# 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* |

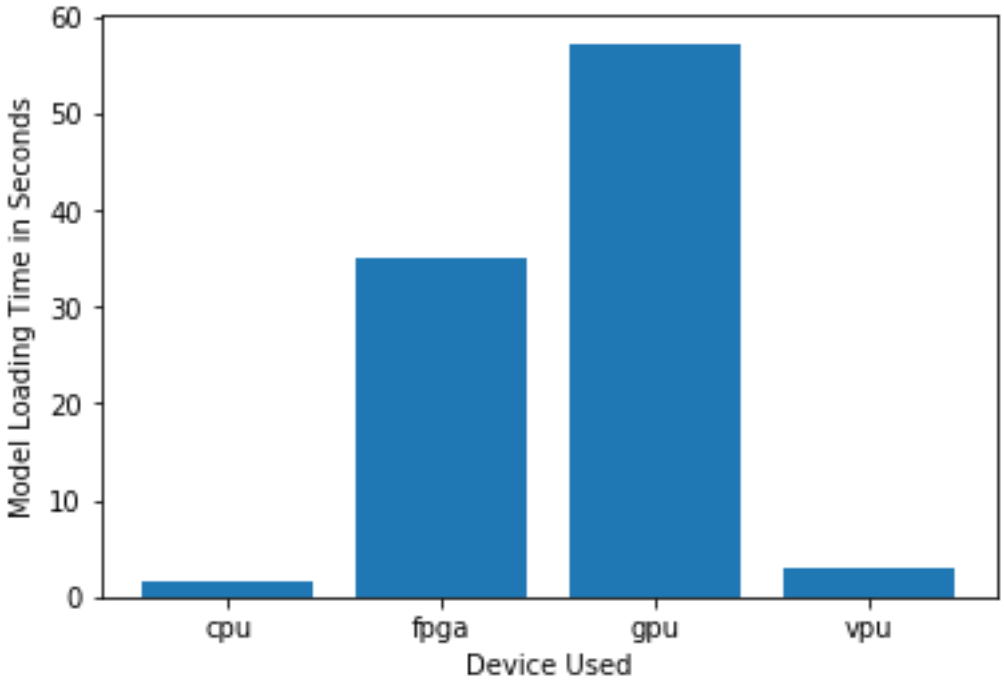
| **Requirement Observed (Include at least two.)** | **How does the chosen hardware meet this requirement?** |
|---|---|
| *Example requirement:* The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device. | *Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| *The client wants to reduce the slow-down in a production line due to the shift transitions. For this, he requires a system to be able to monitor the number of people in the factory line. The new system should use already installed cameras on each production line, which are able to record video at 35 FPS. Workers alternate shifts to keep the floor running 24/7.* | *FPGA, Field Programmable Gate Arrays. These devices will be the best option for the implementation of the required system. They are designed to continuously run 24 hours a day, 7 days a week, 365 days of the year. Additionally, one programmed with a suitable bitstream can be executed with high performance and very low latency, fulfilling the required monitoring of people in the production line of 5 times per second.* |
| *The budget is not a problem, and it is expected to be installed on a quality system that can be in use for the next 5-10 years.* | *FPGAs are not cheap, the cost of the proposed device, the Mustang-F100-A10, is around $1600 USD. Nonetheless, the budget is not a problem. Moreover, FPGAs have a long lifespan, around 10 years from the start of production.* |
| *Furthermore, the system needs to be able to adapt to new problems once the present one is solved.* | *FPGAs are reprogrammable devices, meaning they can be used for different purposes depending on their programming.* |

## Queue Monitoring Requirements

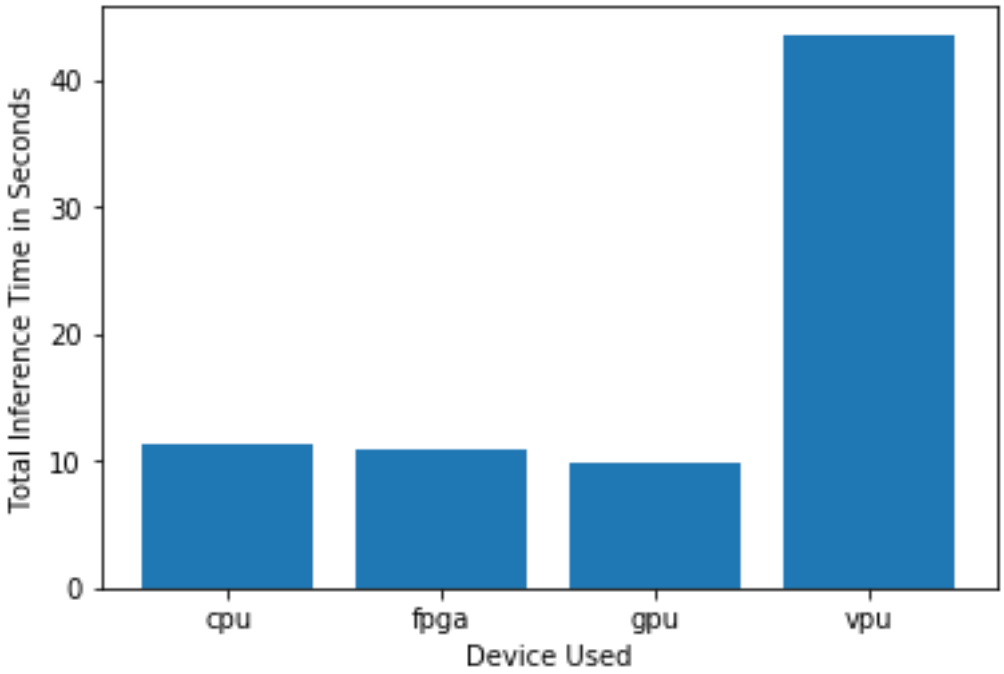
| | |
|---|---|
| **Maximum number of people in the queue** | *5* |
| **Model precision chosen (FP32, FP16, or Int8)** | *FP16* |

## Test Results

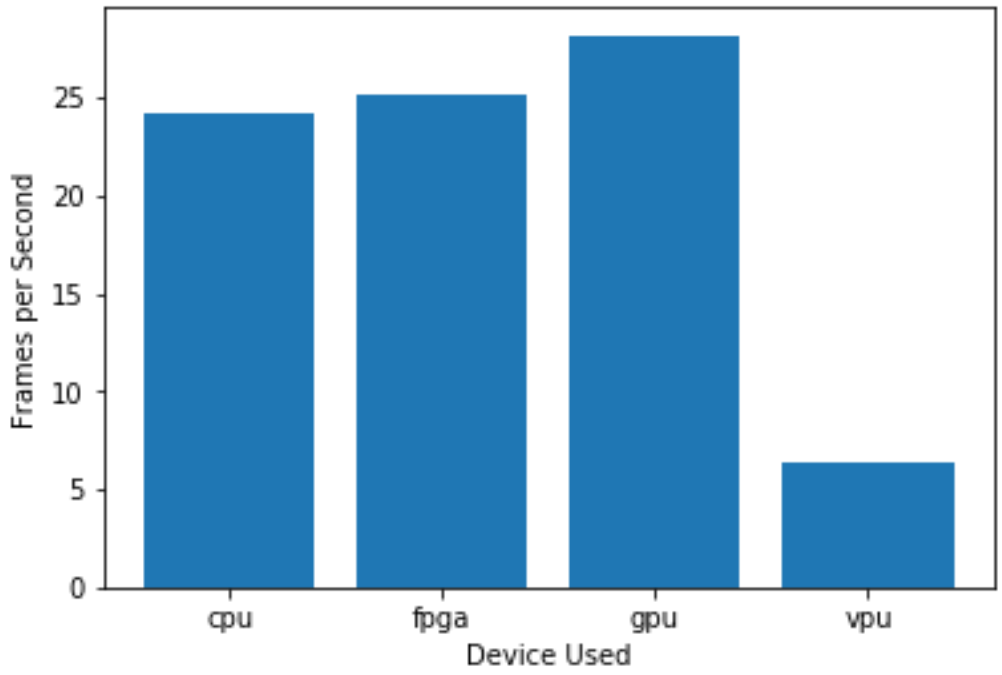
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).




***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 |
| --- |
| *Mr. Vishwas presented a problem that can be solved efficiently using the Mustang-F100-A10 device, an FPGA. The cost of the proposed device can be absorbed during the long lifespan associated with the proposed device, around 10 years. More importantly, these devices can run non-stopping and with low latency, allowing the monitoring of people as required by Mr. Vishwas. Additionally, once the employees have learned how to avoid the production slow-down, the selected FPGA device can be used for other purposes, as desired by the client. Therefore, the Mustang-F100-A10 is the best device to solve the required problem.*<br><br>*The performance of the system can be compared to systems based on other devices. It can be seen that the system based on the FPGA is able to make 10 inferences per second, double of the required by Mr. Vishwas.* |

---

# 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) |
| --- |
| *IGPU* |

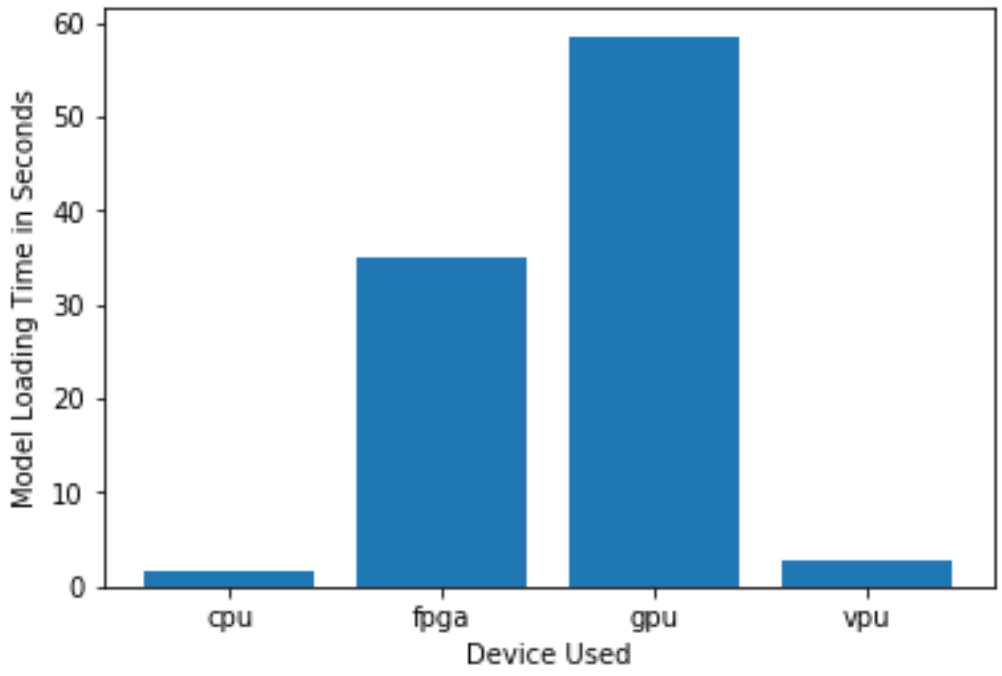
| Requirement Observed<br>(Include at least two.) | How does the chosen hardware meet this requirement? |
| --- | --- |
| *Example requirement:*<br>The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device. | *Example explanation:*<br>VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| *Mr. Lin, a store manager of a small PriceRight Singapore outlet, wants to maximize profit using AI. He needs to control the checkout queues to avoid congestions. Each store checkout counter is equipped with a not fully used i7 processor base computer, it is assumed each processor has an integrated GPU.* | *The best device to solve this problem is the use of a GPU, given that they are already incorporated into the available hardware and have native support for video playback functions. Additionally, since the CPU is not fully used, the shared memory with the incorporated GPU should be enough to implement the solution required by Mr. Lin.* |
| *Limited budget, preferably the system should be based on existing hardware.* | *A lot of money can be saved if using a GPU to solve the problem since they are already incorporated into the available i7 processors.* |
| *Saving energy* | *Incorporate GPUs devices have configurable power consumption. Unused sections of the GPU can be powered down to reduce power consumption.* |

## Queue Monitoring Requirements

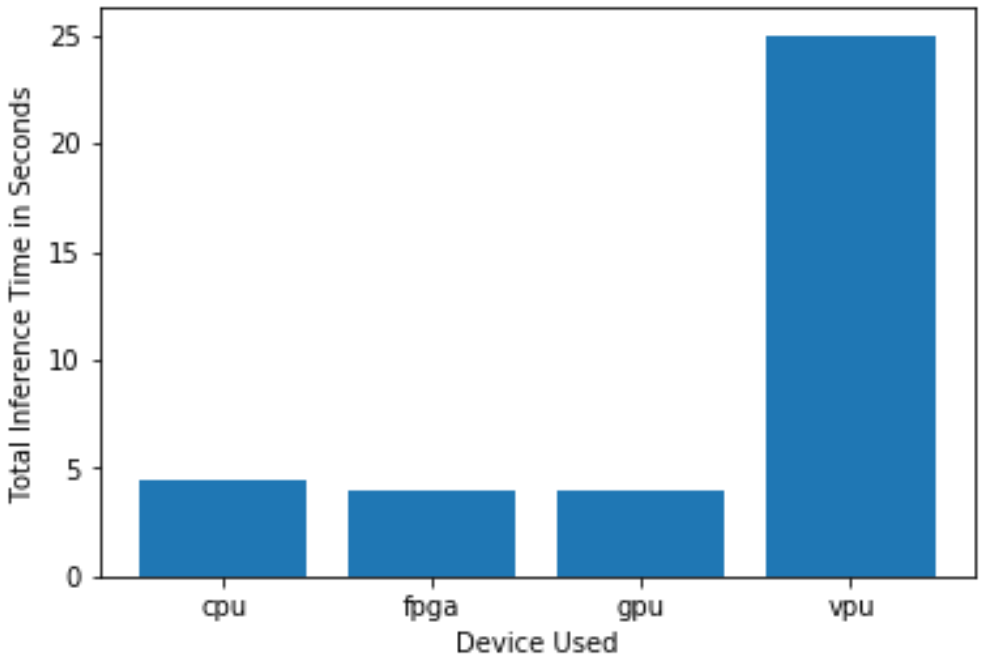
| | |
|---|---|
| **Maximum number of people in the queue** | *3* |
| **Model precision chosen (FP32, FP16, or Int8)** | *FP16* |

## Test Results

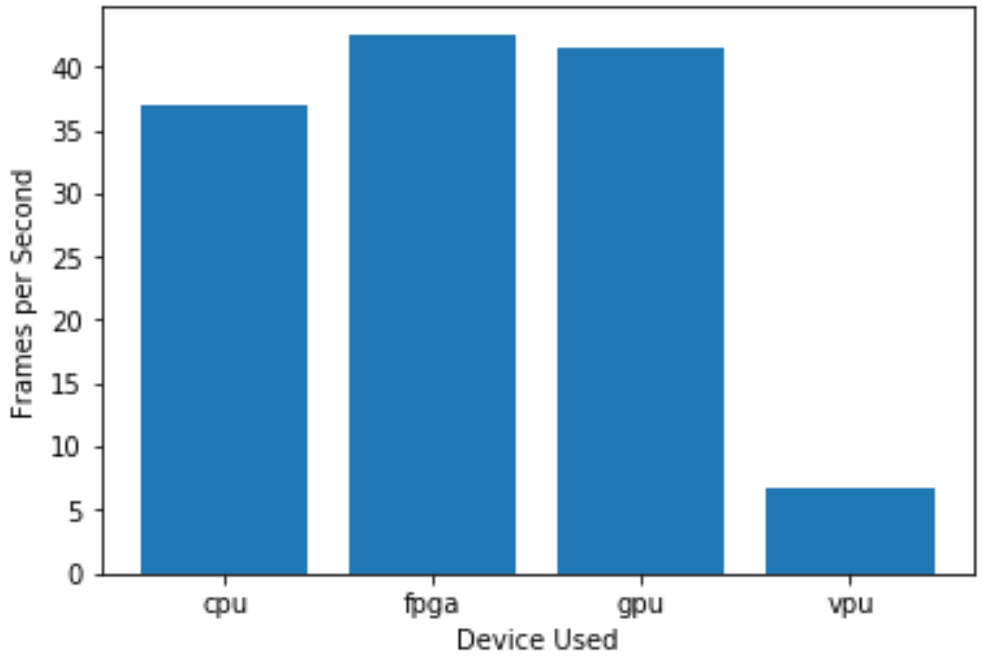
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).




***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 |
| --- |
| *Thus, GPU is the best device to solve the problem presented by Mr. Lin. As mentioned before, due to the limited budget, the available hardware and the needs to save energy, a system based on an i7 processor integrated GPU is the best option to solve the problem.*<br><br>*Nonetheless, as it can be seen in the previously presented images, the number of inferences per second that a system based on a GPU is the smaller among all analyzed devices. However, for the problem is more than enough since the idea is to analyze the number of persons on a queue, therefore, it should be enough to cover the variance in the number of people present in the queues.* |

---

# 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* |

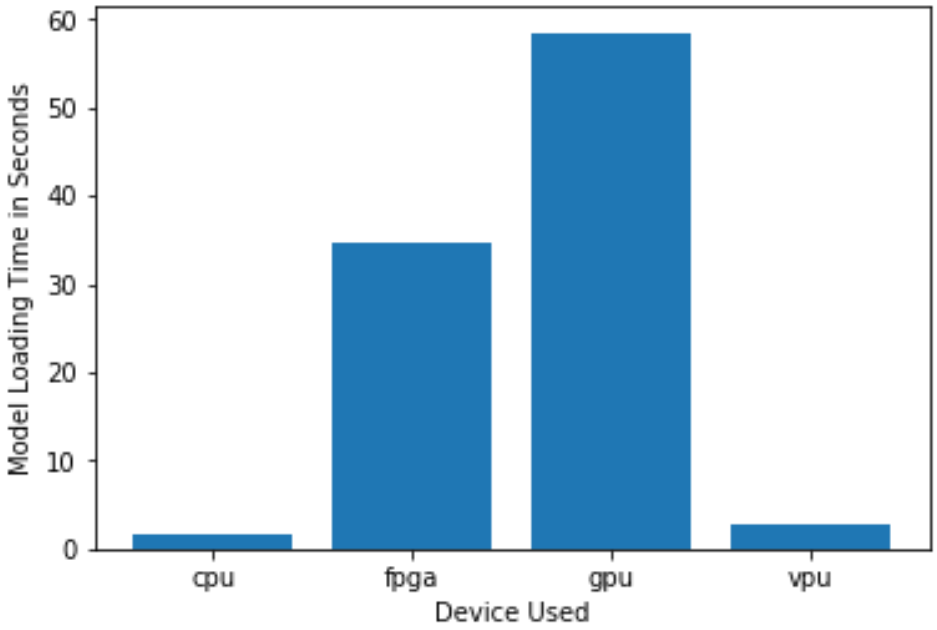
| Requirement Observed<br>(Include at least two.) | How does the chosen hardware meet this requirement? |
| --- | --- |
| *Example requirement:*<br>The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device. | *Example explanation:*<br>VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| *Ms. Leah, the Innovation head for Delhi Metro Rail Services, needs a solution to improve the handling of congestions during rush hours. The solution needs to direct passengers to less congested areas during peak time. Currently, the Delhi Metro Rail Services has already implemented 7 CCTV cameras on the platform, controlled by All-In-One PCs that process the video footage, PCs that cannot handle extra-processing needed for this problem.* | *VPUs based systems give a good solution for the presented problem, especially when considering that the currently used hardware is running at full processing capacity. Therefore, the use of an AI accelerator can help to make an inference from the video footage obtained from the 7 CCTV system. VPUs are useful for this problem since they are specialized in AI tasks related to computer vision problems.* |
| *Ms. Leah's budget allows for a maximum of $300 per machine.* | *The chosen VPU device for the solution of the problem is the Intel Neural Compute Stick 2, NCS2, with a cost of around $70 USD per device.* |
| *The budget should consider future implementations, to save in future requirements.* | *If required, additional NCS2 devices can be used to run in parallel to solve more complex problems.* |

## Queue Monitoring Requirements

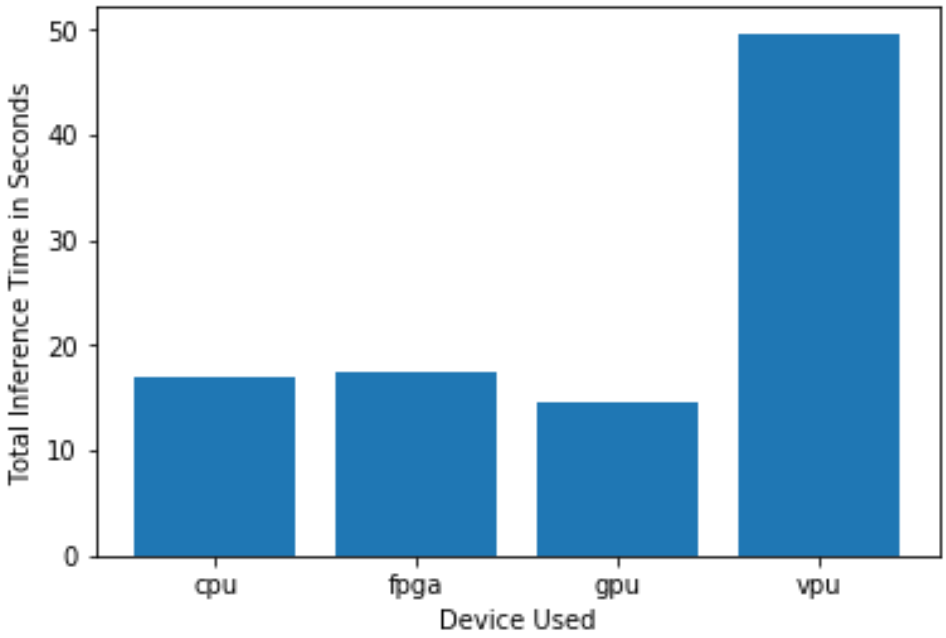
| | |
|---|---|
| **Maximum number of people in the queue** | *7* |
| **Model precision chosen (FP32, FP16, or Int8)** | *FP16* |

## Test Results

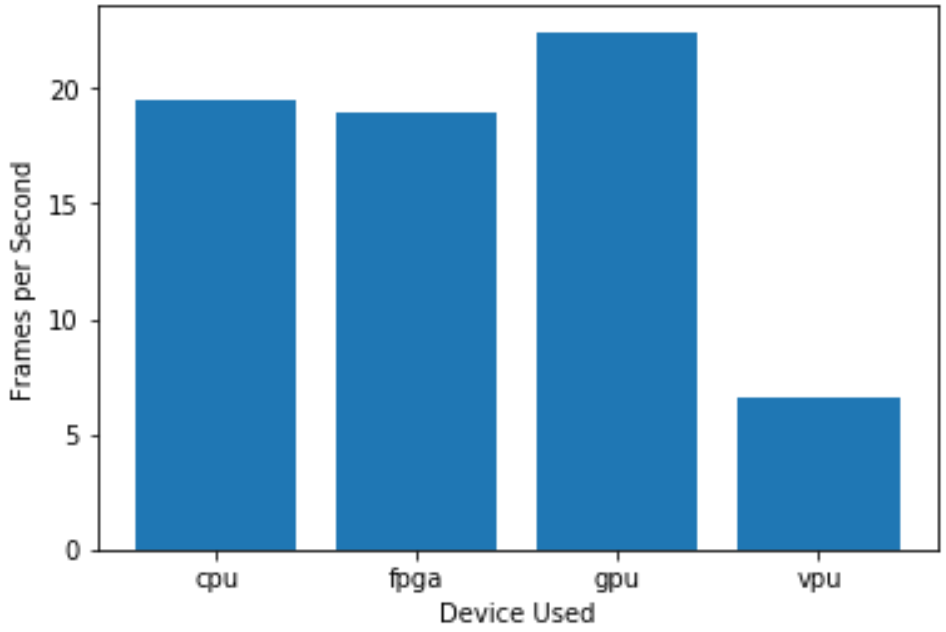
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).




***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 solution for Ms. Leah problem should consider the use of VPUs as the core for the inferences required, especially when considering the limited budget and that the currently used hardware is running at full processing capacity. Furthermore, the NCS2 characteristics allow the use of different devices running in parallel, allowing more processing in case needed in the future. Finally, the power consumption on these devices is very small, around  1W. Therefore, is the perfect device for the implementation of a solution for Ms. Leah problem* |