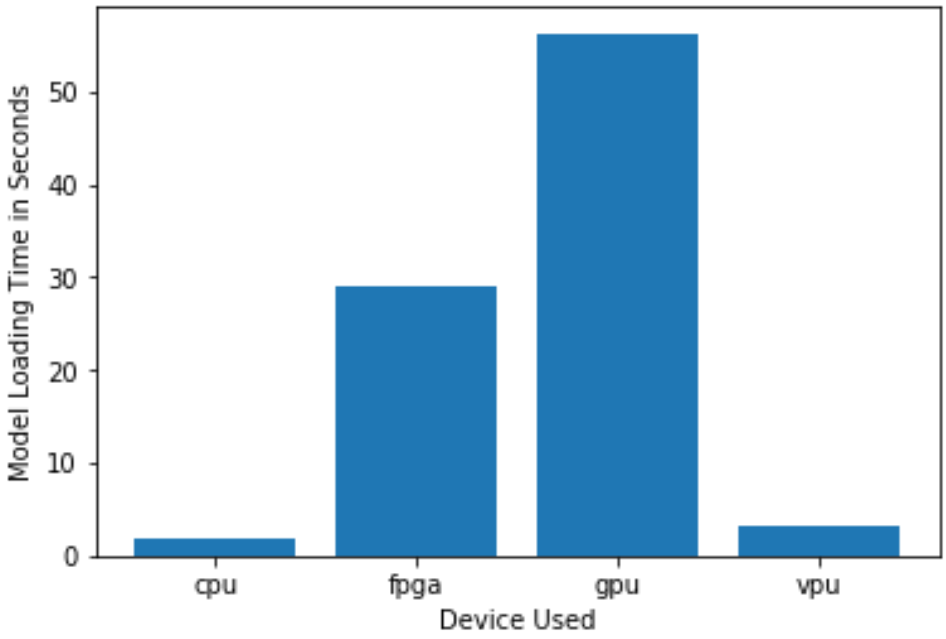
The client wants to monitor the queues in real-time and quickly direct the crowd in the right manner	VPU's have low latency hence real time processing is possible

Queue Monitoring Requirements

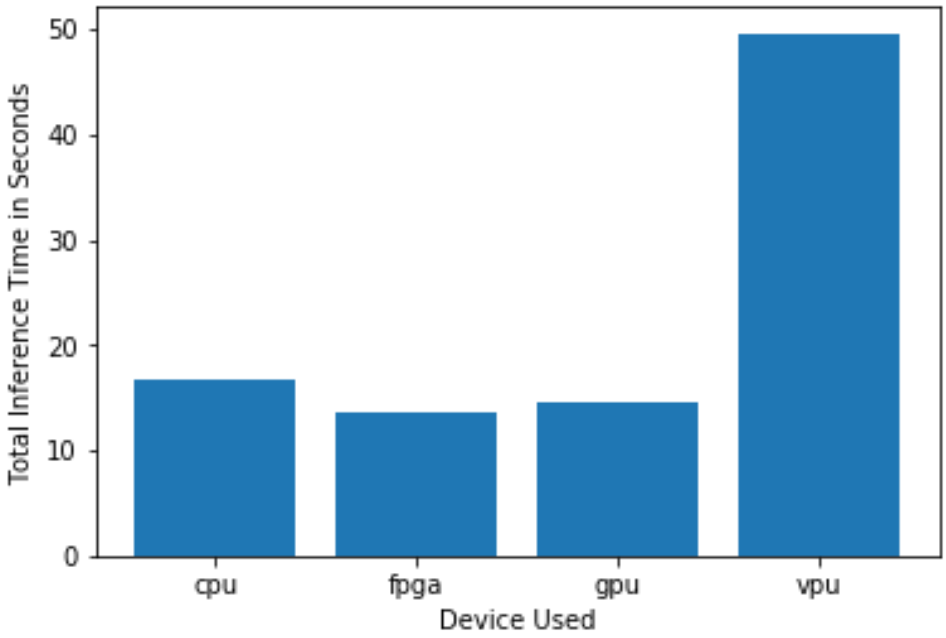
Maximum number of people in the queue	7
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

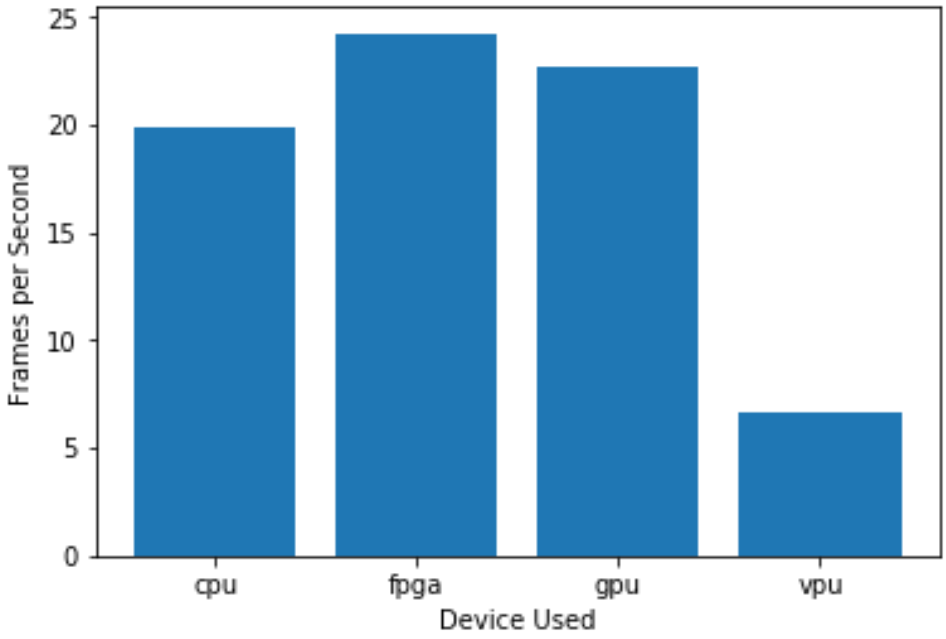
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

It is seen from the above graph VPU has the low model load time less than 5 seconds .But it has higher inference time of around 50seconds and has low processing frames per second. Using an inbuilt CPU for this scenario would have provided better results, but the CPU's are currently being used to process and view CCTV footage for security purposes and no significant additional processing power is available to run inference. Since CPU is not available we can use FPGA to provide good results but the client budget is a maximum of $300 per machine using VPU is the best option here