

# 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)
<a href="#">IEI Mustang-F100-A10</a> FPGA is most appropriate for this scenario.

Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
The factory has a vision camera installed at every belt. Each camera records video at 30-35 FPS that can be used to monitor the number of people in the factory line.	<a href="#">IEI Mustang-F100-A10</a> FPGA can server at this processing speed. With various precision options supported, developers can balance between speed and accuracy.
Client wants the image processing task to be completed five times per second.	<a href="#">IEI Mustang-F100-A10</a> FPGA can meet this requirement.
To be able 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.	<a href="#">IEI Mustang-F100-A10</a> FPGA meets this requirement. It can execute neural networks with high performance and very low latency. The high performance comes from the ability to run many sections of the FPGA in parallel.
While Naomi Semiconductors 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.	<a href="#">IEI Mustang-F100-A10</a> FPGA is robust and designed to have a guaranteed availability of 10 years, from start of production.

the system would also need to be flexible so that it can be reprogrammed and optimized to quickly detect flaws in different chip designs.

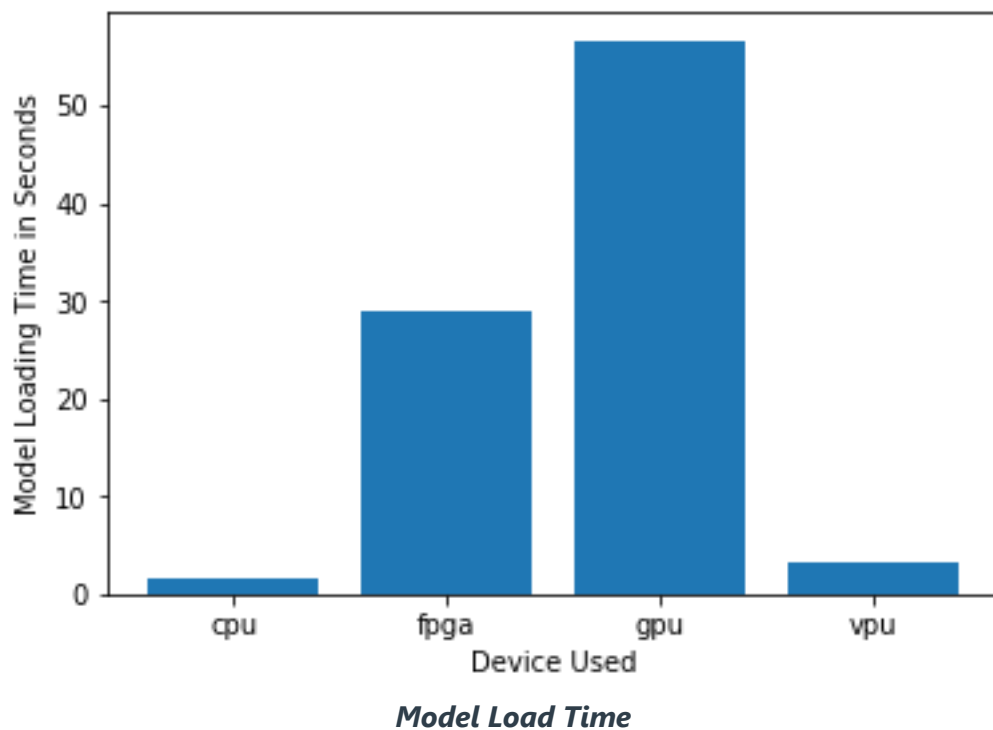
FPGAs are field-programmable; they can be reprogrammed to adapt to new, evolving, and custom networks

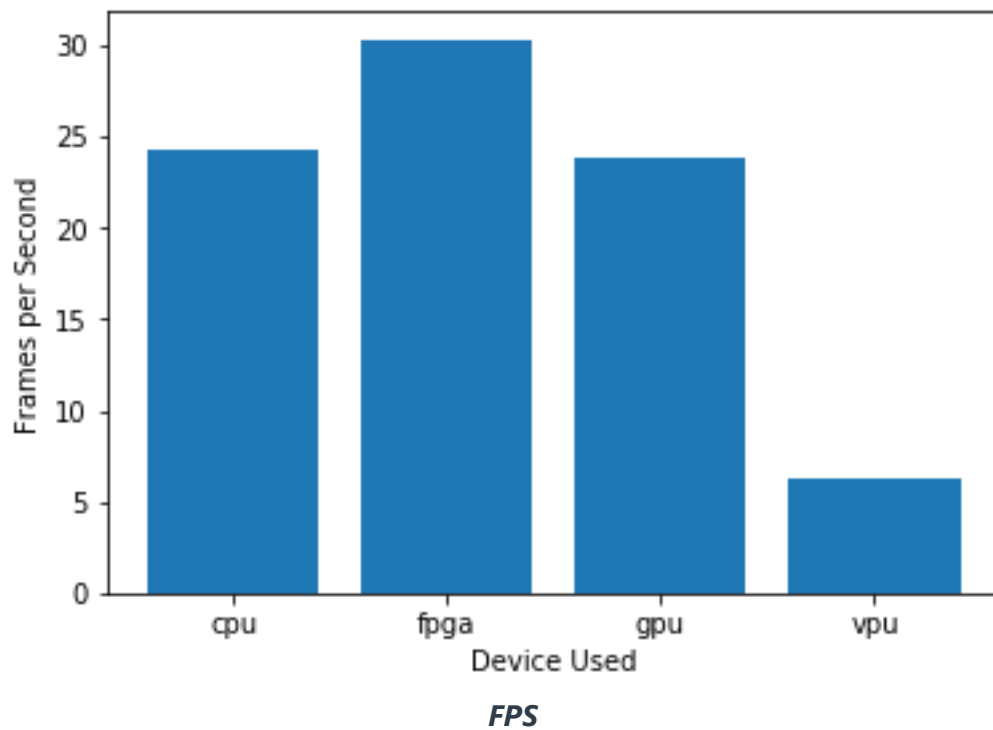
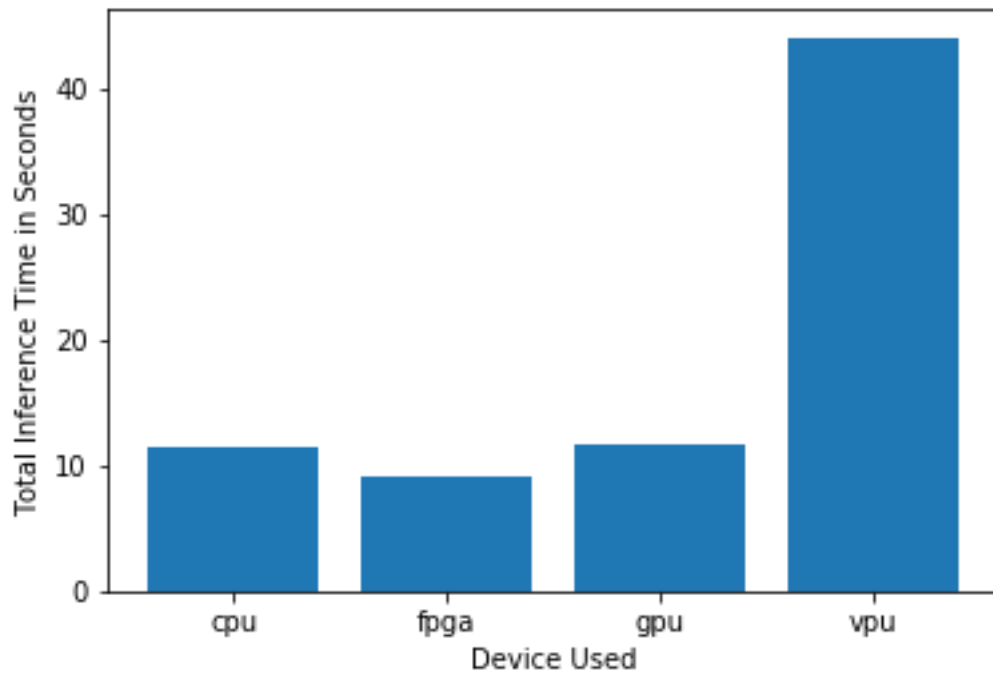
## Queue Monitoring Requirements

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).





## 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 this scenario. The test result shows that inference time is the lowest in FPGA that meets client's requirement. Also, FPS is highest for FPGA that meets the requirement (30-35 FPS). Client wants the system to be flexible so that it can be reprogrammed. FPGA provides this facility. Also, client wants this system to have a long lifespan. [IEI Mustang-F100-A10](#) FPGA is robust and designed to have a guaranteed availability of 10 years, from start of production.

## 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 - Existing Intel i7 core processor.

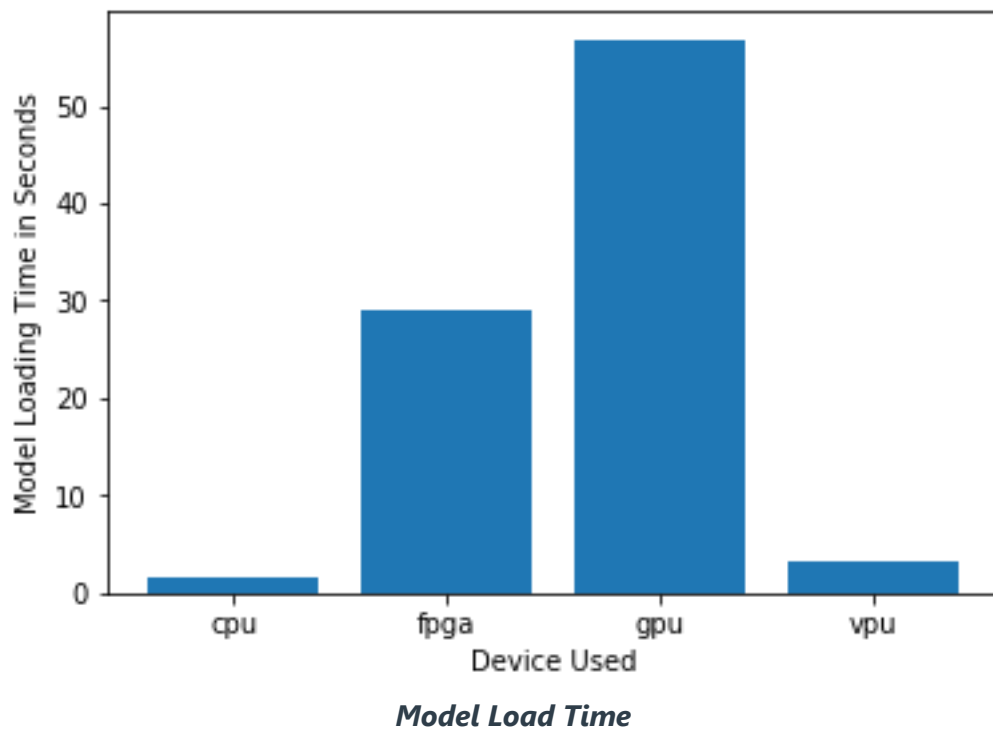
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Each of the store's checkout counters has an Intel i7 core processor. Currently these processors are only used to carry out some minimal tasks that are not computationally expensive.	Existing Intel i7 core processors are not heavily used. Hence these processors can be used for the Edge AI application.
Client does not have much money to invest in additional hardware.	Existing Intel i7 core processors can be used. No extra hardware is required.
Client would like to save as much as possible on electric bill.	Intel i7 core processors are having 65W TDP. Hence power consumption is not too much.
Client does not have any stringent requirement on speed.	Existing Intel i7 core processors with 3 GHz base frequency are enough to meet the requirement.

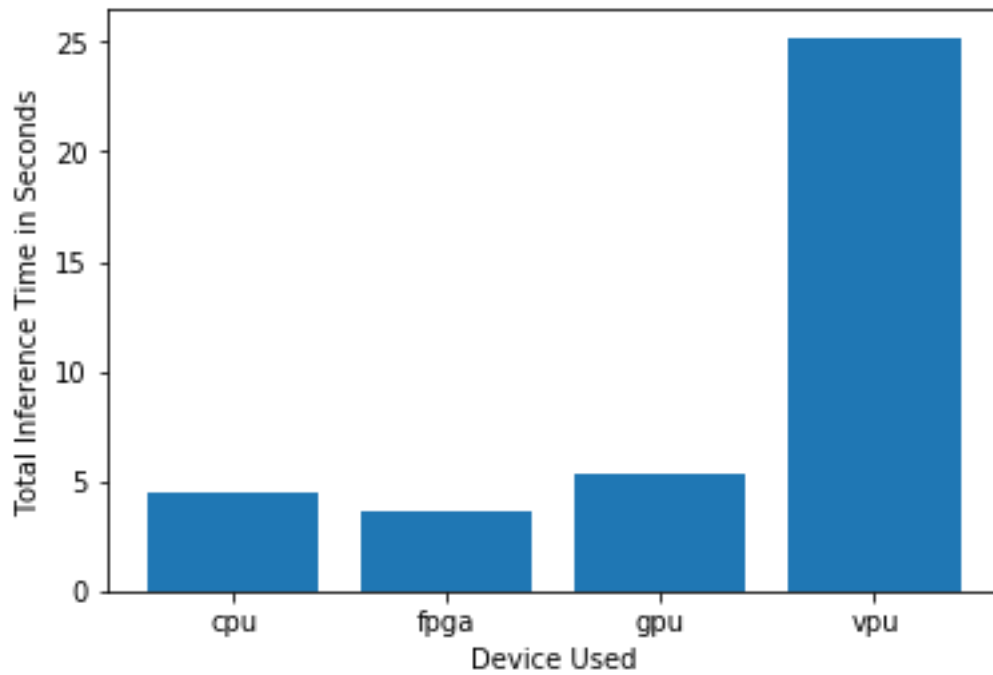
## Queue Monitoring Requirements

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

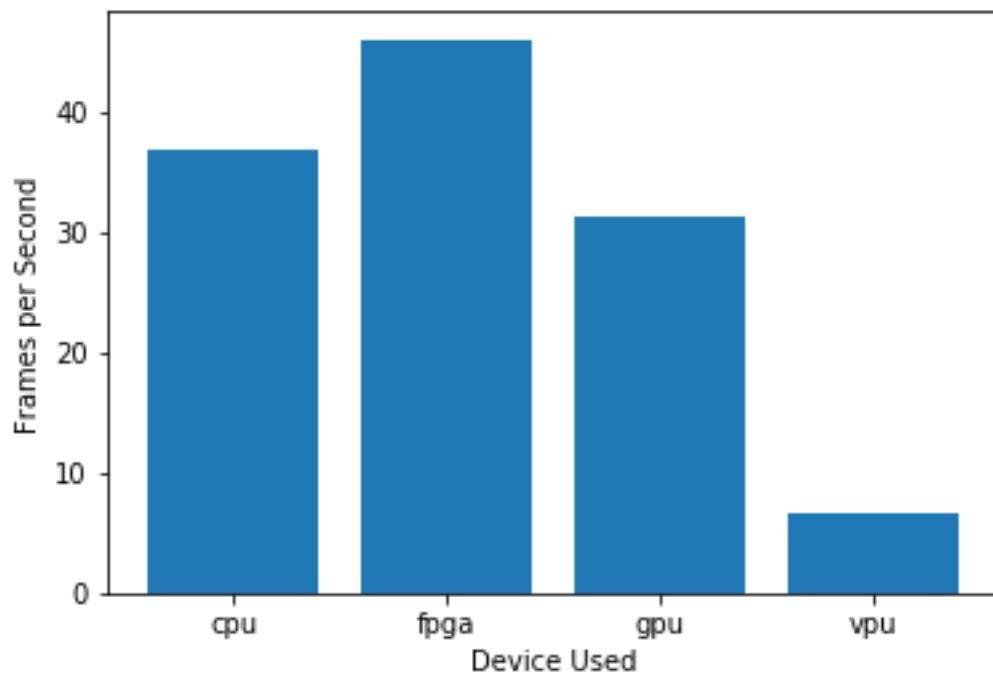
## Test Results

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).





*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

Client has no stringent requirement on speed. From the test result it is visible that CPU provides second best inference time which is quite good to meet the requirement. Frames per second is also second best which meets client's requirement. Hence considering client's low budget to be invested, existing Intel i7 processors with 3 GHz base frequency can be used for the Edge AI application and no other hardware needs to be bought. So, CPU is the final hardware for the retail 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 ([Intel Neural Compute Stick 2](#))

Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
No significant additional processing power is available to run inference in the CPUs of existing machines.	VPU can work as accelerators specialized for AI tasks and add performance to a pre-existing system (CPU).
Client's budget allows for a maximum of \$300 per machine and client wants to save on hardware.	VPU ( <a href="#">Intel Neural Compute Stick 2</a> ) is an inexpensive option, typically costing around \$70 to \$100.
Client wants to save on power requirement too.	VPU ( <a href="#">Intel Neural Compute Stick 2</a> ) is extremely low power device but this can come at some cost to performance. However, client does not have any stringent processing FPS requirement.
They monitor the entire situation with 7 CCTV cameras on the platform. These are connected to closed All-In-One PCs that are in a nearby security booth.	VPU ( <a href="#">Intel Neural Compute Stick 2</a> ) can hit 4 teraflops of compute and 1 trillion operations per second of

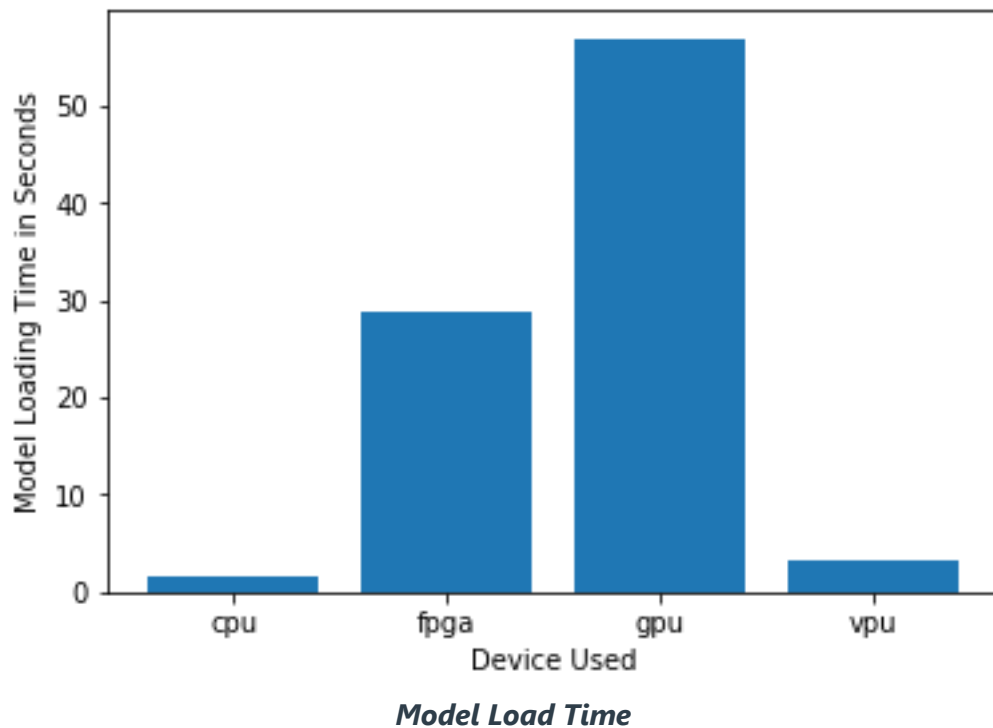
	dedicated neural net compute at full blast. This can meet the requirement.
--	--

## Queue Monitoring Requirements

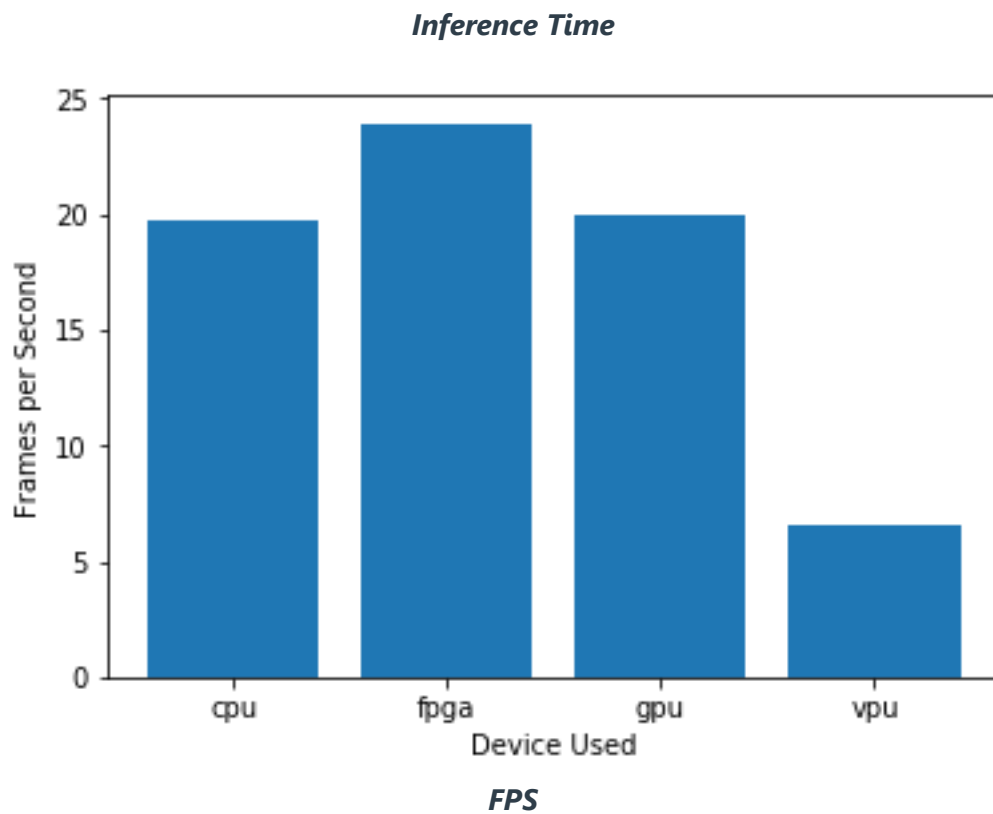
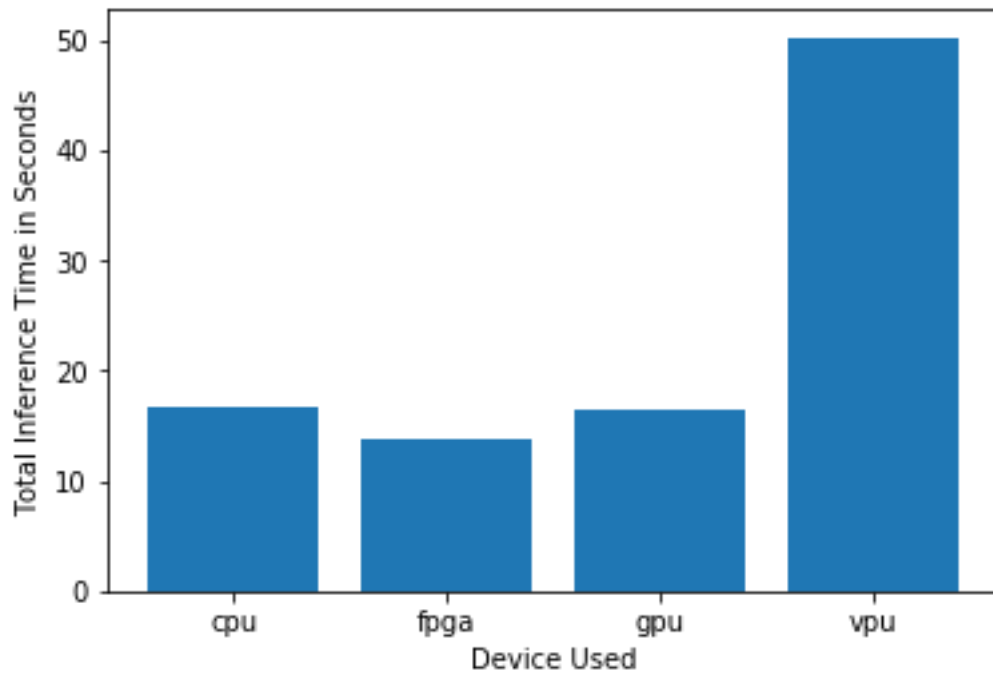
Maximum number of people in the queue	15. In peak hours they currently have over 15 people on average in a single queue
Model precision chosen (FP32, FP16, or Int8)	FP16

## 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).







## 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

Final hardware requirement is VPU ([Intel Neural Compute Stick 2](#)). This is suitable since client has a limited budget (\$300/machine) and wants to save on hardware and power requirement. NS2 is a low power device and costs between \$70 - \$100. Test result shows that FPS and inference time are enough to meet the requirement.