

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) |
|--|
| FPGA |

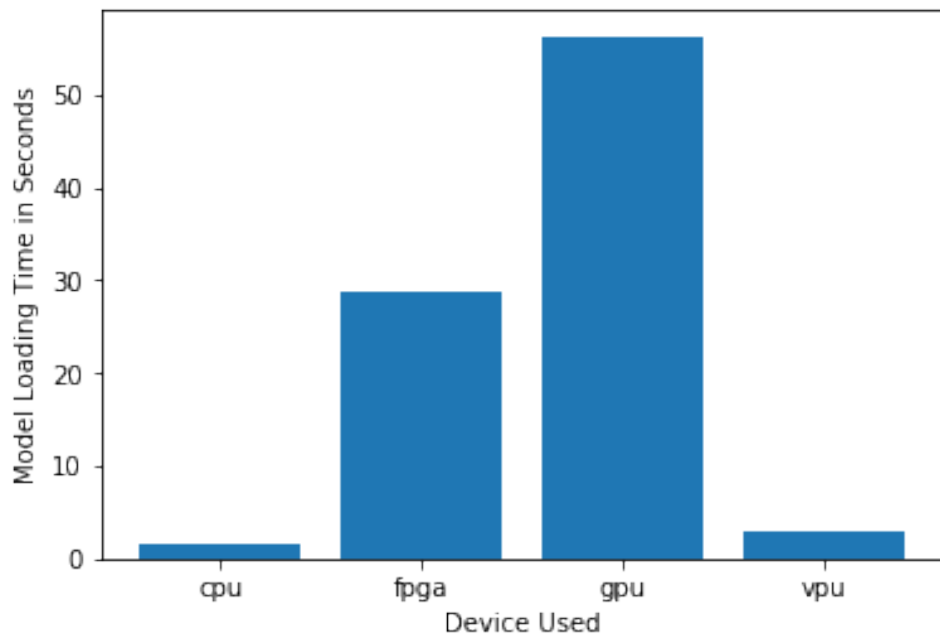
| Requirement Observed (Include at least two.) | How does the chosen hardware meet this requirement? |
|---|---|
| <i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device. | <i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| <i>Must be able to be reprogrammed on the field for a future application.</i> | FPGAs can be reprogrammed as needed in the field after deployment. |
| <i>30-35 FPS and DSP pipeline to execute 5 times per second. Should be able to perform very high speed inference on a custom model</i> | FPGAs are high performance and low latency. |
| <i>System should be reliable for 5 – 10 years</i> | FPGAs are guaranteed to last at least 10 years |

Queue Monitoring Requirements

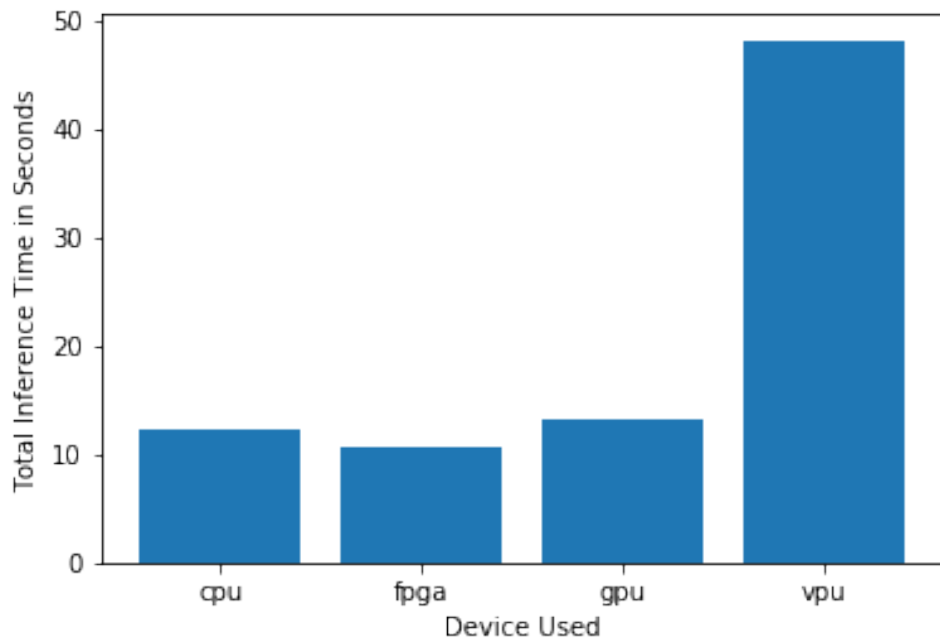
| | |
|--|---------------------|
| Maximum number of people in the queue | 6 per conveyor belt |
| Model precision chosen (FP32, FP16, or Int8) | FP32 or FP16 |

Test Results

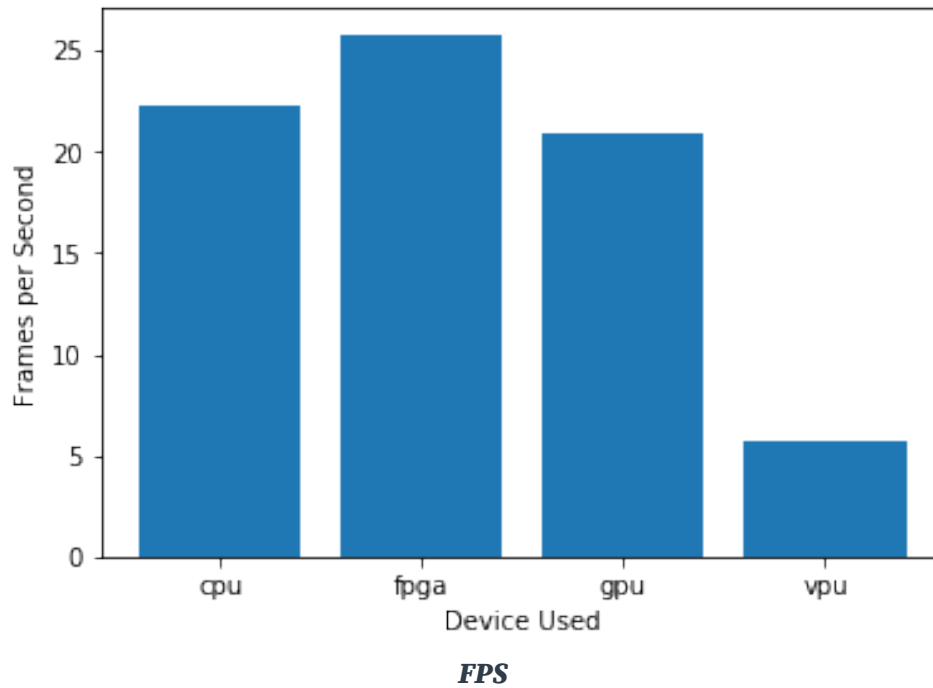
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

FPGA is still the recommended option as it proved it provides the most frames per second, the fastest inference and a reasonable loading time. It can also be reprogrammed for future applications as the customer specified.

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?

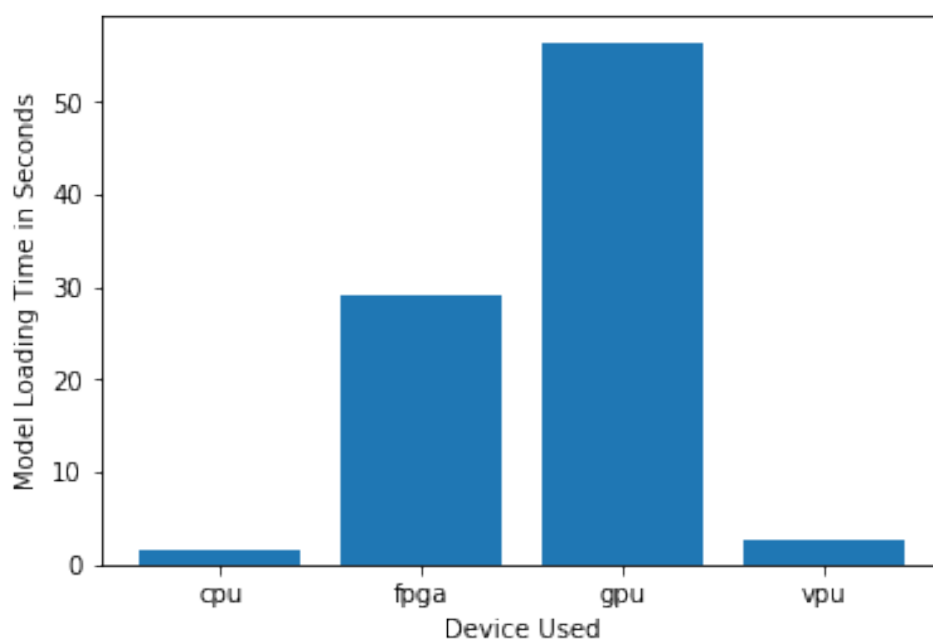
| | |
|---|---|
| <i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device. | <i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| <i>Save as much as possible</i> | <i>Customer already owns several i7 CPUs</i> |
| <i>Must help minimize waiting time</i> | <i>CPUs have reasonable fast inference</i> |
| <i>[TODO: Type your answer here]</i> | <i>[TODO: Type your answer here]</i> |

Queue Monitoring Requirements

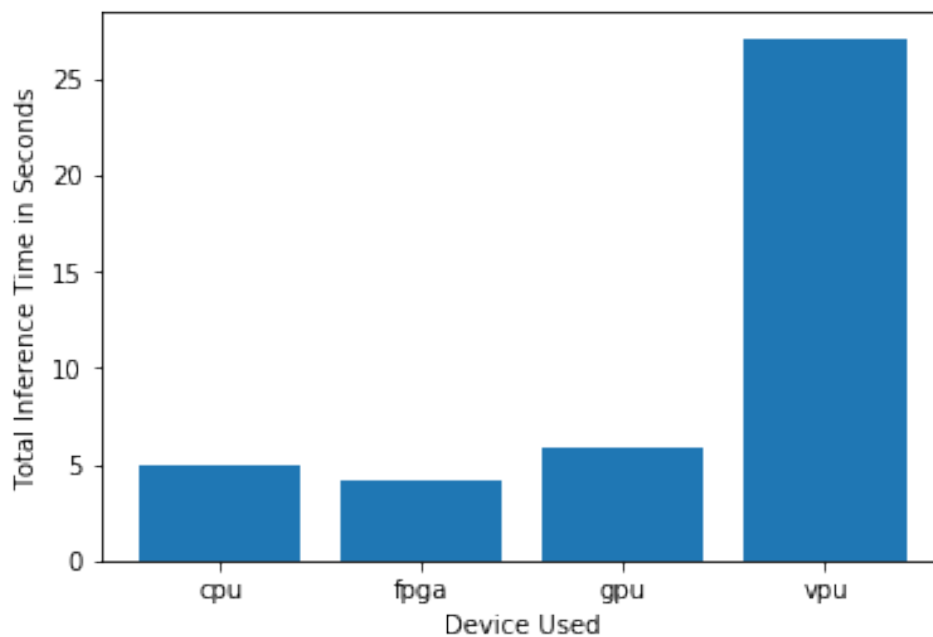
| | |
|---|--------------------|
| Maximum number of people in the queue | <i>2 per queue</i> |
| Model precision chosen (FP32, FP16, or Int8) | <i>FP32</i> |

Test Results

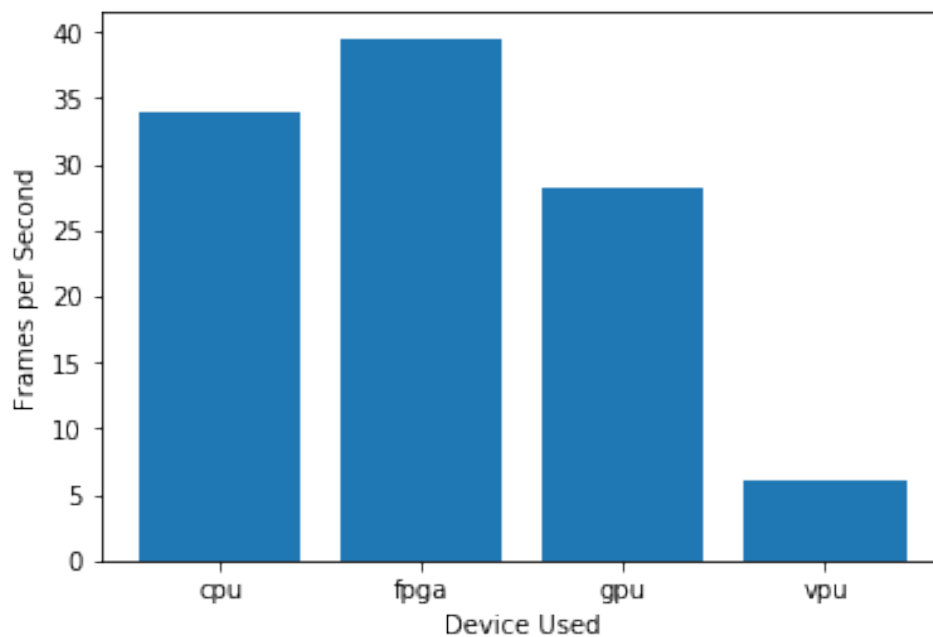
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



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 recommendation is still the CPU since it outperforms every other device except the FPGA and the customer already has i7 CPUs which are more powerful than the i5 used for testing

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 |

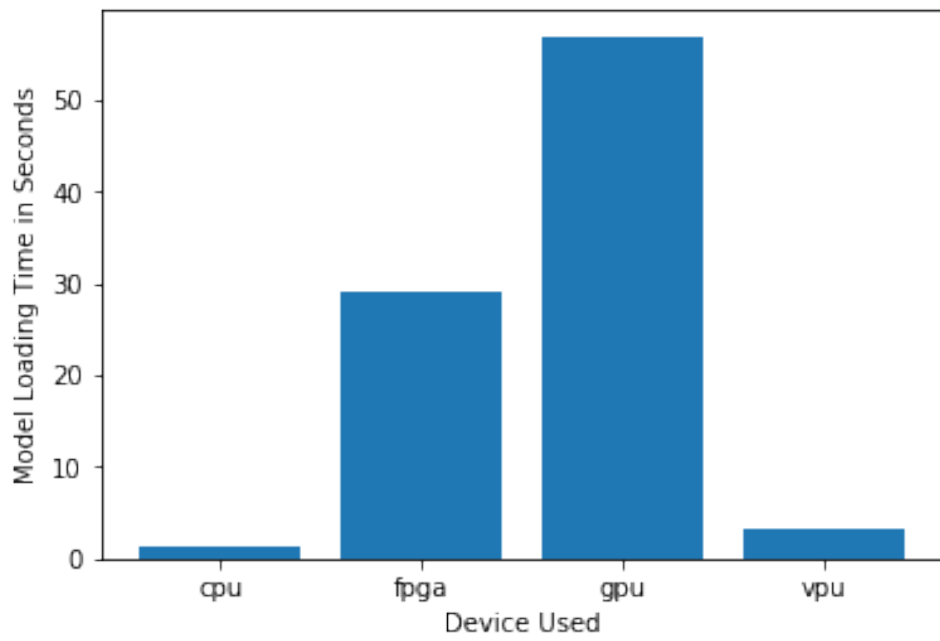
| Requirement Observed (Include at least two.) | How does the chosen hardware meet this requirement? |
|---|---|
| <i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device. | <i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| Max 300 USD per machine | NCS2 is cheap <\$100 |
| Can't upgrade all-in-one's hardware. | VPU is a USB3.1 |
| Inference should be reasonably fast and all processing should be done within 2 minutes | Optimized for image processing |

Queue Monitoring Requirements

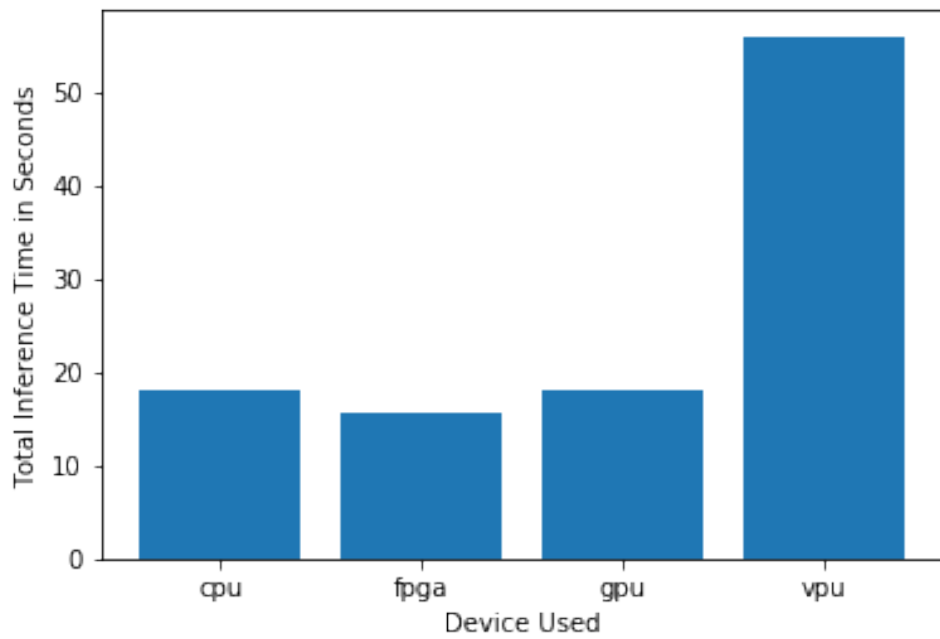
| | |
|--|--|
| Maximum number of people in the queue | 7 in order to keep people flowing through queues |
| Model precision chosen (FP32, FP16, or Int8) | FP16 (required by VPU) |

Test Results

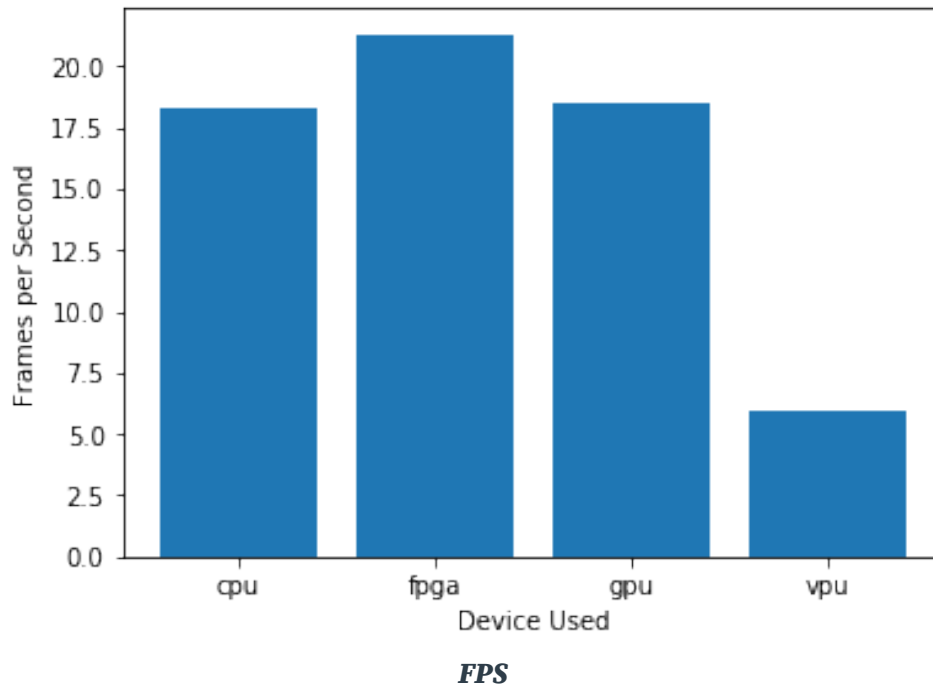
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

The final recommendation is the Neural Compute Stick 2 since it has enough FPS capability to comply with the customer requirements and represents a small investment.