

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)
Field Programmable Gate Array (Intel Mustang-F100-A10)

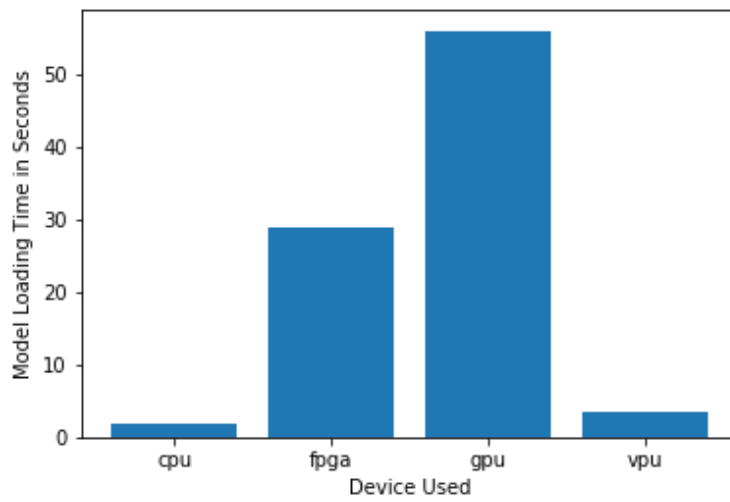
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
The client wants the hardware to be flexible so that it can be reprogrammed and optimized to quickly detect flaws in different chip designs.	FPGAs are known to be flexible and can be reprogrammed quickly depending upon our needs.
The client wants to monitor the number of people in the factory line and their camera can record video at 30-35 FPS which means each frame needs to be processed in $1/30 = 0.03s$ i.e 30ms (microseconds)	Intel Mustang-F100-A10 is capable of providing such processing.
The client also wants the device to last for 5-10 years	An FPGA is guaranteed to have a life span of 10 years.

Queue Monitoring Requirements

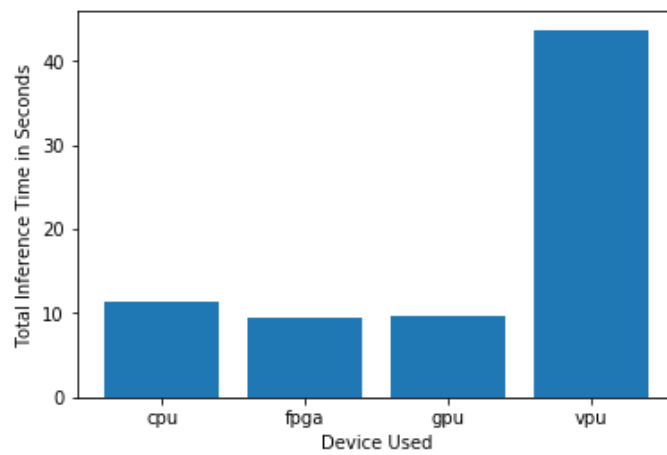
Maximum number of people in the queue	5
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

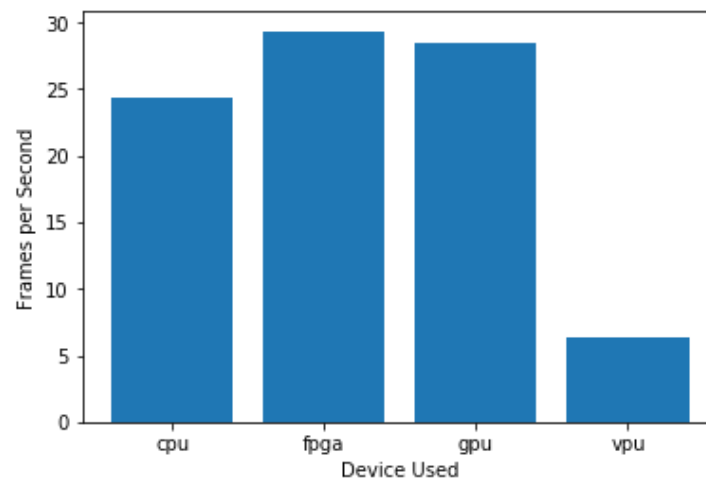
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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 client needs the system to be flexible so that it can be reprogrammed and optimized as and when required. FPGA stands for **Field Programmable** Gate Array, which means they can be customized(reprogrammed) as per the requirements.*
- *The client wants a system to last for 5-10 years and FPGA guarantees a life span 10 years.*
- *The client's camera records video at 30-35 FPS and as we can see from the graph, FPGA read approximately 30FPS, hence meeting the client's requirement.*
- *The model load time of FPGA is less than GPU and greater than CPU and VPU, but this does not meet the client requirement.*
- *The client also wants the system to run inference quickly and as we can see from the above graph, FPGA has the lowest Inference time, therefore meeting the client's requirement. Hence, 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? (CPU / IGPU / VPU / FPGA)
IGPU (Integrated Graphics Processing Unit)

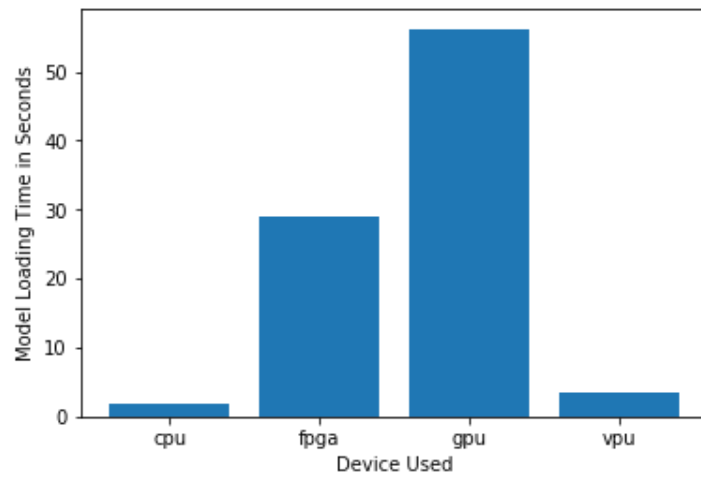
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>The client does not to invest in additional hardware</i>	<i>The client already has a modern computer at every counter, each of which has an Intel i7 core processor and an IGPU comes along with the CPU core</i>
<i>The client mentions saving as much as possible on his electric bill.</i>	<i>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.</i>

Queue Monitoring Requirements

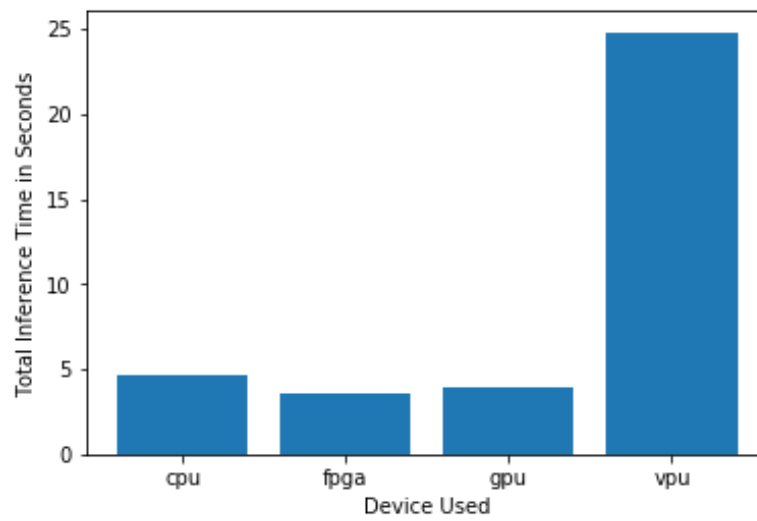
Maximum number of people in the queue	2 (during normal daily hours) - 5 (during rush hours).
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

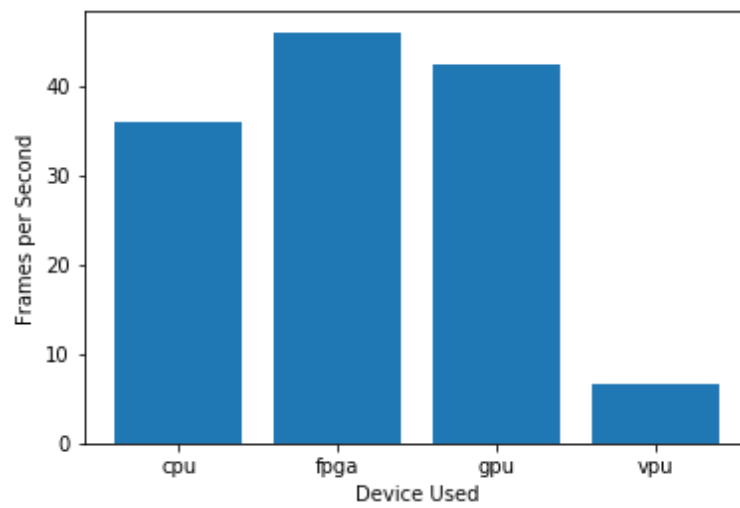
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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 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), therefore, no need for additional hardware investment.*
- *The client would like 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, hence this requirement is satisfied.*
- *It can observe from the graph that an IGPU has less inference time than a CPU which ensures quick processing.*
- *Also we can observe from the graph that an IGPU reads the video with more FPS than a CPU. Although the GPU takes more loading time than CPU, it does not meet the client's requirement. Hence, an IGPU meets the client's requirements.*

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)
<i>Vision Processing Unit(Intel Neural Compute Stick 2)</i>

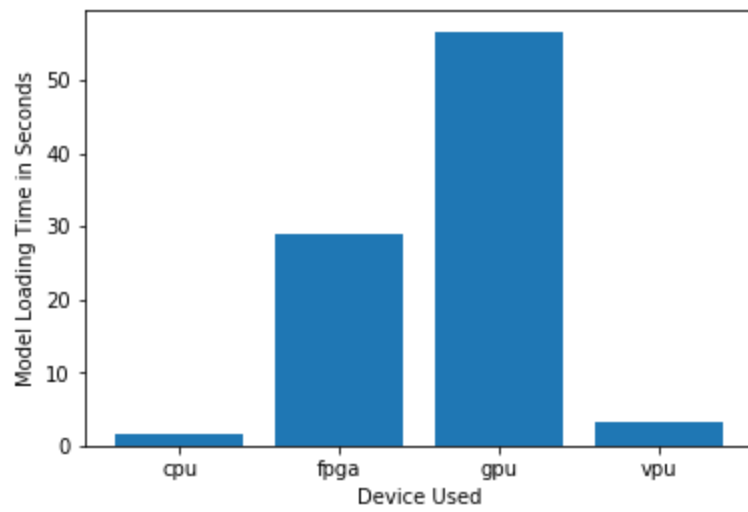
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>The client can spend \$300 per machine</i>	<i>The Neural Compute Stick 2 would meet the requirements here typically costing around \$70 to \$100.</i>
<i>The client mentions the feed from 7 CCTV cameras is managed by one PC, hence it requires high processing speed with low power consumption.</i>	<i>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. Also, has a very low power consumption of only 1-2 watts.</i>

Queue Monitoring Requirements

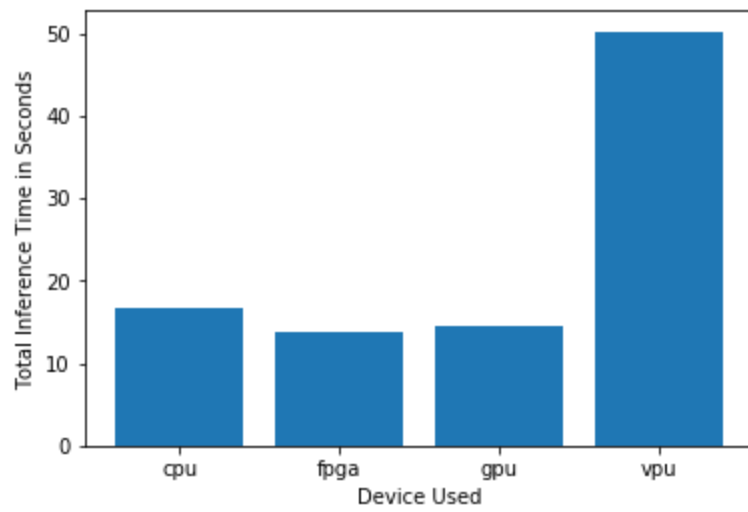
Maximum number of people in the queue	<i>7(non-peak hours) - 15(peak hours)</i>
Model precision chosen (FP32, FP16, or Int8)	<i>FP16</i>

Test Results

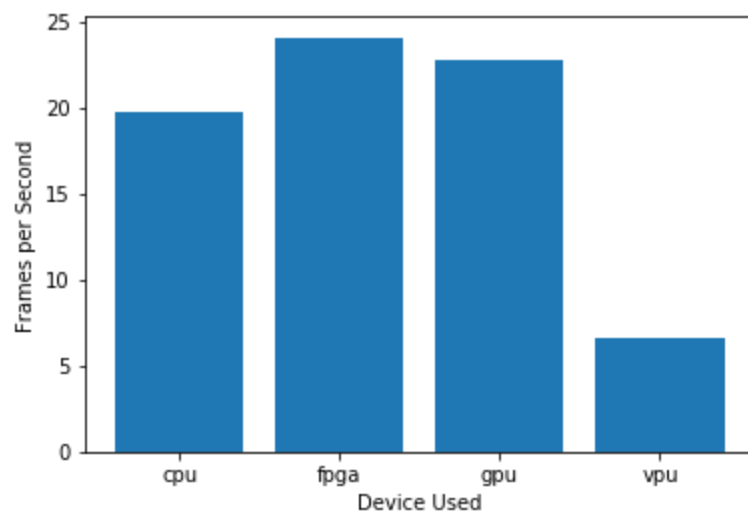
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

- *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. Hence this meets the client's requirement.*
- *The client uses 7CCTV 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. Hence, the client's requirement to high processing speed is satisfied.*
- *The client would also require a reduction in power usage. The NCS2, has a very low power consumption of only 1-2 watts.*
- *As we can see from the graph, the inference time of VPU is the highest. But the other hardwares(i.e. CPU, GPU and FPGA) does not meet the client's requirement. FPGA has the lowest Inference time, but it is very expensive costing greater than \$300.*
- *It can be observed from the graph, the model loading time of VPU is less than GPU and FPGA but it is greater than CPU. Using a CPU for this scenario is not recommendable, because the CPU requires high power for high computing(processing and running inference on data from 7 CCTV cameras).*
- *Also, VPU's FPS is less than all the other hardwares(i.e. CPU, GPU, FPGA) which does not meet the client's requirement. Hence, VPU i.e, Neural Compute Stick 2 meets all the client's requirements.*