

# Choose the Right Hardware

## Proposal Template

Student: Rubenia Borge

### 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 requires a solution that will last over time, hardware that could last from 5 to 10 years.</i>	<i>FPGAs can last up to 10 years, so they are adequate for long time investments. It has the longest time span compared with CPUs, IGPU and VPUs. Though they are much more expensive.</i>
They system requirements are changing constantly and there are different processes that need to be supported. There are multiple chip designs and new designs are created regularly. The new system requires the possibility of being reprogrammed.	<i>FPGAs are designed with maximum flexibility so that they can be reprogrammed as needed in the field.</i>
The company had an increase in revenue and expects to keep growing both in production capacity and in profits. This means that right now they can afford expensive hardware.	<i>FPGAs can be very expensive in the cases that customers want to invest heavily in hardware. An extreme could be \$100,000. So, the company can afford the investment. The only concern I have is that FPGAs are for prototyping and low volume production. If the company</i>

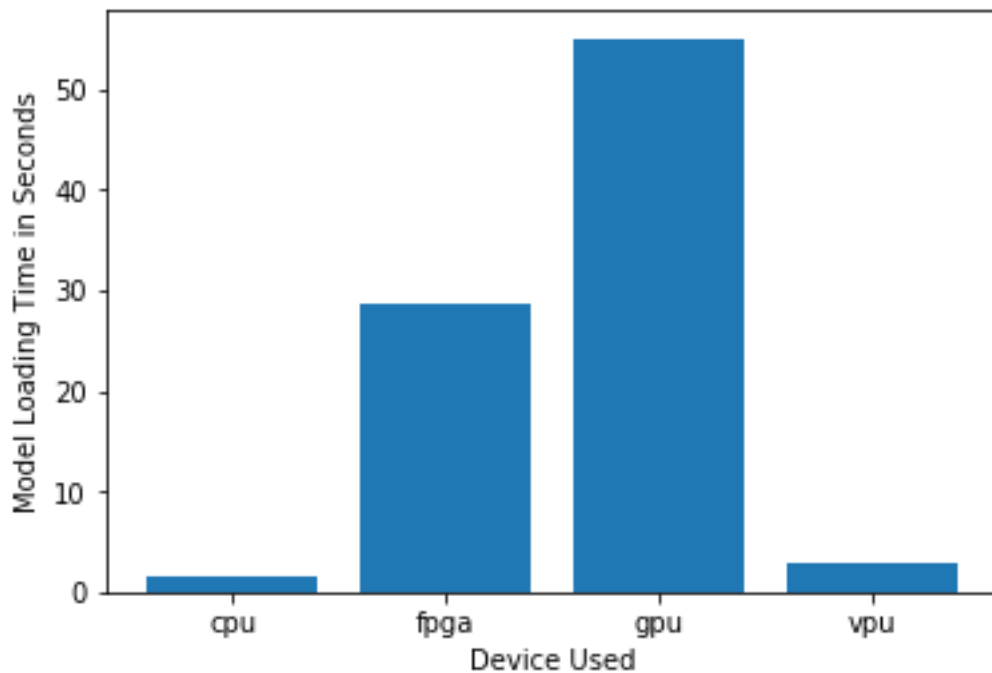
wants to increase production capacity then the highest performance can be achieved with ASICs but requirements in the company are changing and ASICs are used when requirements won't change over time. The details that catches my attention is that the customer requires fast loading time and the fastest loading time is given by the GPU not the FPGA however, my understanding is FPGA shows higher FPS so gives higher performance so I believe that FPGA has more pros than cons in this case.

## Queue Monitoring Requirements

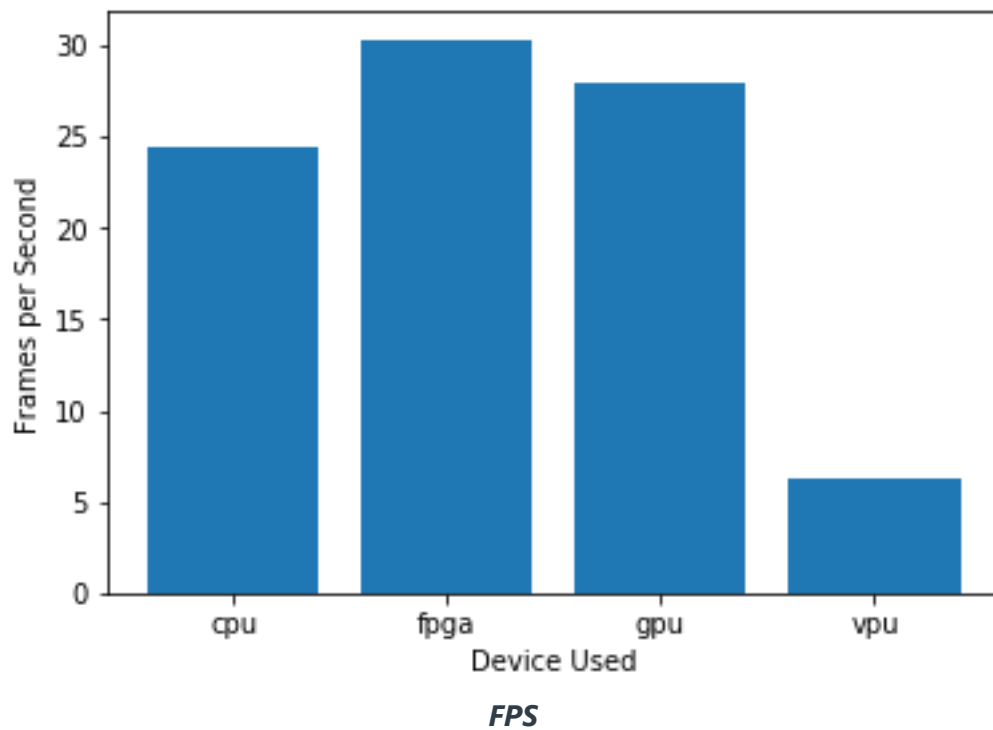
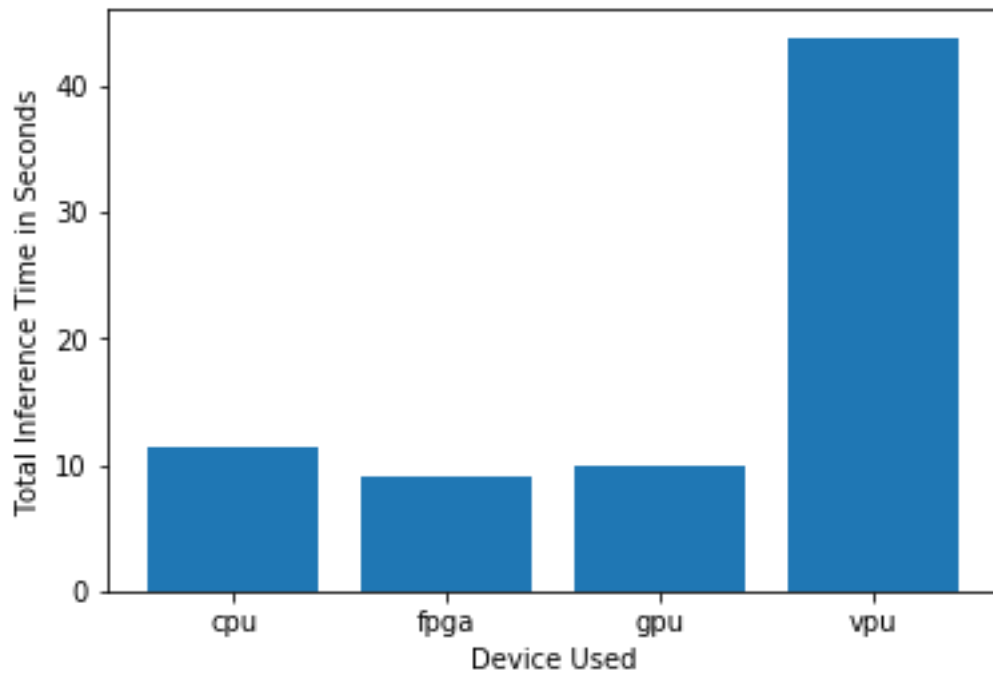
Maximum number of people in the queue	5
Model precision chosen (FP32, FP16, or Int8)	FPGAs run on a 16-bit precision so it would be FP16.

## 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**



## 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 FPGA has the lowest inference time and the customer wants the lowest possible inference time. It also has the highest FPS which I interpret as the highest performance and the loading time is not so bad. So, I still believe that the FPGA is the best option for this client. The cost of FPGAs is high but at this moment the customer can afford it. My other concern was that the FPGAs are suitable for prototyping and low volume production and I understand that the client wants to increase production not reduction, but the other alternative is ASICs and that can be too expensive and it's rigid. ASICs will not provide the flexibility that the client requires. Overall, FPGAs are high performance-low latency devices and have the higher life span which is a requirement for the customer so I still believe the most suitable edge device is 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)

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>Mr. Lin does not have much money to invest in new hardware.</i>	<i>VPUs come in the form of a USB stick. The computers have already i7 processors which makes it feasible to just adapt a VPU like a NCS2. A VPU does not require an upgrade in hardware and it will just accelerate the performance of the pre-existing i7 processors.</i>

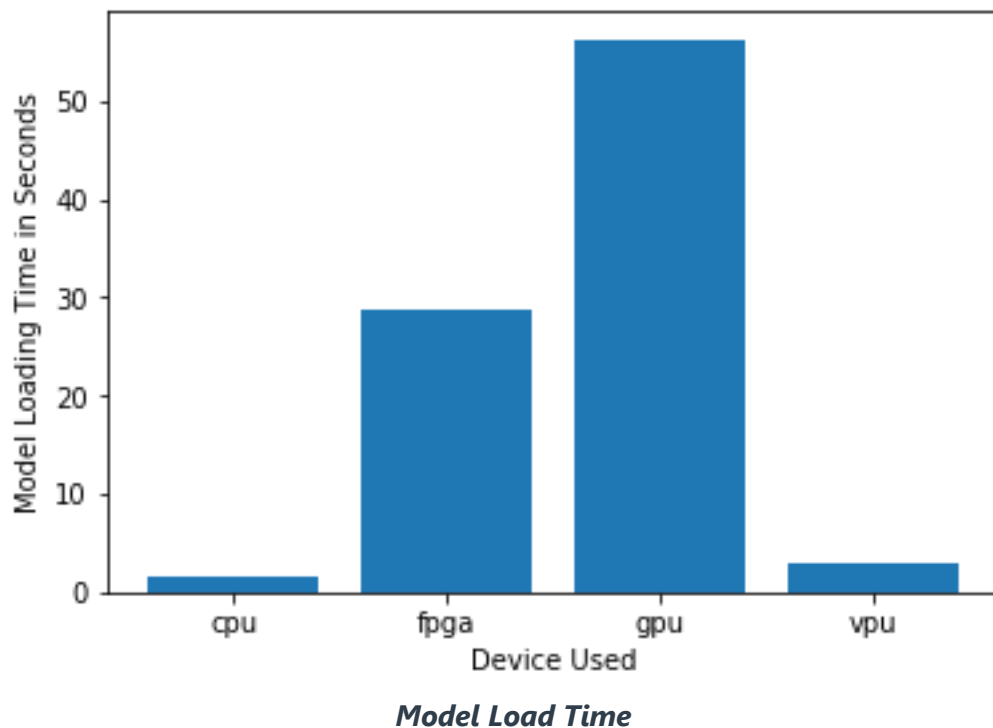
<i>Mr. Lin wants to save as much as possible in his electricity bill.</i>	<i>VPUs are small, low-cost, cost-efficient and low-power devices.</i>
<i>Mr. Lin already has modern computers with i7 processors.</i>	<i>A VPU is cheap, can be just added and he can add more than one if needed.</i>

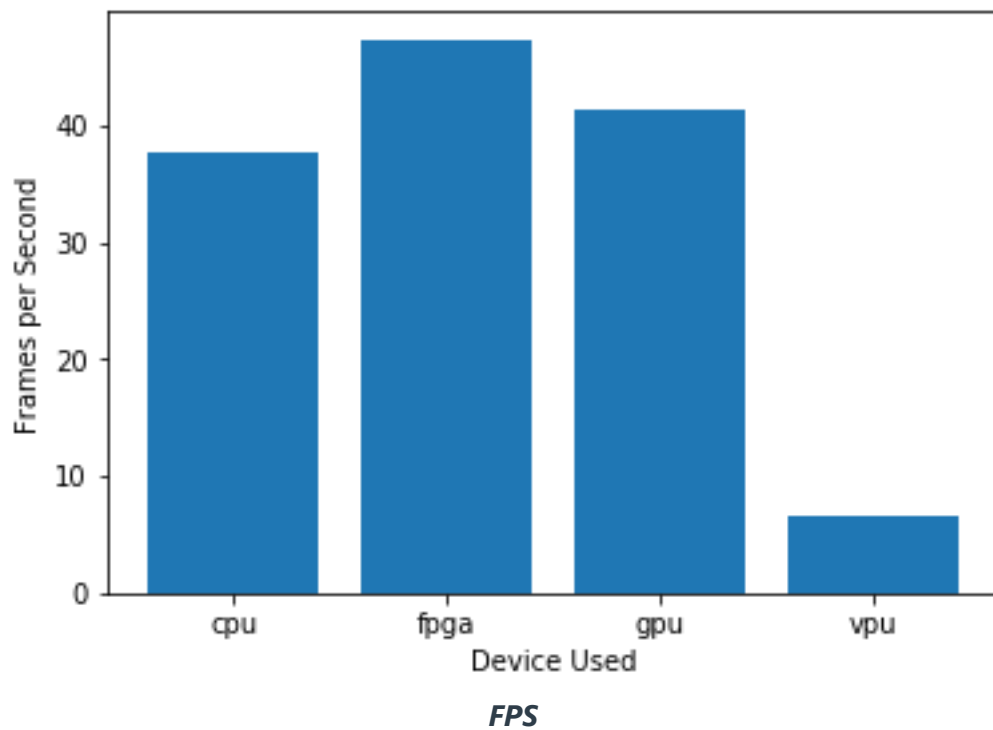
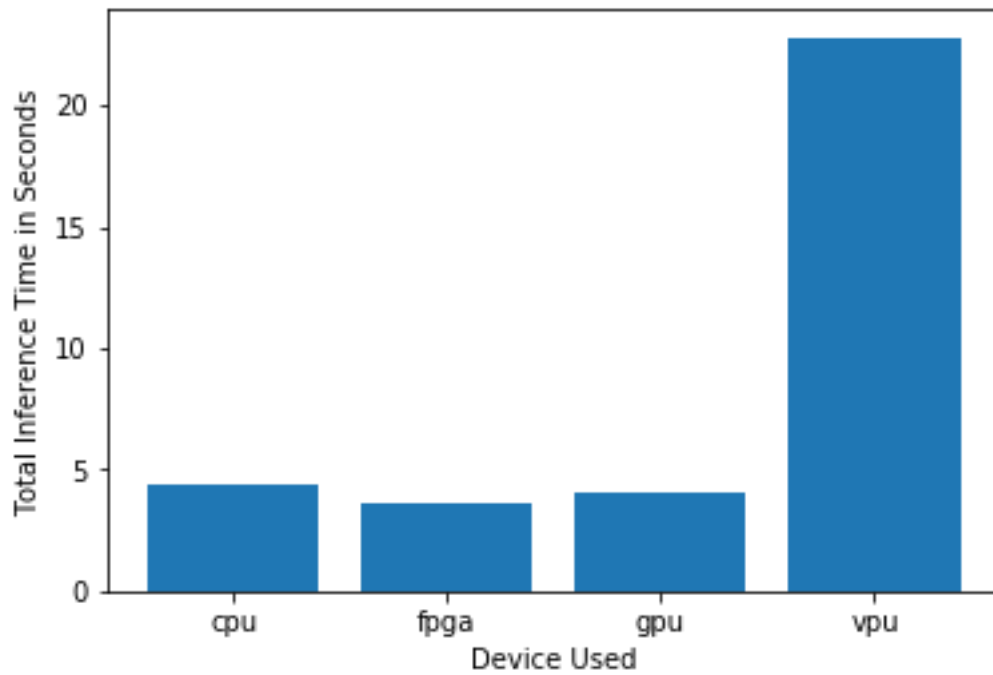
## Queue Monitoring Requirements

<b>Maximum number of people in the queue</b>	5
<b>Model precision chosen (FP32, FP16, or Int8)</b>	VPUs run on a 16-bit precision so it would be 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

*When I see the charts, I believe that the best option for Mr. Lin are either FPGA or GPU because those have the highest performance in the FPS chart. However, he is very limited in budget so definitely cross out FPGA. The GPU has a significantly smaller inference time than the VPU and although the loading time is higher for the GPU than for the VPU I understand that the business context of Mr. Lin has room for flexibility in loading time. However, if I assume that the computers that Mr. Lin has do not have an Integrated GPU (IGPU) that means that he will have to buy an external GPU, and that also will mean that the system will need more power consumption to run and those are realities that Mr. Lin can afford so though I believe the GPU is the most adequate edge device, I believe the client can only afford VPUs and he will have to do the tradeoff of FPS vs. Power consumption.*

## 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 CPUs of the machines connected to the 7 CCTV cameras have no significant additional processing power available.</i>	<i>So, this means to me that Ms. Leah will have to invest in new hardware capabilities since the hardware available right now cannot handle an extra workload that could even cause a collapse.</i>

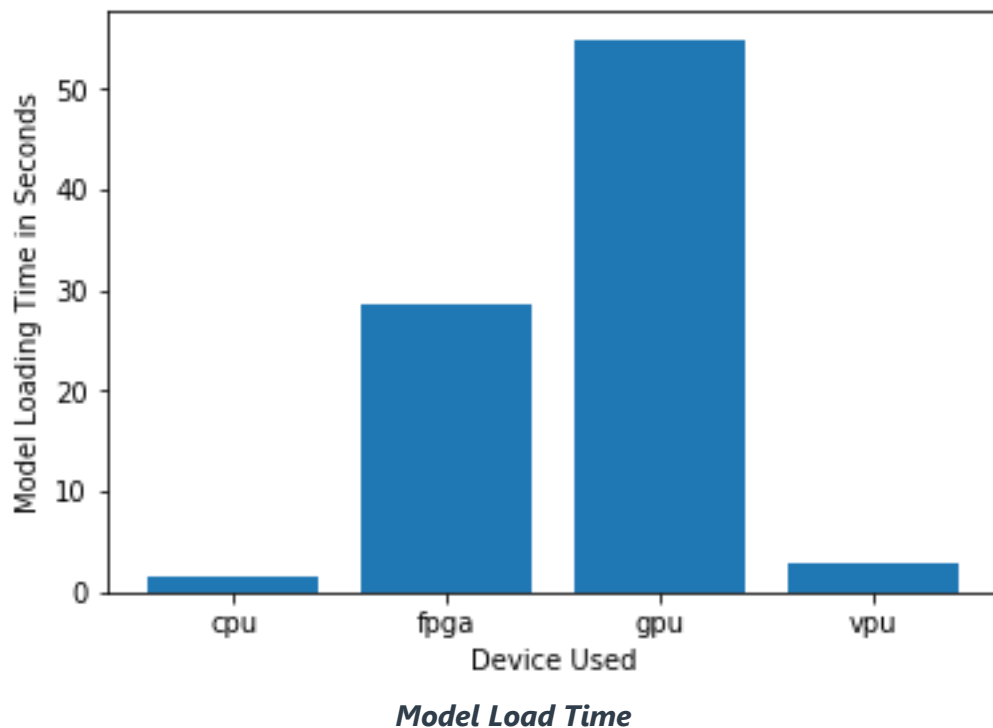
<i>Ms. Leah's budget allows a maximum of U\$300 per machine.</i>	<i>I believe that with that budget Ms. Leah's can only afford VPUS. GPUs are more than U\$300 usually.</i>
<i>Ms. Leah wants to save as much as possible on hardware and future power requirements.</i>	<i>VPUs are low power devices and do not require a hardware update they can be added on as USB sticks and she can add more than one on a computer if that is needed.</i>

## Queue Monitoring Requirements

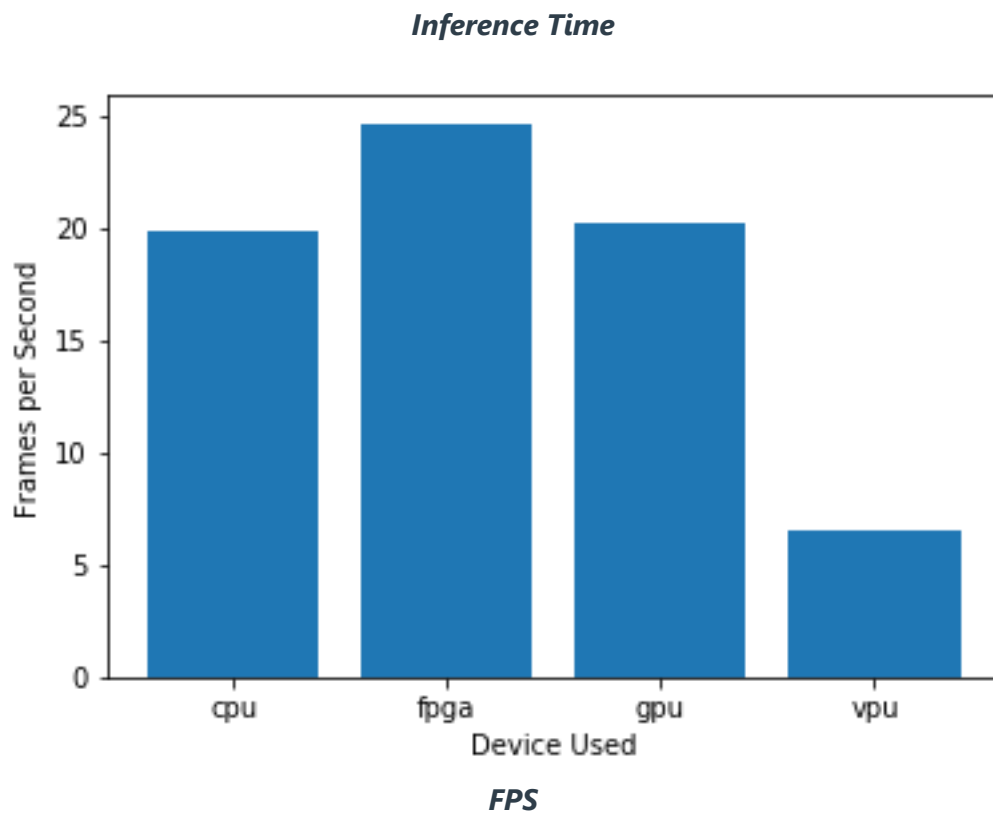
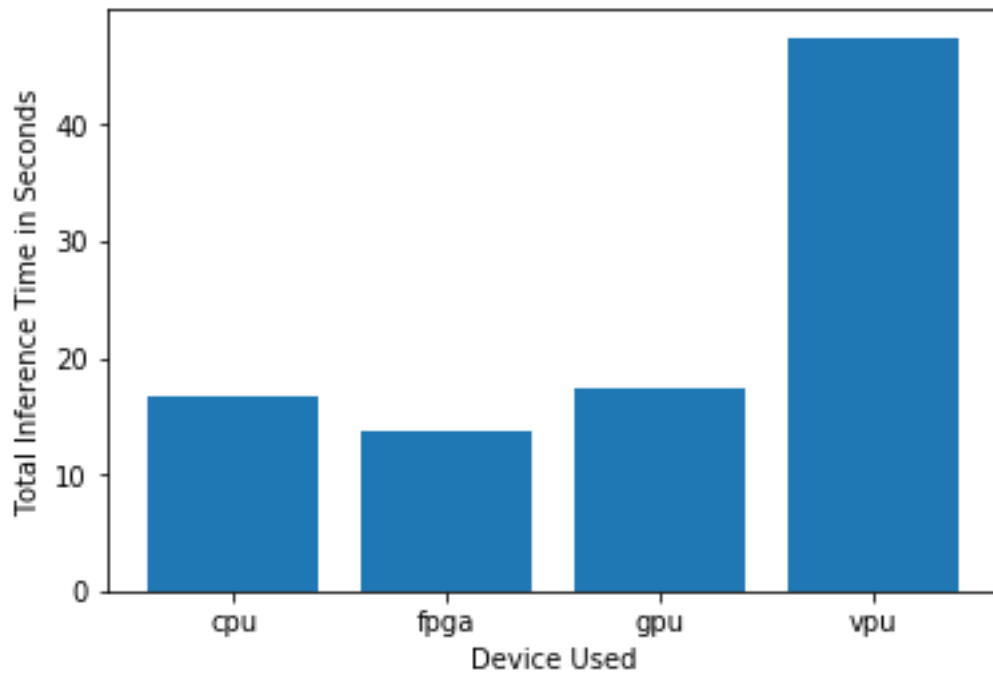
<b>Maximum number of people in the queue</b>	6
<b>Model precision chosen (FP32, FP16, or Int8)</b>	GPUs run on a 16-bit precision so it would be FP16.

## 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

*After seeing the hardware experiments results, my recommendation is the GPU. I believe the GPU is the edge device that will perform better. It gives a performance of about 20 FPS while the VPU gives just a little bit over 5 FPS that's a big difference in performance. Though the VPU has a smaller loading time than the GPU, its inference time is almost double, so I believe the tradeoff of loading time vs performance and inference time makes the GPU a more efficient option than the VPU. However, a GPU can be around \$700 dollars whereas a VPU can be about U\$100 and Ms. Leah has a small budget so most likely she can only afford the VPU, the other aspect is that the GPU might require additional power consumption to run and Ms. Leah wants to save in power costs so again the VPU beats the GPU in power consumption. Summary, I find the GPU more efficient but given the reality of the costumer the edge device she can afford is the VPU and she will have to trade performance over budget constraints.*