# Choose the Right Hardware

*Nnamdi Ajah*

---

## 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* |

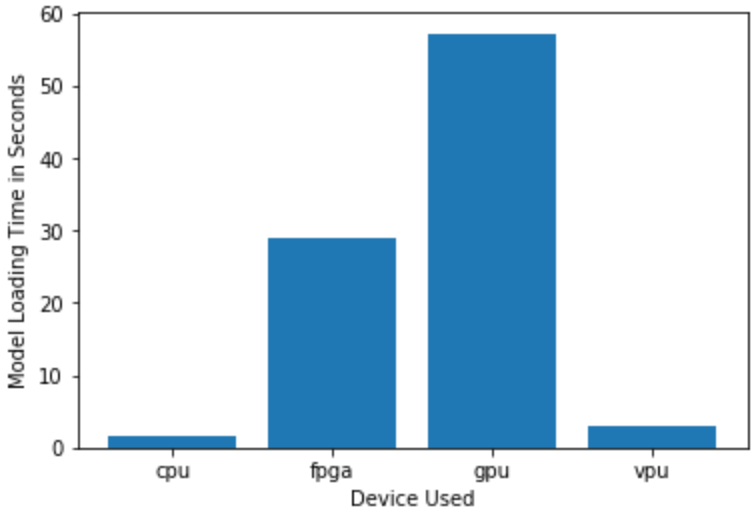
| **Requirement Observed (Include at least two.)** | **How does the chosen hardware meet this requirement?** |
|---|---|
| *Example requirement:* The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device. | *Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| *The client wants to install a system that is flexible to different use cases in his production line* | *FPGAs provide the flexibility of being able to be adapted to whatever scenario. I.e. easy configurability* |
| *The client intends to spend to make a quality system with long lifespan* | *FPGAs provide high performance and have a guaranteed availability of 10 years from start of production* |

### Queue Monitoring Requirements

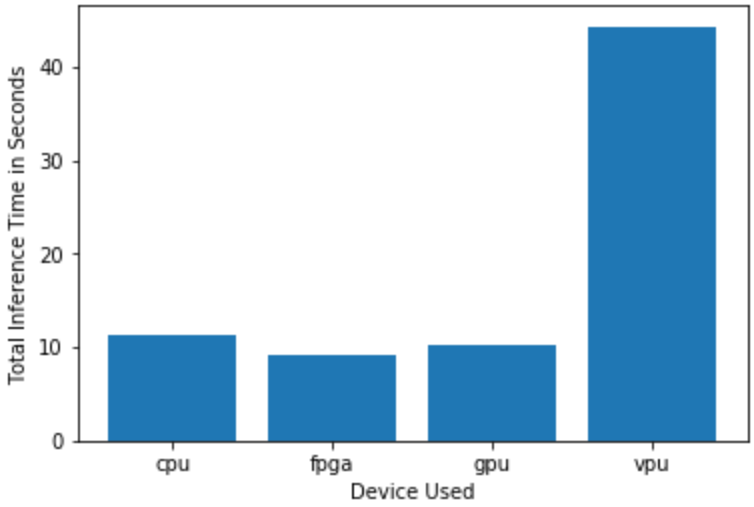
| | |
|---|---|
| **Maximum number of people in the queue** | *6* |
| **Model precision chosen (FP32, FP16, or Int8)** | *FP16* |

### Test Results

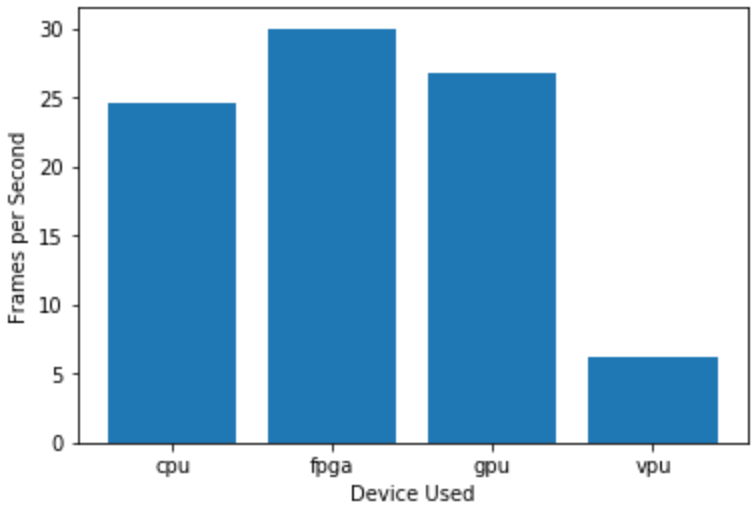
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).




***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 |
| --- |
| *Given that the client's important criteria are centred on flexibility of the system and a long lifespan, an FPGA becomes the obvious choice. Comparing the FPGA's performance with other hardware, it can be seen that the FGPA has a slow model loading time. Considering that the FPGA outperforms other hardware in inferencing time and frames per second, the slow loading time (which is a one-time event) is a good trade off for inferencing speed, FPS, long lasting and quality hardware provided by 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?<br>(CPU / IGPU / VPU / FPGA) |
| --- |
| *CPU* |

| Requirement Observed<br>(Include at least two.) | How does the chosen hardware meet this requirement? |
| --- | --- |
| *Example requirement:*<br>The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device. | *Example explanation:*<br>VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| *The client is on budget for his electrical bills* | *Using the existing CPUs in place will not add additional electrical bills* |
| *The client does not have much money to invest in additional hardware* | *The existing computionationally under-utilized CPUs can perform the required inferencing job.* |

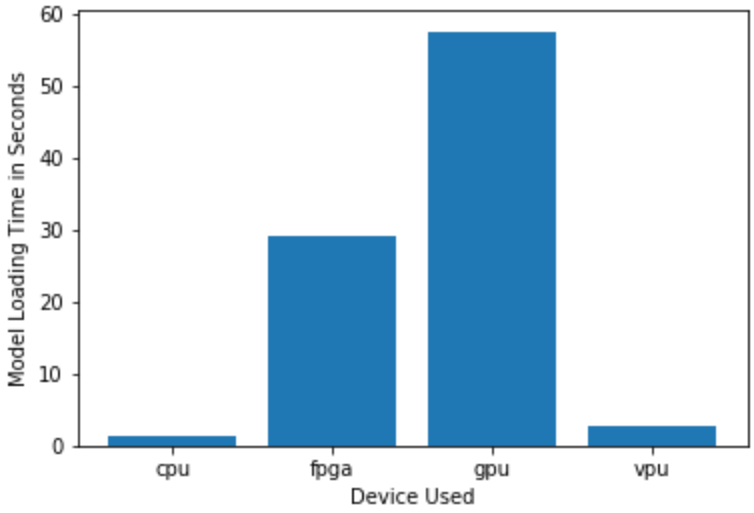
## Queue Monitoring Requirements

| | |
| --- | --- |
| **Maximum number of people in the queue** | *5* |

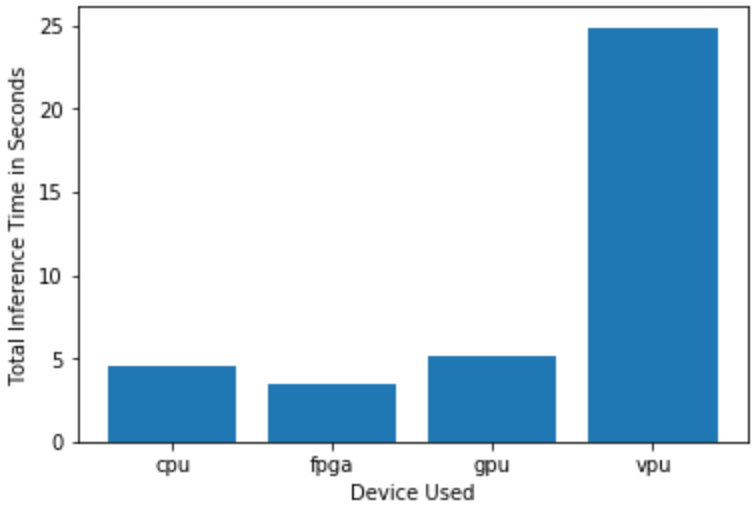
| Model precision chosen (FP32, FP16, or Int8) | *FP32* |
|---|---|

## Test Results

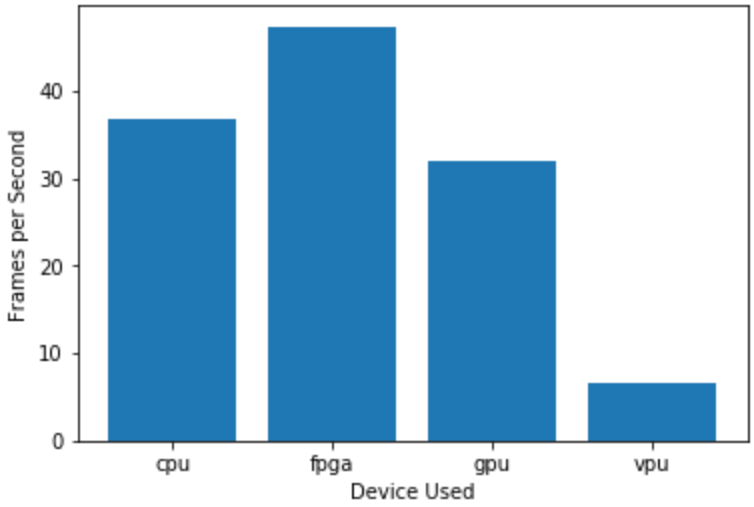
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).




***Model Load Time***




***Inference Time***




***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 |
| --- |
| *Since the client has computationally under-utilized CPUs at his facility and is concerned about electrical bills, making use of the existing CPUs makes the most sense. Especially given that the CPU is competitive in performance to other hardware - fgpa and igpu slightly outperform the cpu with respect to inference time and speed. But considering the capital and electrical cost that can be obtained from using the existing CPUs, it is just ideal to use the CPUs.* |

---

# 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?<br>(CPU / IGPU / VPU / FPGA) |
| --- |
| *VPU* |

| Requirement Observed | How does the chosen hardware meet this |
| --- | --- |

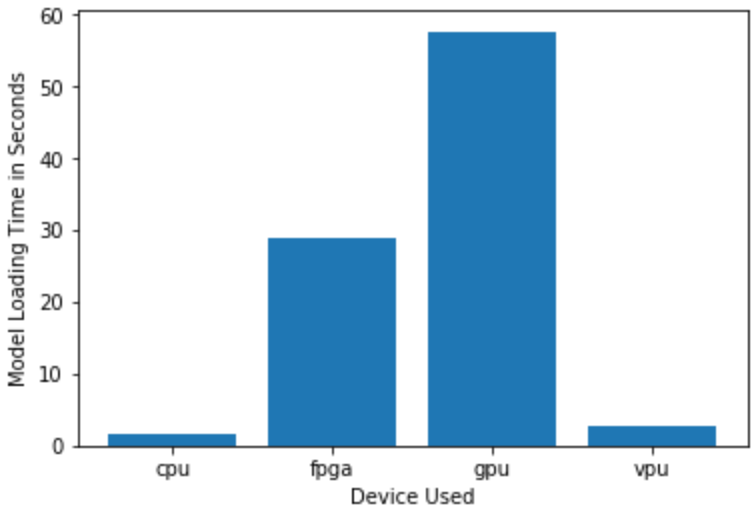
| (Include at least two.) | requirement? |
|---|---|
| *Example requirement:*<br>The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device. | *Example explanation:*<br>VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| *The client requires a device that can perform the inferencing task with minimum power* | *VPU has the lowest power consumption and can easily be added to any cpu.* |
| *While the client has a budget of $300 per machine, the client would like to save as much as possible both on hardware and future power requirements* | *A VPU costs at most a $100 and has low power requirements. Hence, the client will realize savings.* |

## Queue Monitoring Requirements

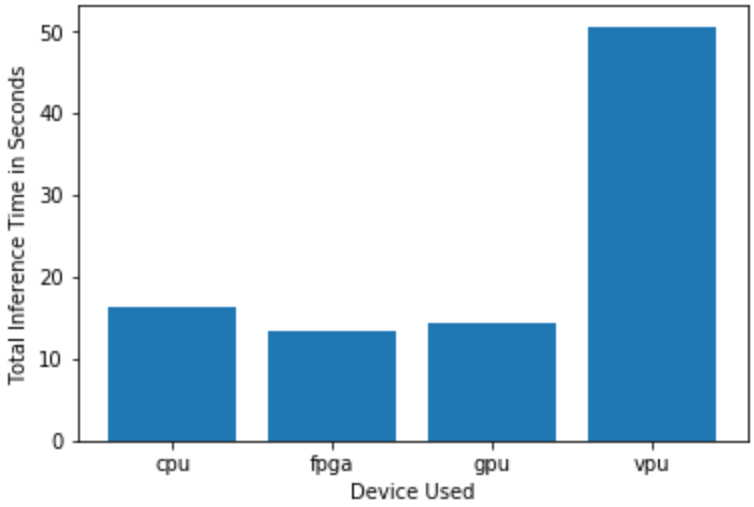
| | |
|---|---|
| **Maximum number of people in the queue** | *15* |
| **Model precision chosen (FP32, FP16, or Int8)** | *FP16* |

## Test Results

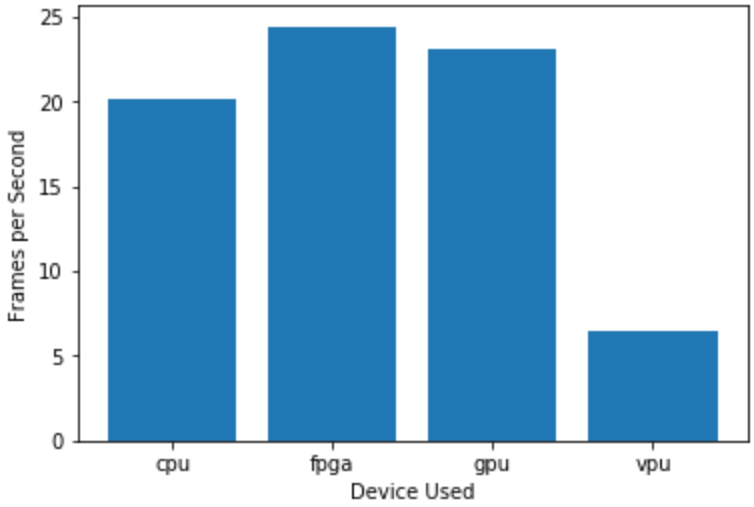
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).




***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 client requires a budget device that can perform inference and save on electric power. While the client has PCs on site, the computational resources on the PCs are scarce. Hence, the need for a low power and affordable device for inference - VPU. While the VPU lags other hardware in frames per second and inference time metric, it meets the client's core requirement on budget and energy consumption.* |