

Choose the Right Hardware

Proposed by Muhammad Haris

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)
<i>FPGA</i>

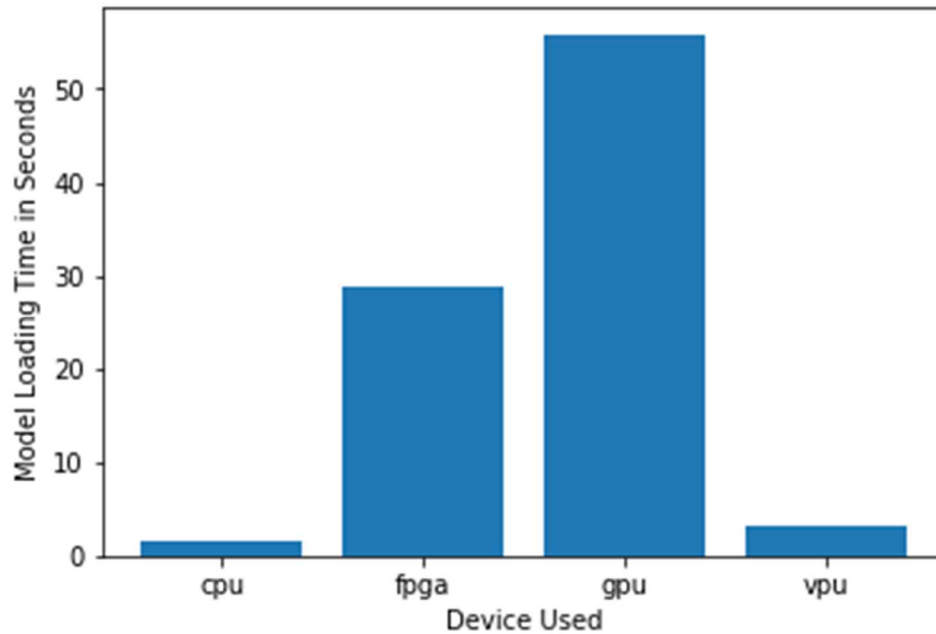
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device.	<i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
<i>To detect chip flaws without slowing down the process quick inference is required</i>	<i>FPGAs have low inference time and offers high throughput</i>
<i>Because of multiple chip designs a flexible system is required</i>	<i>FPGAs are reprogrammable devices and can easily be reprogrammed for different chip designs</i>
<i>System should last for at least 5-10 years and cost is not an issue</i>	<i>FPGAs are expensive but have longer lifespan compared to other options</i>

Queue Monitoring Requirements

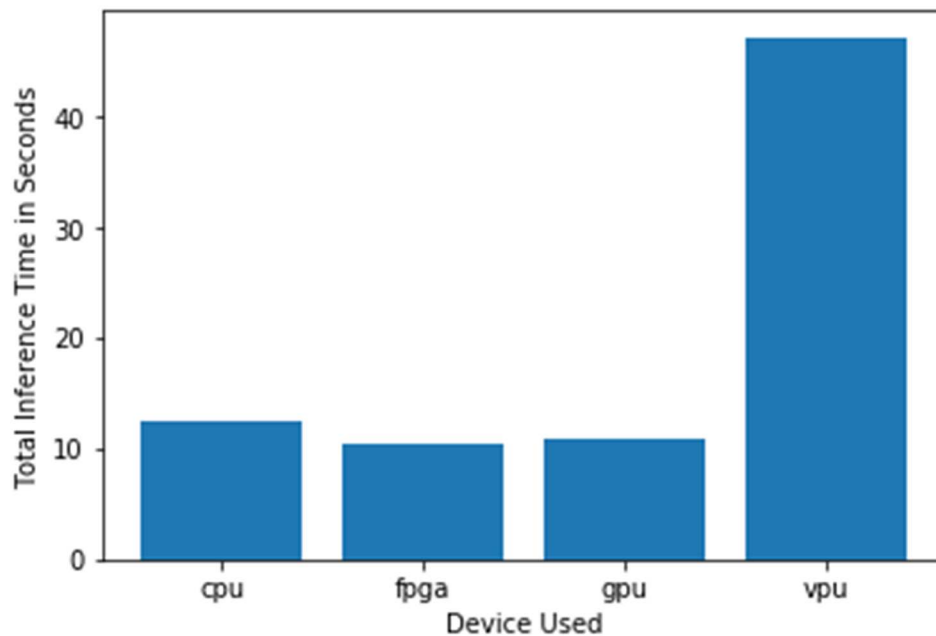
Maximum number of people in the queue	<i>5</i>
Model precision chosen (FP32, FP16, or Int8)	<i>FP16</i>

Test Results

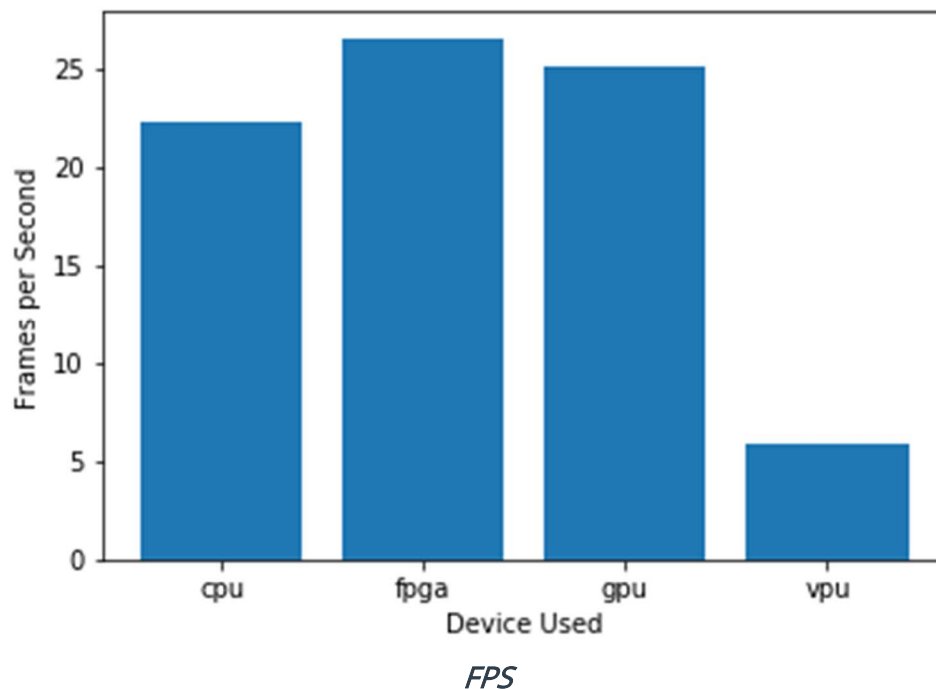
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

Based on the results the final recommended device is FPGA. The graph results shows the low inference time and high throughput of higher than 25 FPS which meets the user demand. Also the FPGAs are reprogrammable devices and can easily be reprogrammed for different chip design scenarios which fulfills the user flexibility demand. Finally FPGAs have longer life span and designed to run 24 hours.

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)
<i>CPU</i>

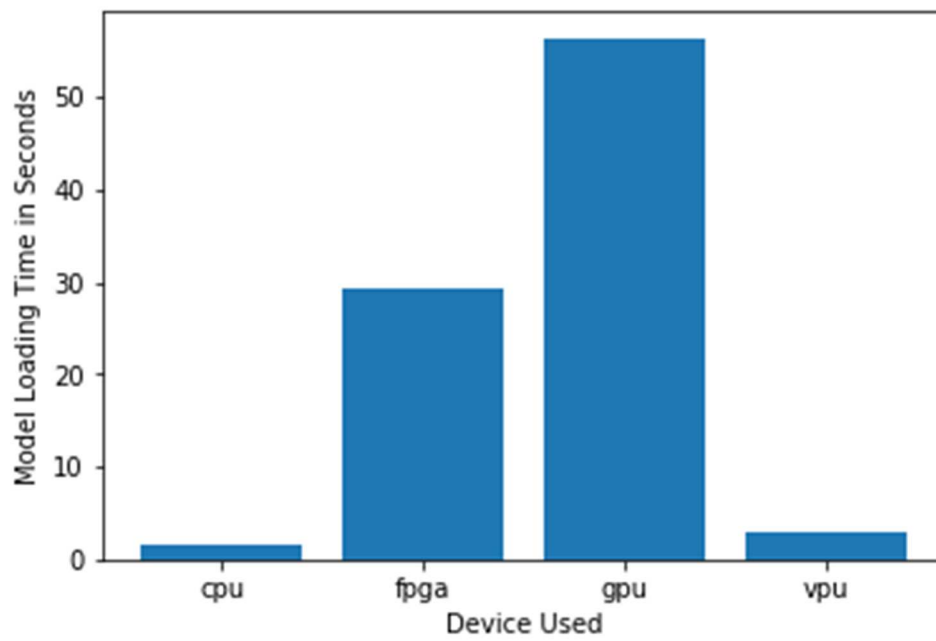
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device.	<i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
<i>In Normal scenario average wait time of a person in que is 230 seconds. In rush hours wait time increases to 350-400 seconds</i>	<i>Inference time is not an issue here. Also no need for additional hardware like FPGA or VPU to increase inference. CPU will work fine.</i>
<i>Client has not much money to invest in additional hardware</i>	<i>Client already has modern intel i7 core processor systems used for some minimal tasks. That's why we can utilize the existing CPU resources.</i>
<i>Like to save as much as on electric bill</i>	<i>CPUs are designed for less power consumption</i>

Queue Monitoring Requirements

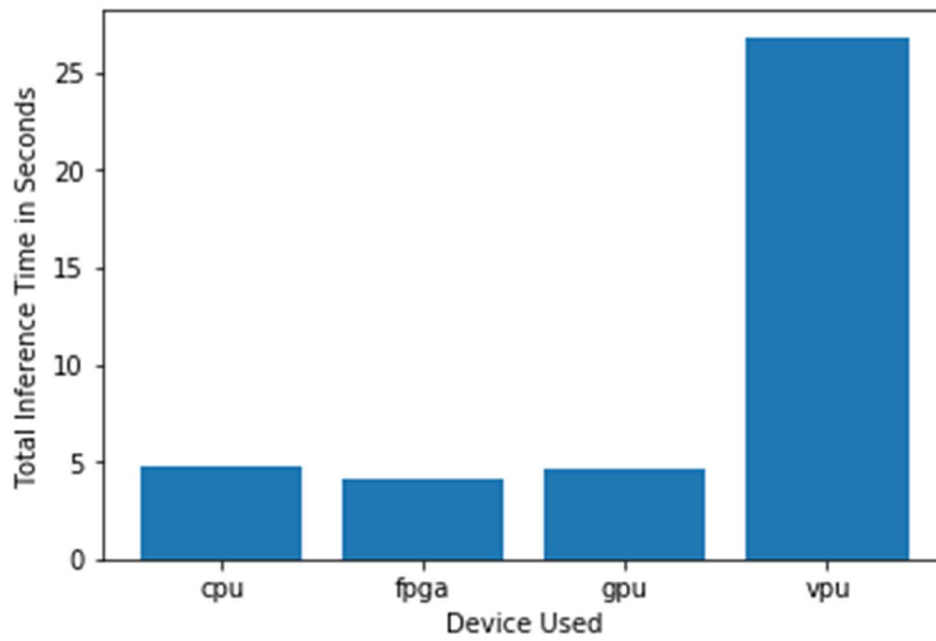
Maximum number of people in the queue	5
Model precision chosen (FP32, FP16, or Int8)	FP32

Test Results

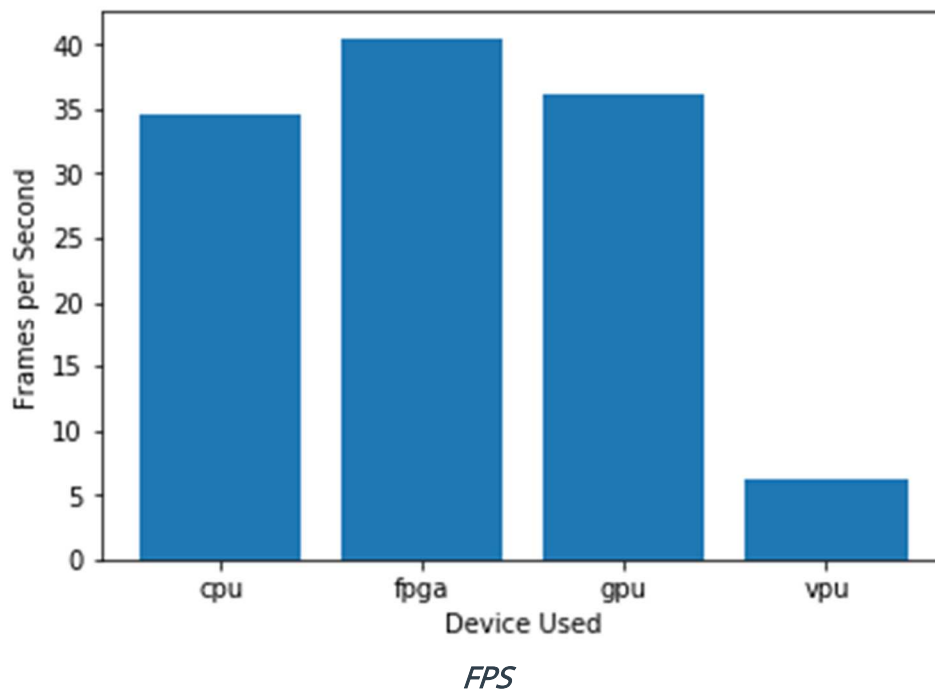
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

Based on results CPU is final hardware recommendation. Client does not want to invest money on additional hardware and they already have modern intel i7 core processor systems which doesn't currently run any computationally expensive tasks. Also CPU inference time is good to handle average 230 seconds average queue wait time. Finally CPUs designed for less power consumption which meets the user electricity consumption demand.

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

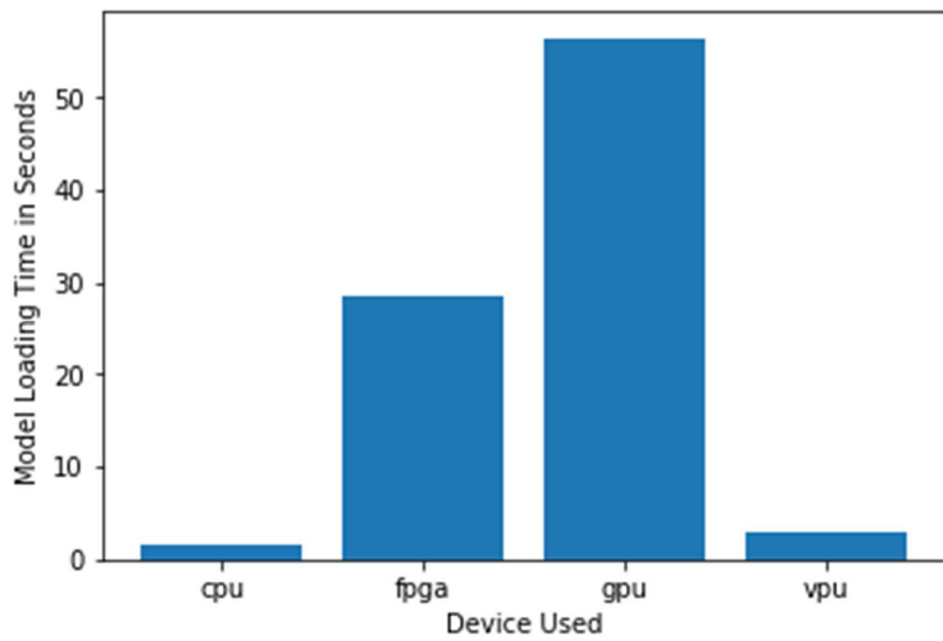
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device.	<i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
<i>No additional processing power is available to run inference</i>	<i>CPUs are currently being used to process and view CCTV footage. VPU will provide additional processing power and multiple VPUs can be used to accelerate the performance of preexisting system.</i>
<i>Client has maximum budget of \$300 per machine</i>	<i>NCS2 with Myriad X processor costs around \$100. Which falls under customer budget</i>
<i>Save as much as possible both on hardware and power requirements</i>	<i>NCS 2 is a USB device which runs on low power.</i>

Queue Monitoring Requirements

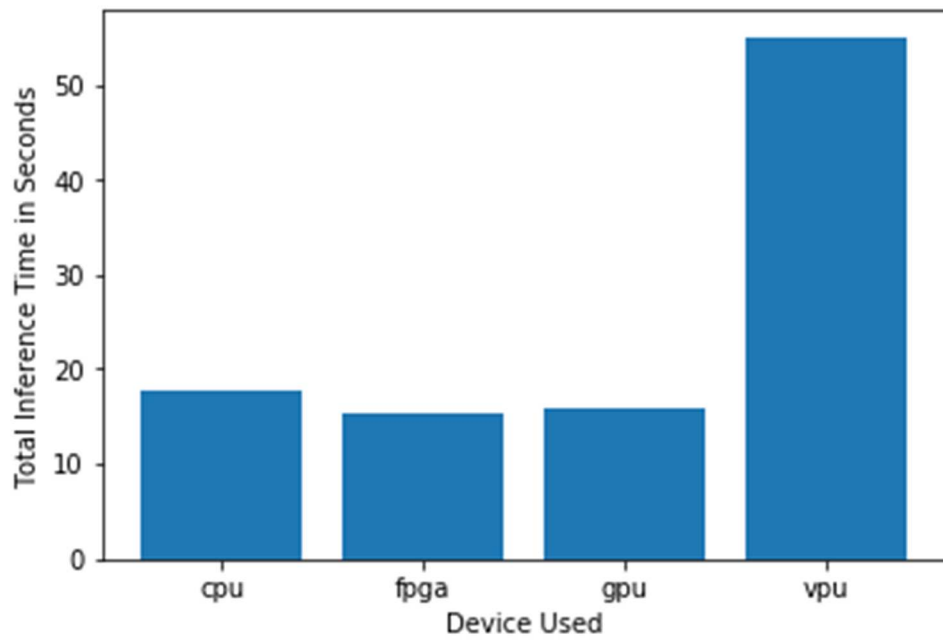
Maximum number of people in the queue	10
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

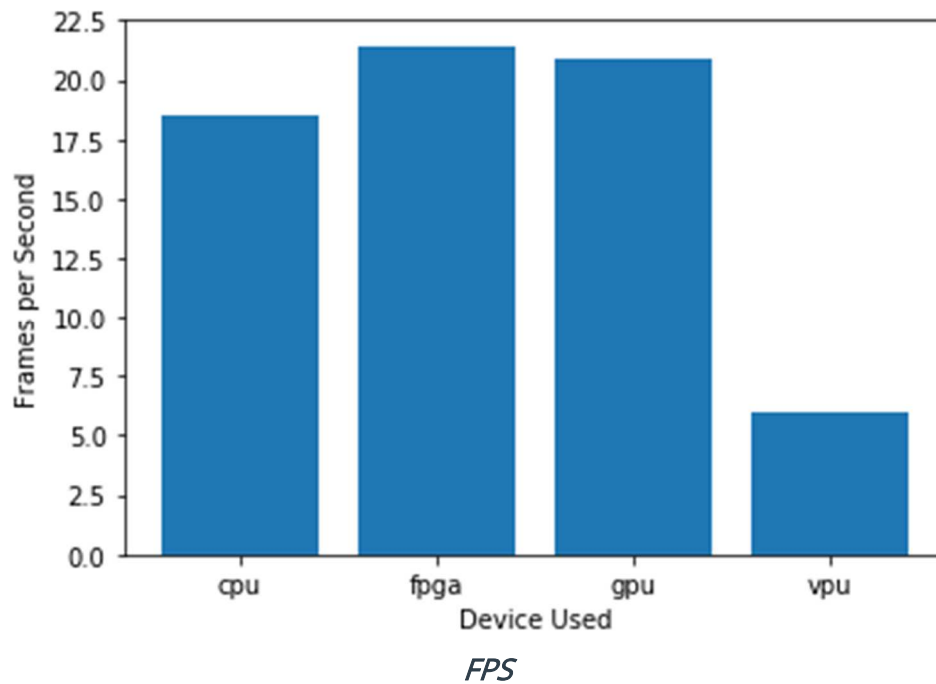
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



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

Based on results final recommended device is VPU. CPUs are currently being used to process CCTV footage. In this case VPUs will provide additional processing power and multiple VPUs up to 3 under user budget should be used to accelerate the performance and provide the required inference time. Also NCS 2 VPU is a USB device which is designed to run on low power and costs around \$100. Which meets the user budget and power requirements.