

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 |

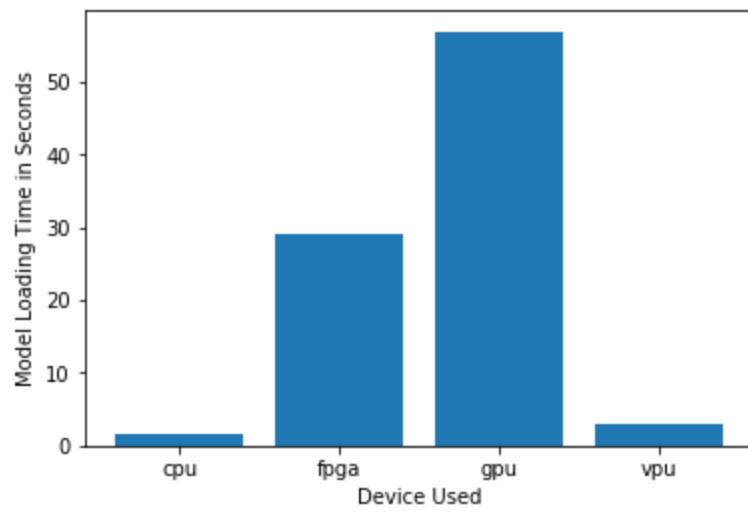
| Requirement Observed (Include at least two.) | How does the chosen hardware meet this requirement? |
|---|---|
| <i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device. | <i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| The client requires the installed system to last for at least 5-10 years. | FPGAs have a long lifespan. For example, FPGAs that use devices from Intel's Internet of Things Group have a guaranteed availability of 10 years, from start of production. |
| The client requires the system to be reprogrammable and have low inference time. | FPGAs have low inference time because of the ability to run many sections of the FPGA in parallel. FPGAs are also reprogrammable. |
| [TODO: Type your answer here] | [TODO: Type your answer here] |

Queue Monitoring Requirements

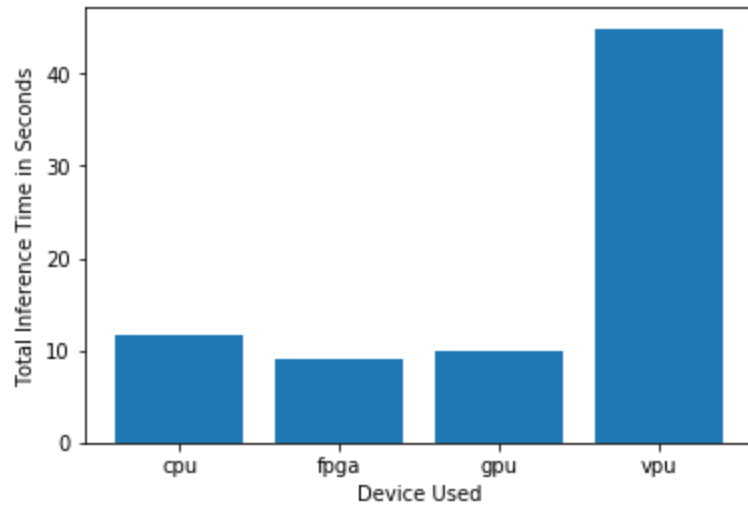
| | |
|--|------|
| Maximum number of people in the queue | 5 |
| Model precision chosen (FP32, FP16, or Int8) | FP16 |

Test Results

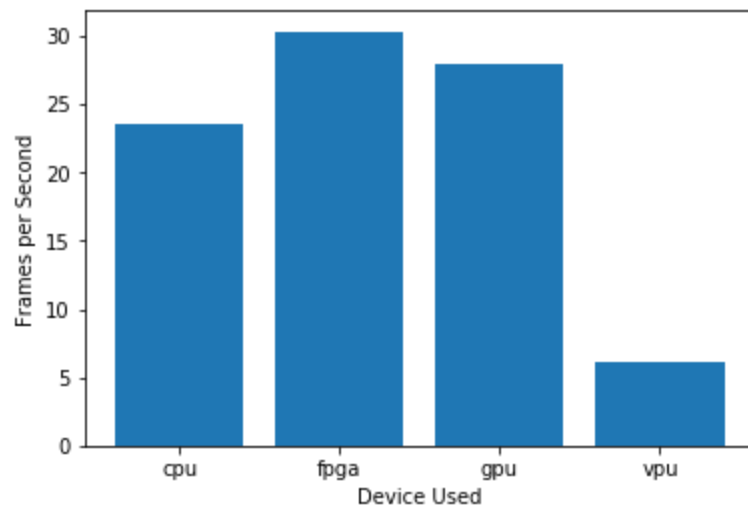
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

The FGPA had the fastest inference time and processed more frames per second, though it had the second lowest model load time but i think we can manage with that. The FPGA fits the clients needs nicely because of the need for reprogrammability and lasting for a long period of about 10 years.

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

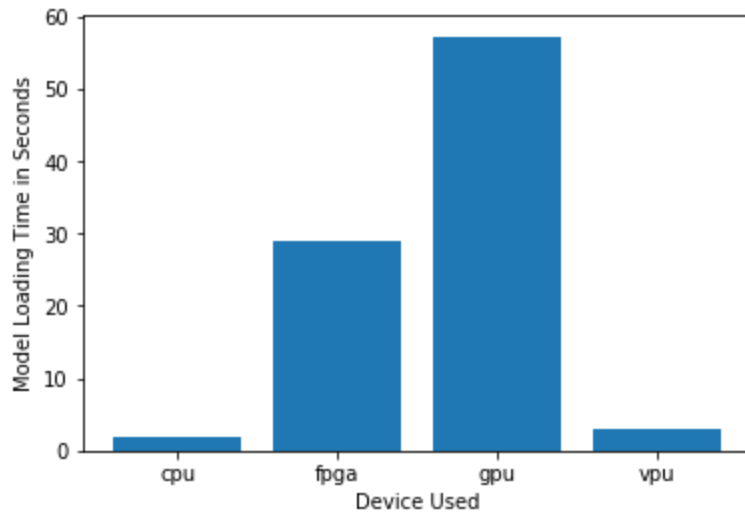
| Requirement Observed (Include at least two.) | How does the chosen hardware meet this requirement? |
|---|---|
| <i>Example requirement:</i> The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device. | <i>Example explanation:</i> VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| The client does not have money to invest in additional hardware. | The client already has systems with core i7 processors. |
| The client does not want to add to his electric bill. | This can be achieved by using the already available hardware. |
| [TODO: Type your answer here] | [TODO: Type your answer here] |

Queue Monitoring Requirements

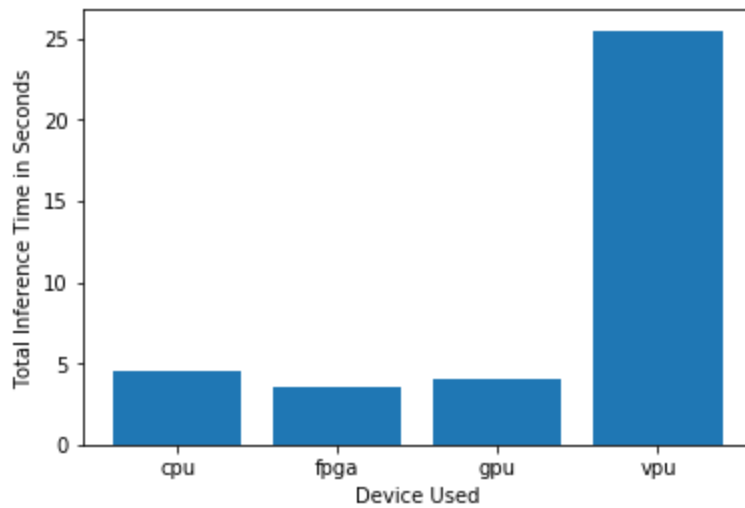
| | |
|---------------------------------------|---|
| Maximum number of people in the queue | 2 |
|---------------------------------------|---|

Test Results

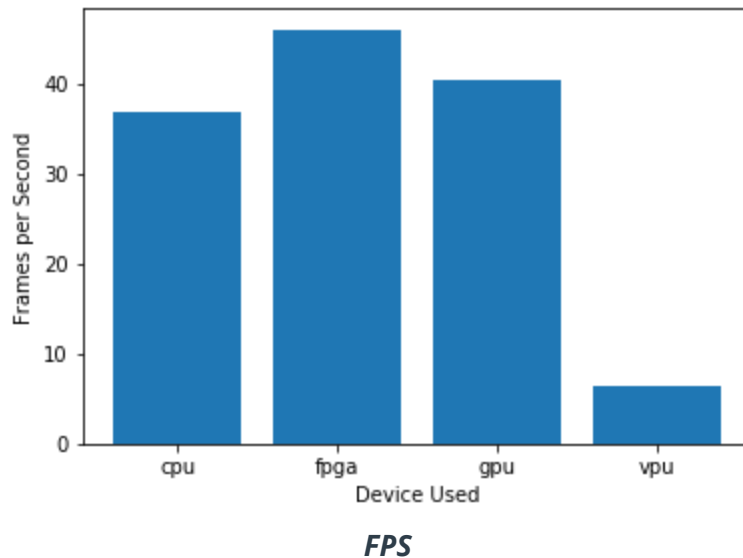
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

The CPU performs very well on all three fronts. It had the fastest model loading time. It's inference time and frames per second were also very good. I believe the CPU fits the clients needs especially as it doesn't require expenses on his part.

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?

Example requirement:

Example explanation:

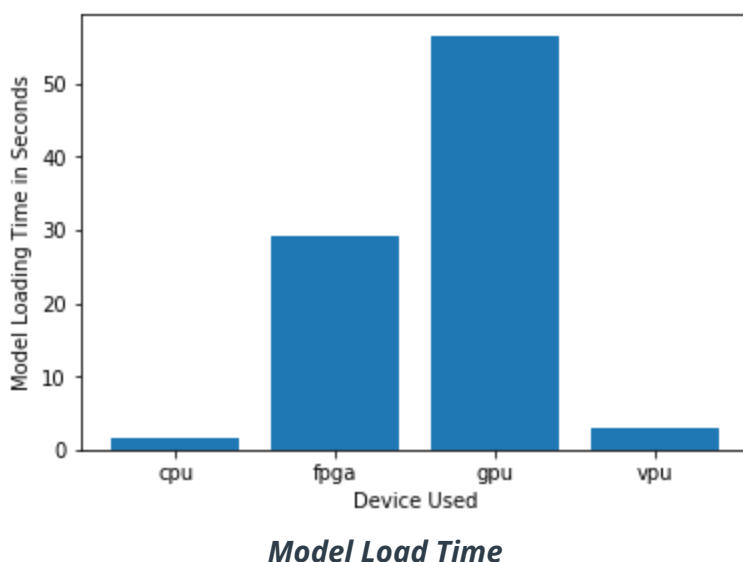
| | |
|--|--|
| The client requires a tiny device to be connected to their CPU—and their budget is only about \$100 for each device. | VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| The client has a budget of \$300 per machine | VPU or NCS2 would fit in this price range |
| The client requires a device with low power consumption. | NCS2 has a low power consumption of 1-2 watts. |
| <i>[TODO: Type your answer here]</i> | <i>[TODO: Type your answer here]</i> |

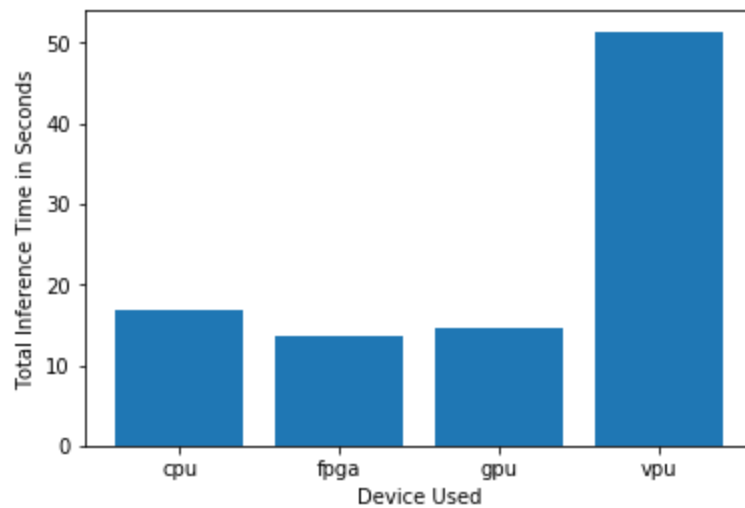
Queue Monitoring Requirements

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

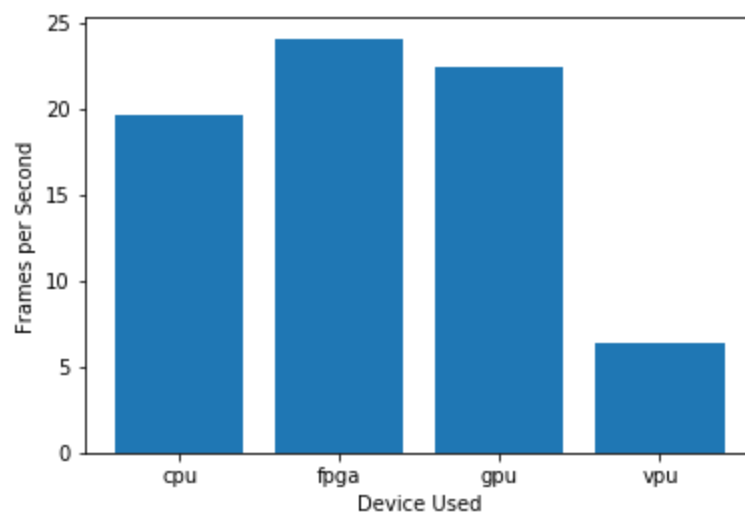
Test Results

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

The VPU had a very fast model load time. Though it performed badly in the frames per second and inference time. A little improvement can be achieved by adding multiple VPUs to run inferences in parallel. Due to the cost and power constraints, the VPU will fit the clients needs.