

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

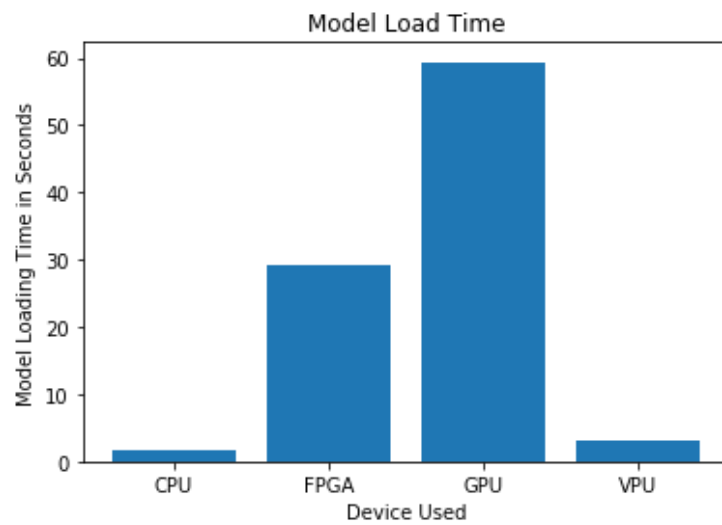
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>The company is expanding and the previous year they made 2 Million dollars in revenue.</i>	<i>The proposed solution fits the clients budget.</i>
<i>The client wants a solution that would be reprogrammable.</i>	<i>The proposed solution is reprogrammable and can be repurposed for both use cases that the client mentioned.</i>
<i>The client wants a solution that will at least last them 5-10 years.</i>	<i>FPGAs have a longer lifespan compared to other alternatives.</i>

Queue Monitoring Requirements

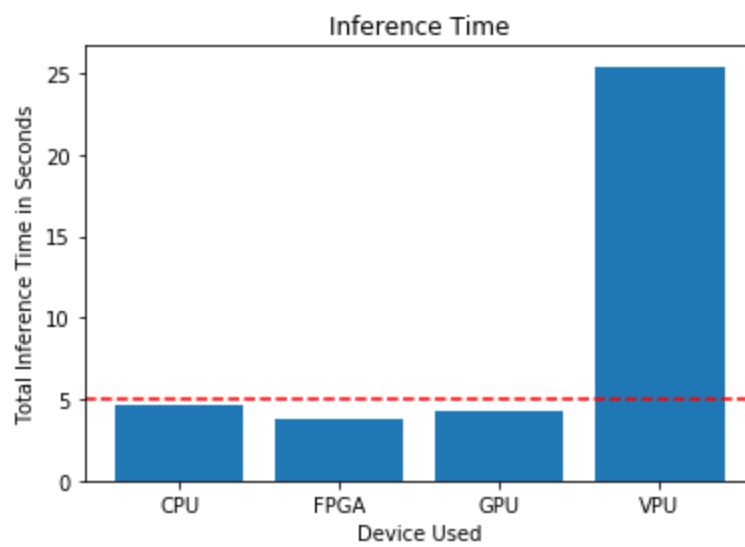
Maximum number of people in the queue	5
Model precision chosen (FP32, FP16, or Int8)	<ul style="list-style-type: none">- CPU: FP32- VPU, GPU and FPGA: FP16

Test Results

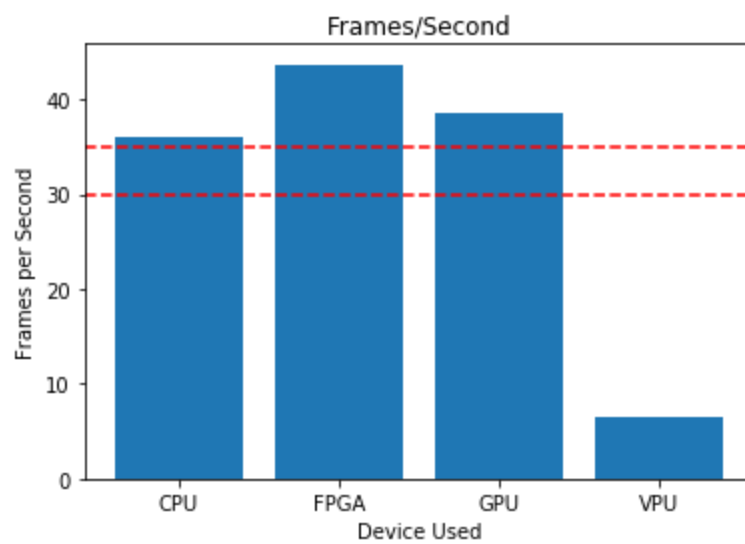
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

*From the results shown above, we can see that the proposed solution (FPGA) as it offers the best performances and it satisfies the FPS requirements of 30-35fps. The client also mentioned that the solution needs to be reprogrammable and lasting between 5-10years.
Thus we can conclude that the client's best choice would be 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)
CPU+IGPU

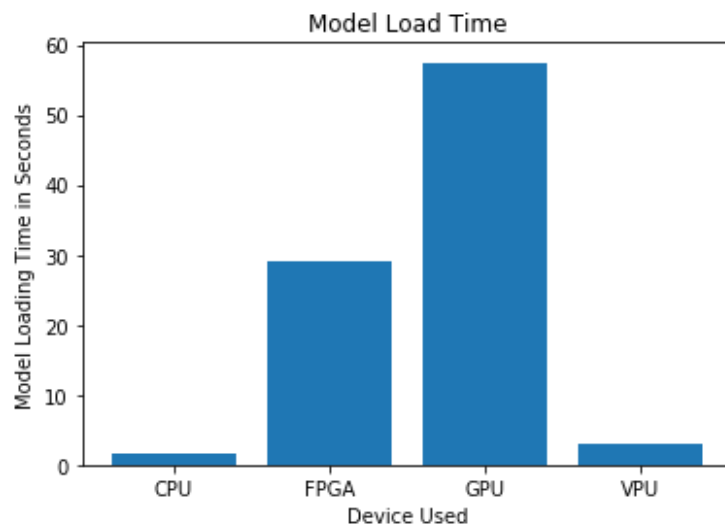
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>The client already has a high-end i7 CPU which is not being fully utilised.</i>	<i>The CPU is not being fully utilised meaning it could be leveraged for inference.</i>
<i>The client does not intend on investing in additional hardware.</i>	<i>Using the i7 CPU for inference means that the client won't need any hardware upgrading related costs.</i>
<i>As per the client, "They have an average wait time of 230 seconds at the checkout counters. But on the weekends, the wait time can increase substantially. The average time spent is 40 mins at the store and 350-400 seconds at the checkout line."</i>	<i>Inference time is not a serious issue as average wait time during normal time is >1 minute.</i>

Queue Monitoring Requirements

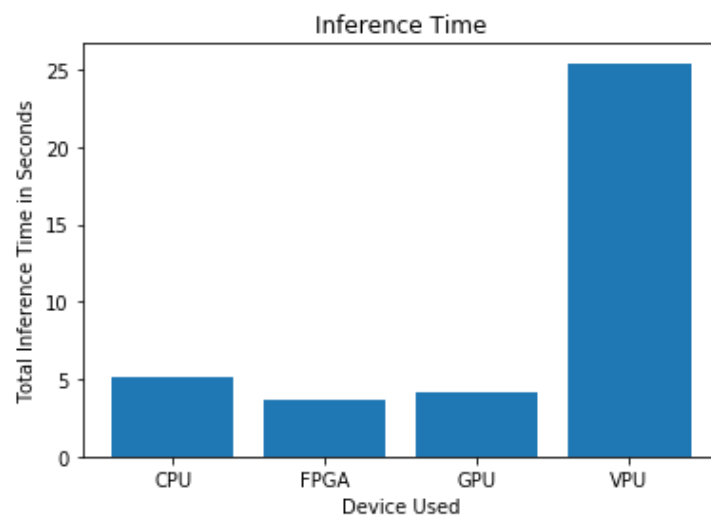
Maximum number of people in the queue	3
Model precision chosen (FP32, FP16, or Int8)	<ul style="list-style-type: none">- CPU: FP32- VPU, GPU, FPGA: FP16

Test Results

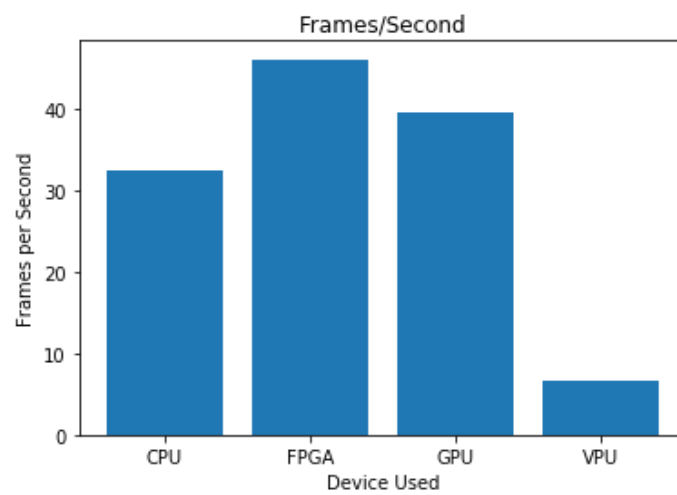
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

The client can leverage using the existing hardware and utilise the IGPU for better performance without purchasing additional hardware. They can expect a better FPS as compared to just utilising the CPU alone and due to its low power consumption, the client will not expect a high electricity bill.

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

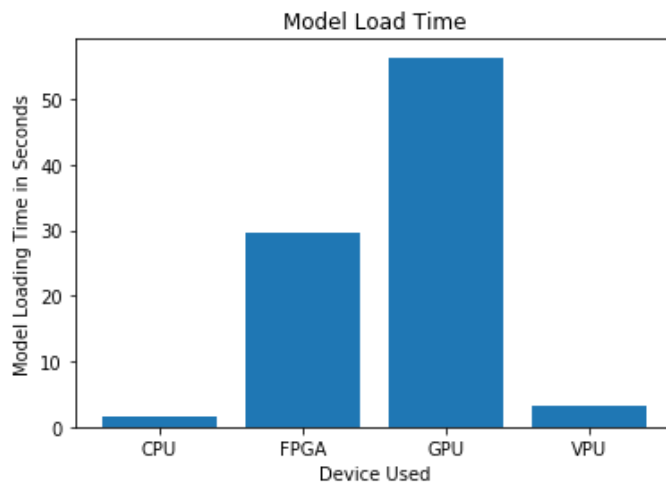
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>The client's budget is \$300/machine</i>	<i>The proposed hardware cost's between \$90-\$120 depending on the retailer.</i>
<i>The client would like to save as much as possible both on hardware and future power requirements.</i>	<i>The proposed hardware consumes less power.</i>
<i>The CPU on the client's machines are currently being used to process CCTV footages.</i>	<i>The proposed hardware can be leveraged for inference without utilising the CPU</i>

Queue Monitoring Requirements

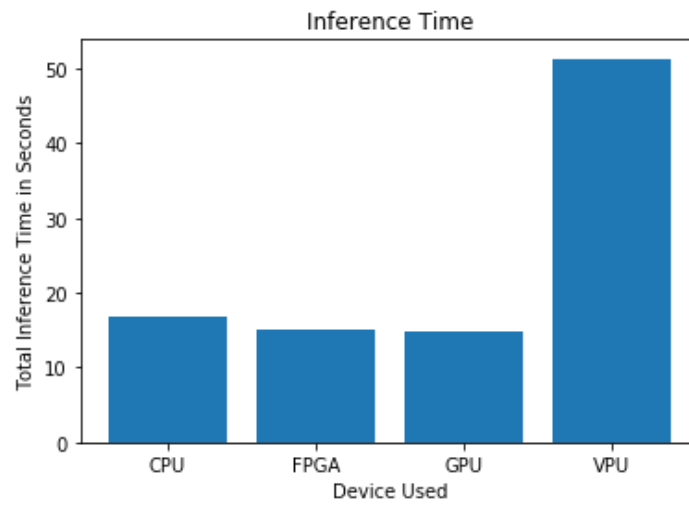
Maximum number of people in the queue	5
Model precision chosen (FP32, FP16, or Int8)	<ul style="list-style-type: none">- CPU: FP32- VPU, GPU, FPGA: FP16

Test Results

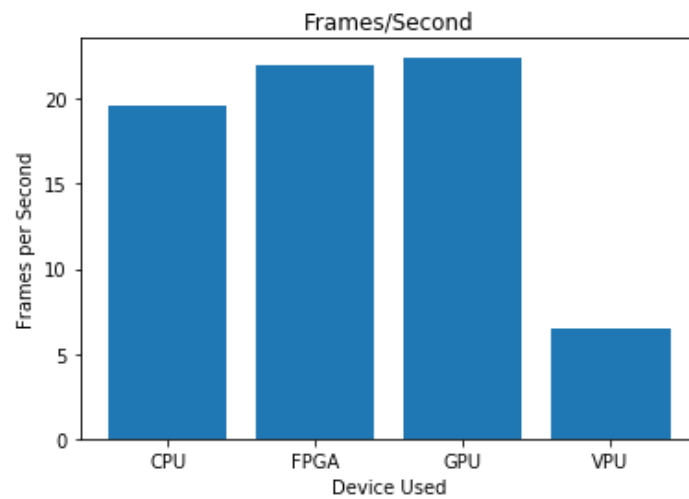
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

Given the client's tight budget and overloaded CPUs, it is advisable for the client to consider the VPU. However, looking at the stats above we can see that the VPU inference time is higher than other alternative hardware and the frames/seconds is the lowest. These are potential drawbacks of using the VPU - should the client increase their budget they could consider upgrading their machines CPUs. At the moment the VPU (eg Neural Compute Stick 2) would be the best option for them.