

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) |
|--|
| <i>FPGA with CPU as a fallback device</i> |

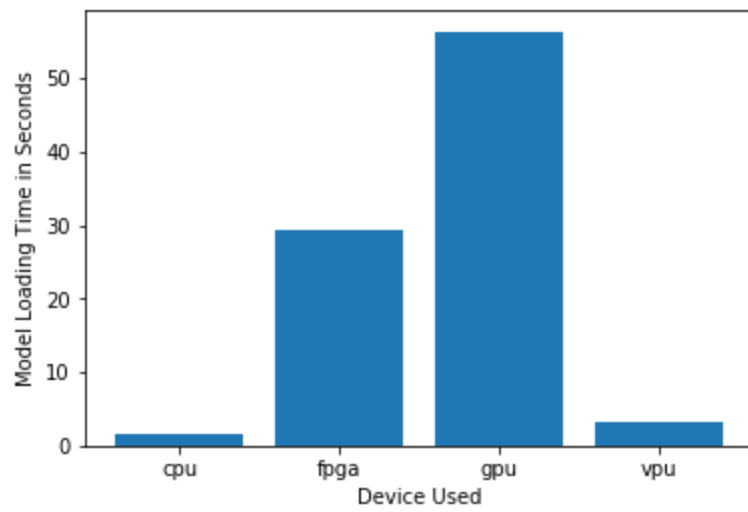
| Requirement Observed (Include at least two.) | How does the chosen hardware meet this requirement? |
|---|---|
| <i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device. | <i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| <i>The client needs to do inference 5 times per second, which mean an inference time below 200ms.</i> | <i>VPU and FPGA have a good low latency and would be fit best for this requirement. CPU and GPU also could fit the requirement.</i> |
| <i>The client needs a system that could be reprogrammed and optimized to quickly detect frames to different chip designs.</i> | <i>FPGA is reprogrammable and could be optimized for certain needs and would definitely fit this constraint.</i> |
| <i>The client would be the system to last for 5 to 10 years.</i> | <i>FPGA last for a long period around 10 years or even more.</i> |

Queue Monitoring Requirements

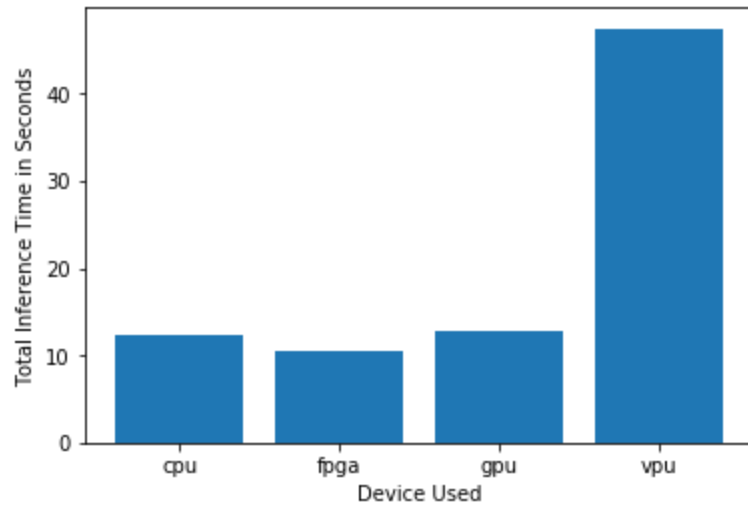
| | |
|--|------|
| Maximum number of people in the queue | 4 |
| Model precision chosen (FP32, FP16, or Int8) | FP16 |

Test Results

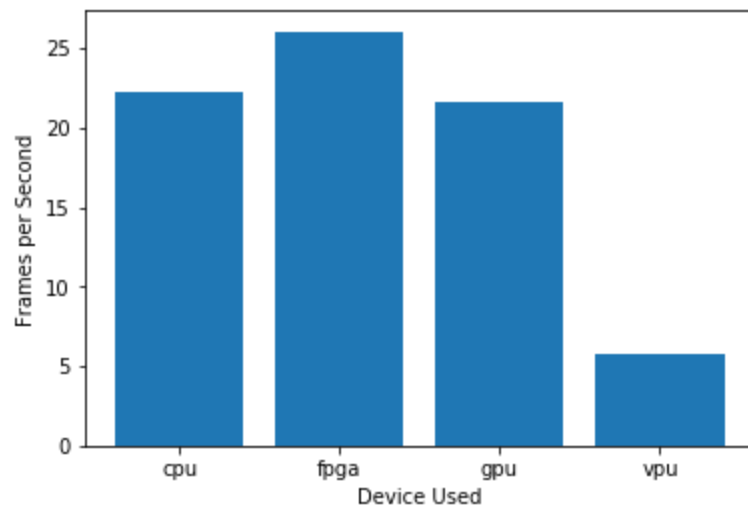
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
FPGA has shown a very good inference time and a very good FPS which is required by the client, furthermore it's the best reprogrammable choice and it lasts for a long period.

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

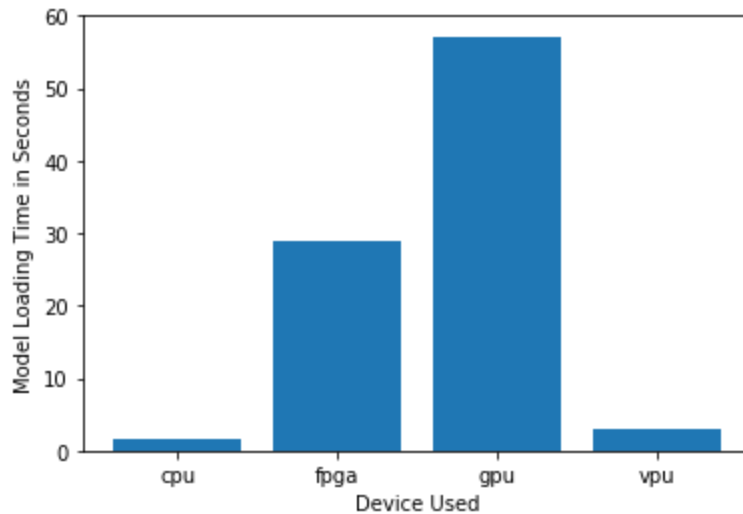
| Requirement Observed (Include at least two.) | How does the chosen hardware meet this requirement? |
|---|--|
| <i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device. | <i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| <i>The client does not have much money to invest in additional hardware</i> | <i>The client already owns non overcrowded modern computers with Intel i7 core processor, not which supports OpenVINO.</i> |
| <i>The client would like to save as much as possible on his electric bill.</i> | The hardware owned by the client Intel i7 using CPU have low TDP |

Queue Monitoring Requirements

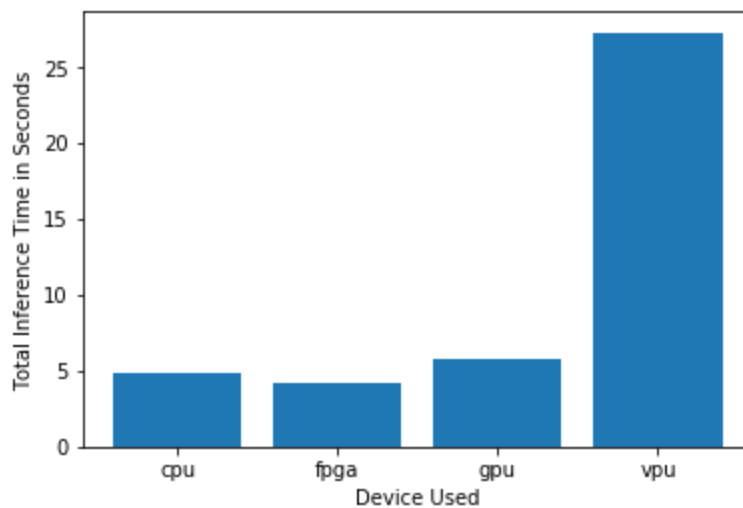
| | |
|---------------------------------------|---|
| Maximum number of people in the queue | 2 |
|---------------------------------------|---|

Test Results

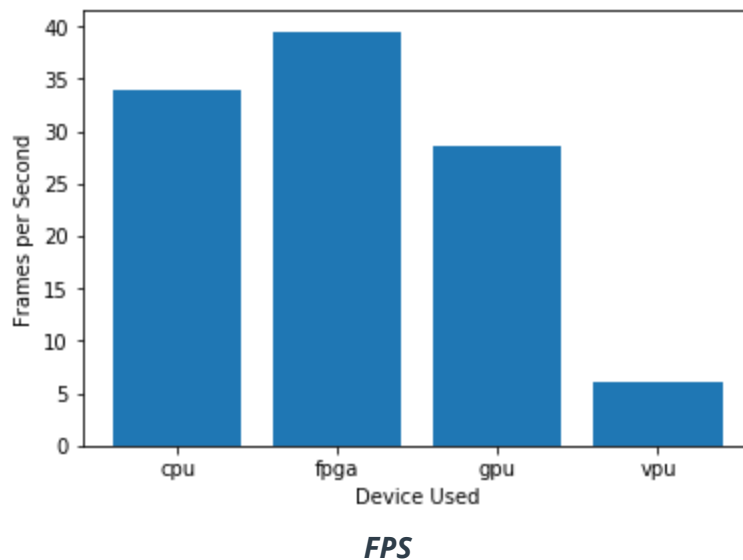
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

CPU
 In term of performance CPU is not the best choice in this case, but according to the situation of the client, who could not afford to buy a new hardware, so the best choice is to use the already existing Intel i7 Core.

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 or NCS2

| Requirement Observed (Include at least two.) | How does the chosen hardware meet this requirement? |
|---|---|
| Example requirement: | Example explanation: |

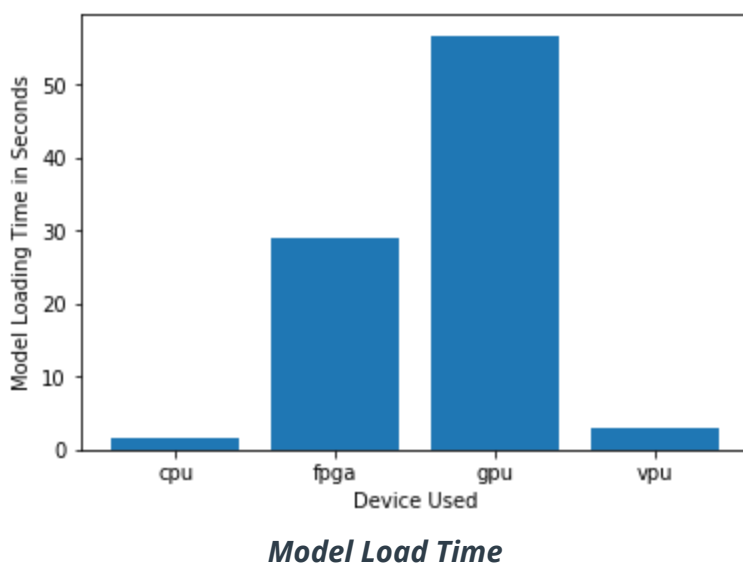
| | |
|--|--|
| The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device. | VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| <i>The client has a small budget of 300\$ per machine.</i> | <i>VPU or NCS2 could cost around 69\$ which is inexpensive and would be a good fit for this requirement.</i> |
| <i>The client would like to save as much as possible power consumption.</i> | <i>NCS2 power consumption is around 1W and will be the best option for saving power.</i> |

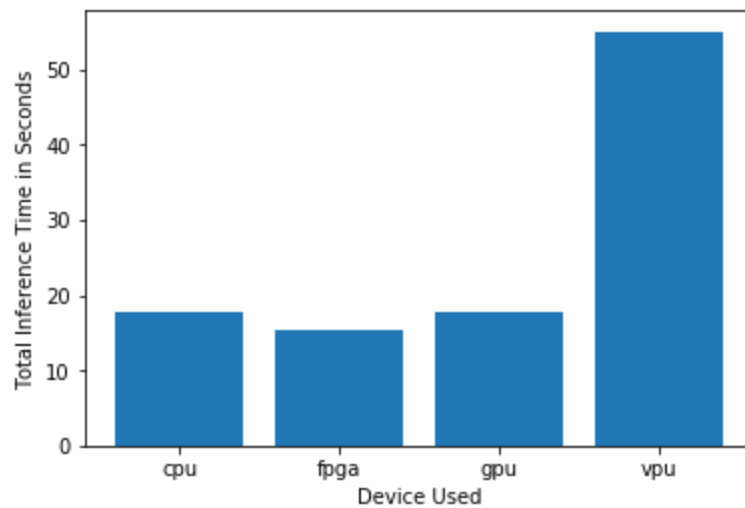
Queue Monitoring Requirements

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

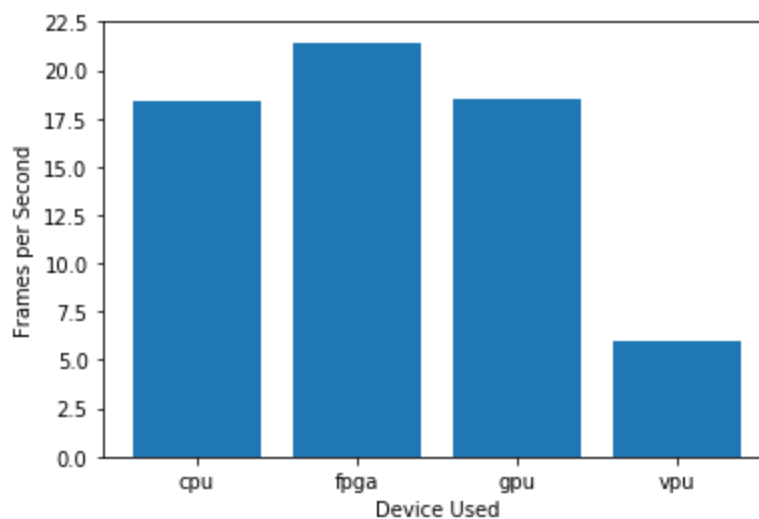
Test Results

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).





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
 VPU was the best in term of time taken for inference and loading the model, but the FPS was low. The client has not expressed any need related to frames per second, so VPU still the best option in this case that could fit the low budget of the client and his willing to save power consumption.