# 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)** |
|---|
| *[TODO: Field Programmable Gate Array (Intel Mustang-F100-A10)]* |

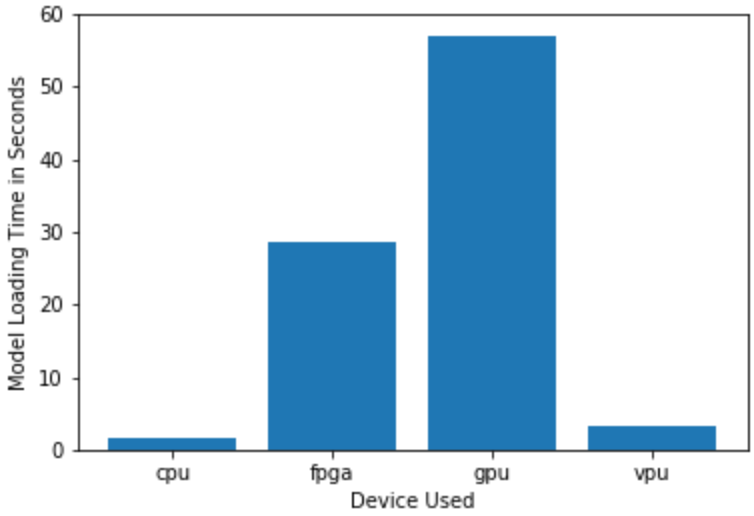
| **Requirement Observed (Include at least two.)** | **How does the chosen hardware meet this requirement?** |
|---|---|
| *Example requirement:* The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device. | *Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| *[TODO: Flexibility]* | *[TODO: FPGAs are known to be flexible and can be reprogrammed quickly depending upon our needs]* |
| *[TODO: Long term solution]* | *[TODO: The solution can last for 5 to 10 years. FPGAs can run 24/7 for a whole year]* |
| *[TODO: Economic constraints]* | *[TODO: Mr. Vishwash has enough for this significant investment]* |

### Queue Monitoring Requirements

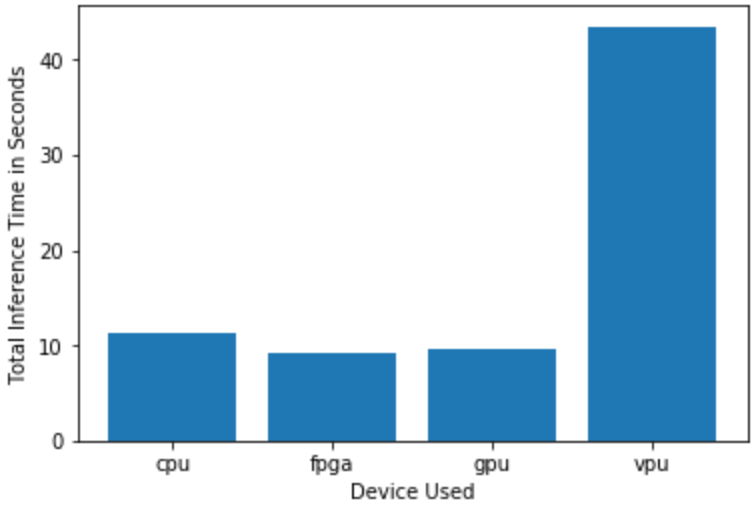
| | |
|---|---|
| **Maximum number of people in the queue** | *[TODO: 5]* |
| **Model precision chosen (FP32, FP16, or Int8)** | *[TODO: FP16 ]* |

### Test Results

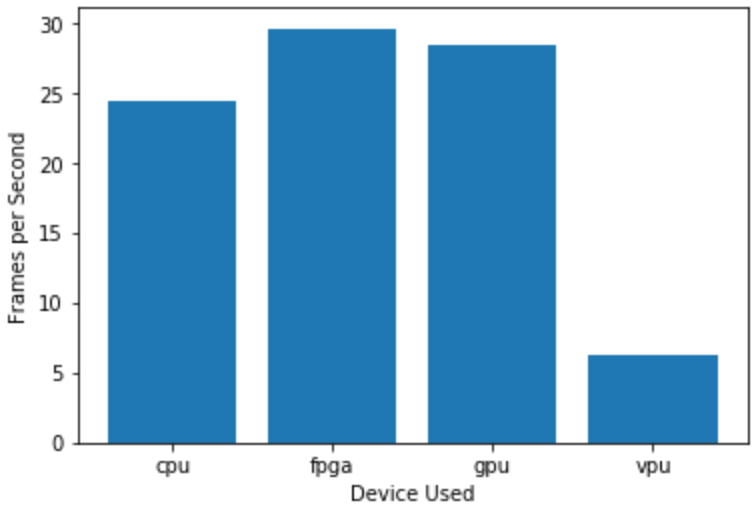
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).




***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 |
| --- |
| *[TODO: - The client needs the system to be flexible so that it can be optimized as required.*<br>*- FPGA stands for Field Programmable Gate Array, which means they can be customized (reprogrammed).*<br>*- The client's camera records video at 30-35 FPS, FPGA reads approximately 30FPS, the client's requirement.*<br>*- The client wants a system for 5-10 years and FPGA guarantees a duration of 10 years.*<br>*- The model load time of FPGA is less than GPU and greater than CPU and VPU.*<br>*- The client also wants the system to run inference quickly: FPGA has the lowest Inference time. So, FPGA meets the client's requirements.]* |

---

# 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) |
| --- |
| *[TODO: IGPU (Integrated Graphics Processing Unit)]* |

| 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. |
| *[TODO: Economic Constraints]* | *[TODO: Most of the store's checkout counters already have a modern computer, each of which has an Intel i7 core processor. Currently these processors are only used to carry out some minimal tasks that are not computationally expensive.]* |

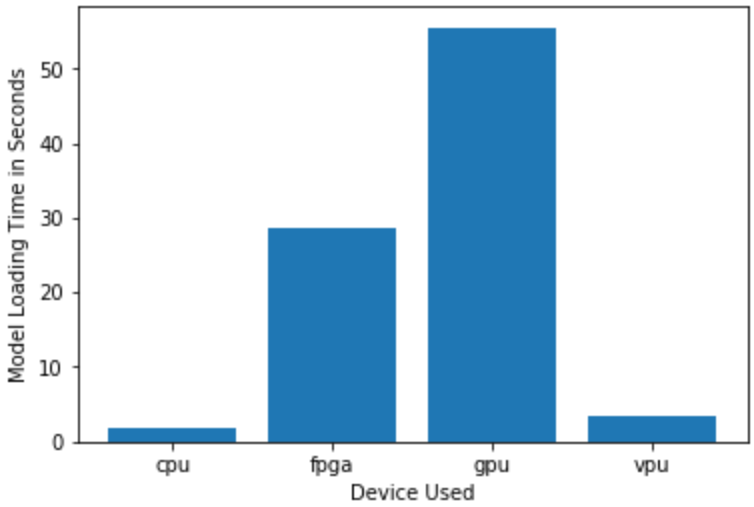
| | |
|---|---|
| *[TODO: Energy Constraints]* | *[TODO: Mr. Lin does not have much money to invest in additional hardware, and also would like to save as much as possible on his electric bill. Pre-installed CPUs are the best solution]* |
| *[TODO: Type your answer here]* | *[TODO: Type your answer here]* |

## Queue Monitoring Requirements

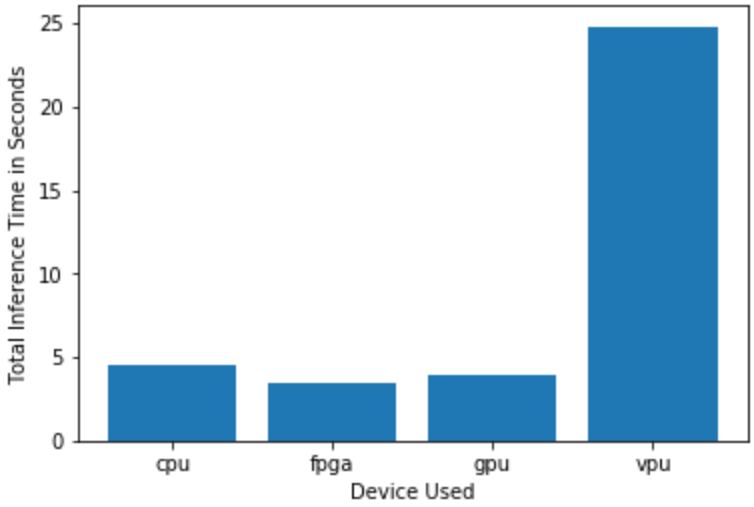
| | |
|---|---|
| **Maximum number of people in the queue** | *[TODO: 2 (during normal daily hours) - 5 (during rush hours)]* |
| **Model precision chosen (FP32, FP16, or Int8)** | *[TODO: FP16 ]* |

## Test Results

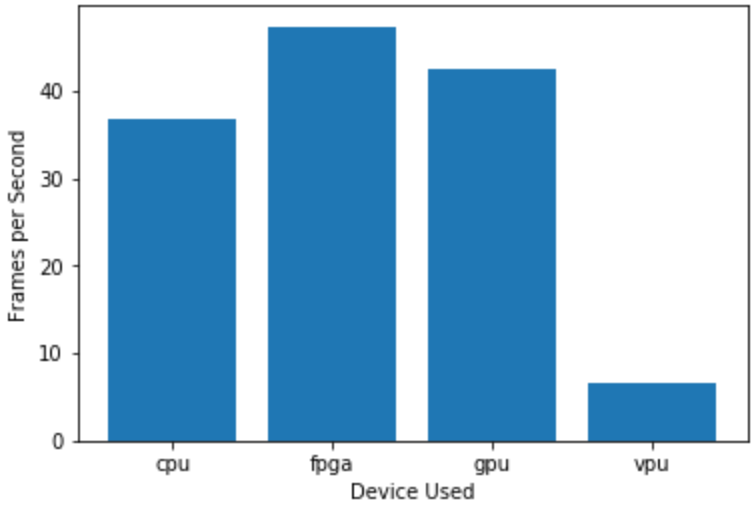
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).




***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** |
| --- |
| *[TODO:*<br>*- The client does not have enough money to invest in additional hardware. An IGPU (Integrated Graphics Processing Unit) is a GPU which comes along with the CPU core(Intel i7 in this case)*<br>*- The client wants to save as much as possible on electric bills. On an IGPU, the clock rate for the slice and unslice can be controlled separately. This means that unused sections in a GPU can be powered down to reduce power consumption.* |

| |
|---|
| *- It can be observed that an IGPU has less inference time than a CPU which produces quick processing. IGPU reads the video with more FPS than a CPU.]* |

---

# 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)** |
|---|
| *[TODO: Vision Processing Unit(Intel Neural Compute Stick 2)]* |

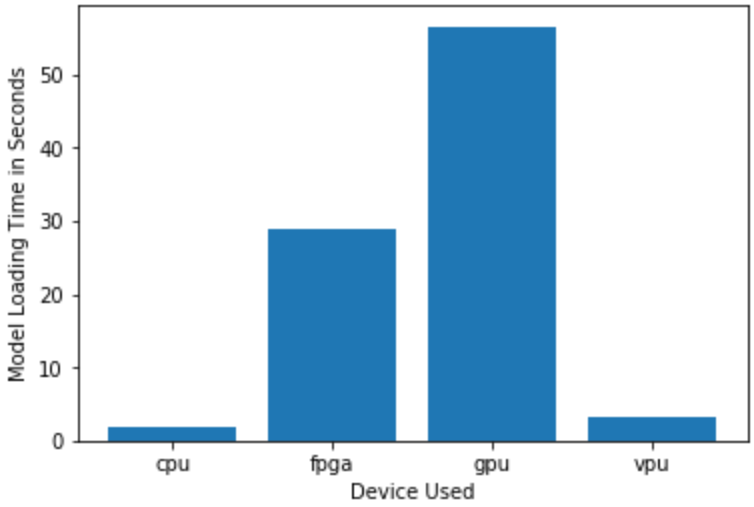
| **Requirement Observed (Include at least two.)** | **How does the chosen hardware meet this requirement?** |
|---|---|
| *Example requirement:* The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device. | *Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range. |
| *[TODO: Economic Constraints]* | *[TODO: The client's maximum cost that can afford it $300 per machine, so the VPU is the best solution for under $100 each]* |
| *[TODO: Energy Constraints]* | *[TODO: The VPU can be run with no additional power using one machine with CCTV footage in use]* |
| *[TODO: Performance]* | *[TODO: The VPU can run in high performance along with the 7 CCTV cameras]* |

## Queue Monitoring Requirements

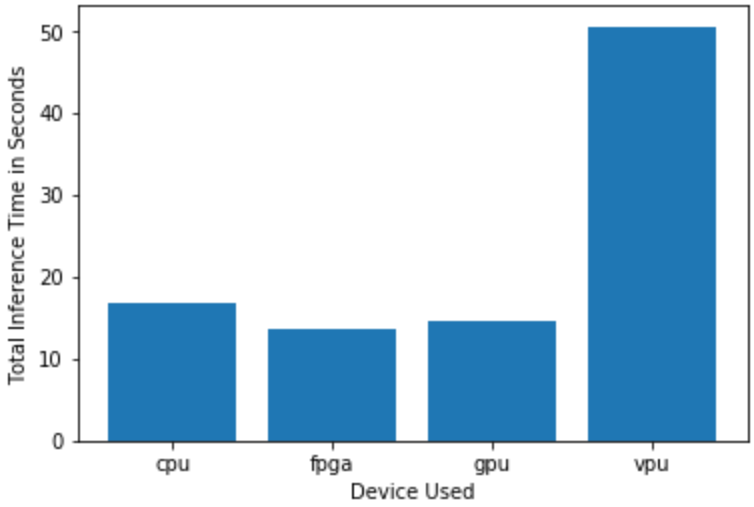
| | |
|---|---|
| **Maximum number of people in the queue** | *[TODO: 7(non-peak hours) - 15(peak hours)]* |
| **Model precision chosen (FP32, FP16, or Int8)** | *[TODO: FP16 ]* |

## Test Results

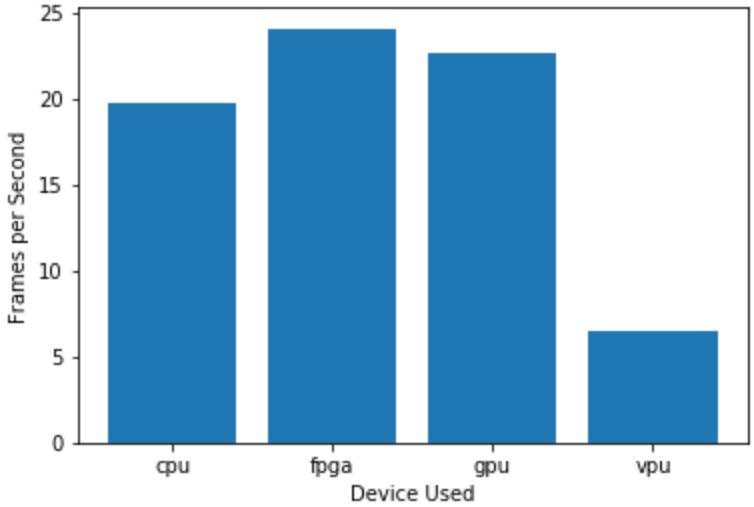
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).




***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** |
| --- |
| *[TODO:*<br>*- The client would like to spend $300 per machine and a VPU i.e. Neural Compute Stick 2, has a price range of $70-$100.*<br>*- The client uses 7 CCTV cameras and the data from these cameras is processed by a single computer (CPU). The processor in the NCS2 is an AI Accelerator specifically designed to handle AI requirements and speed up processes used in AI and machine learning.*<br>*- The client would also require a reduction in power usage. The NCS2, has a very low power consumption of only 1-2 watts.*<br>*- The inference time of VPU is the highest.FPGA has the lowest Inference time, but it is very expensive that can cost more than $300.*<br>*- The model loading time of VPU is less than GPU and FPGA but it is greater than CPU.*<br>*- Also, VPU's FPS is less than all the other hardwares(i.e. CPU, GPU, FPGA) which don't meet the client's requirement. VPU i,e, Neural Compute Stick 2 meets all the client's requirements.]* |