

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

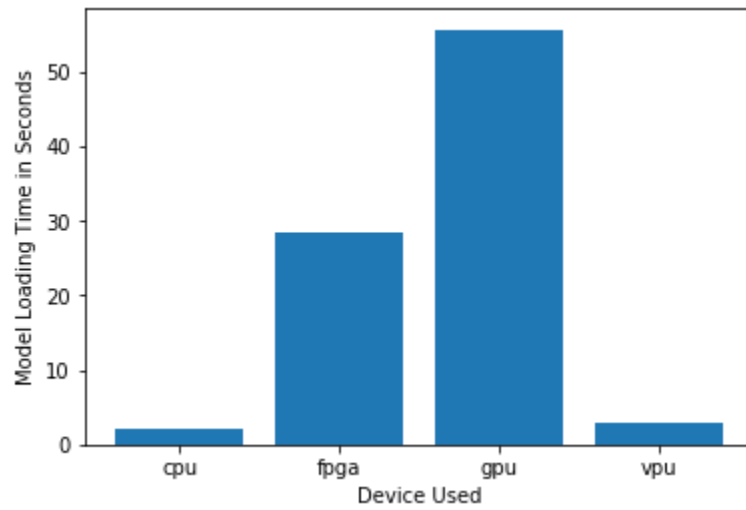
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
<i>The client has cameras installed on the conveyor belt which process 30-35 frames per second and this needs to be improved.</i>	This shows the slow latency in the current system and is a major concern. A FPGA can be programmed for solving this issue since the bitstream can be tweaked accordingly and FPGA's do not send the output back to the cpu which makes inference lot faster.
<i>To detect chip flaws without slowing the production line.</i>	FPGA's can be easily programmed as AI accelerators for the high level abstraction involved.
<i>For customized solutions, and requirements in various components according to the addressed problem</i>	FPGA can be easily be reprogrammed according to the problems and provide customized solutions.
<i>The client expects it to last 5-10 years</i>	FPGA's have a very long lifespan and would easily solve this purpose.

Queue Monitoring Requirements

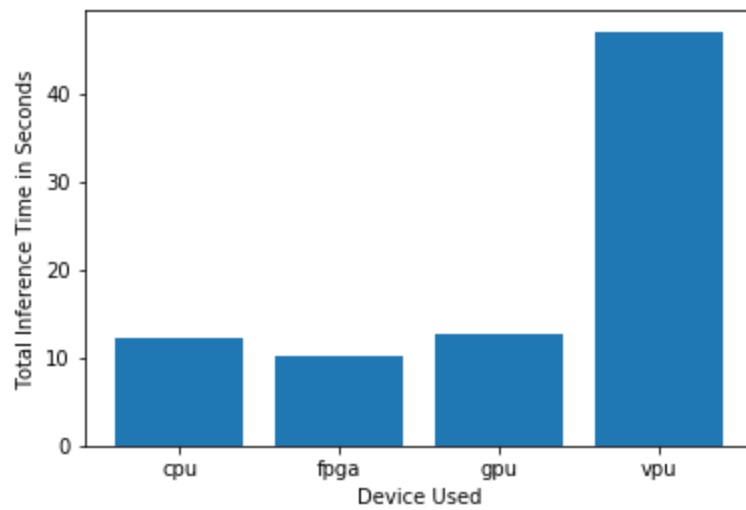
Maximum number of people in the queue	2
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

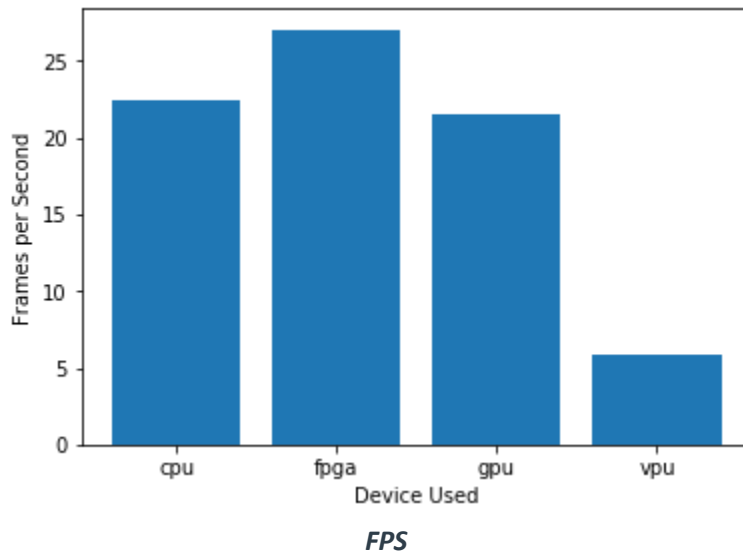
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



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

*Field Programmable Gate Arrays are best suited for this job as shown from the inference graph FPGA has the lowest inference time thereby improving the latency which was previously affecting the client. Greater number of frames are being processed compared to any other devices which results in more productivity for the machine. Furthermore, the client requires long term solution and hence **FPGA are the most suggested for the use.***

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

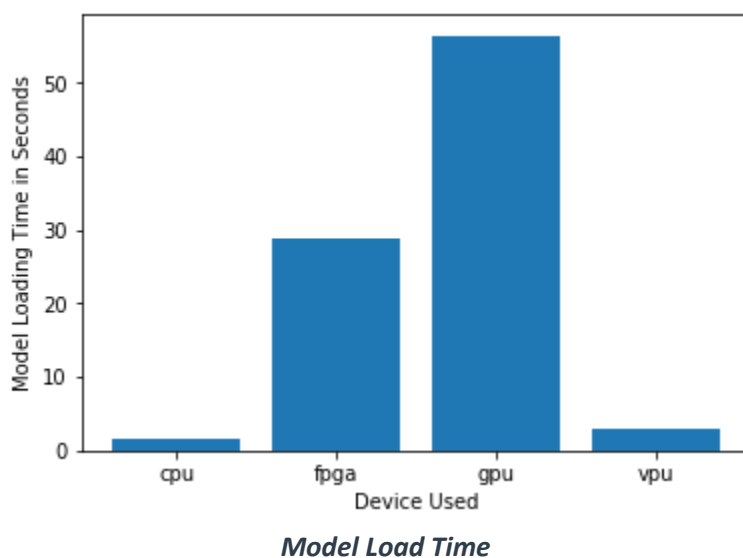
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Most of the queue counters already have new intel i7 processors and are being used to carry minimal tasks.	To enhance the capabilities of the work the installed CPU can easily handle the inference and loading of the models
<i>Does not want to purchase any new hardware.</i>	The client has already invested enough and has high end CPU.
<i>The client would like to save his electricity costs as well.</i>	<i>CPU can meet the hardware requirements at no extra cost.</i>

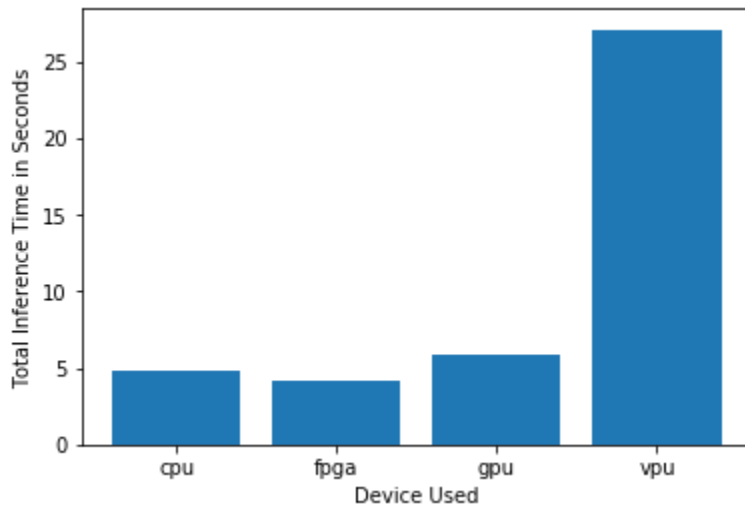
Queue Monitoring Requirements

Maximum number of people in the queue	2
Model precision chosen (FP32, FP16, or Int8)	FP32

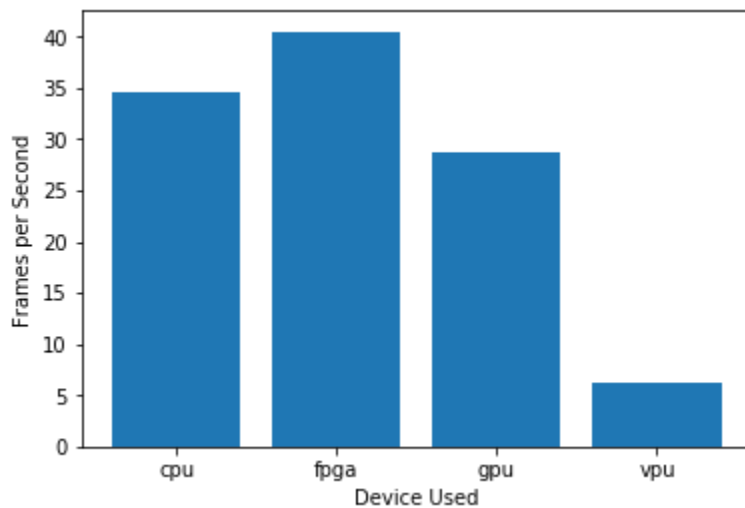
Test Results

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).





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

As seen from the graph, CPU generally has a lower loading time and lowest inference time compared to vpu or gpu which can potentially be helpful for the client. Furthermore, a good amount of frames are processed and also the client does not want to purchase any extra hardware hence **CPU** is the most suggested choice of hardware.

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

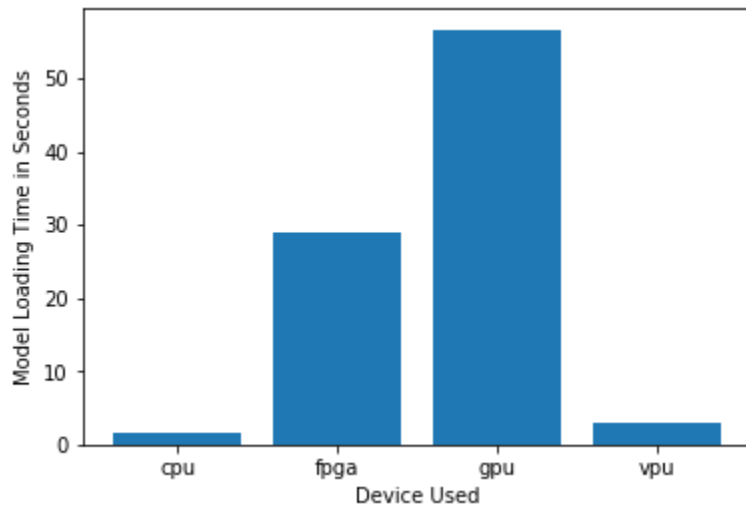
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
The client requires a tiny device to be connected to their CPU—and their budget is only about \$300 for each device.	VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
<i>No requirement to scale hardware.</i>	<i>Since the CPU are already capable an accelerator like VPU is enough in such situation.</i>
The client is required to save electricity and hardware costs.	<i>A VPU is designed to work on minimal power rate and hence is a better choice.</i>
-	-

Queue Monitoring Requirements

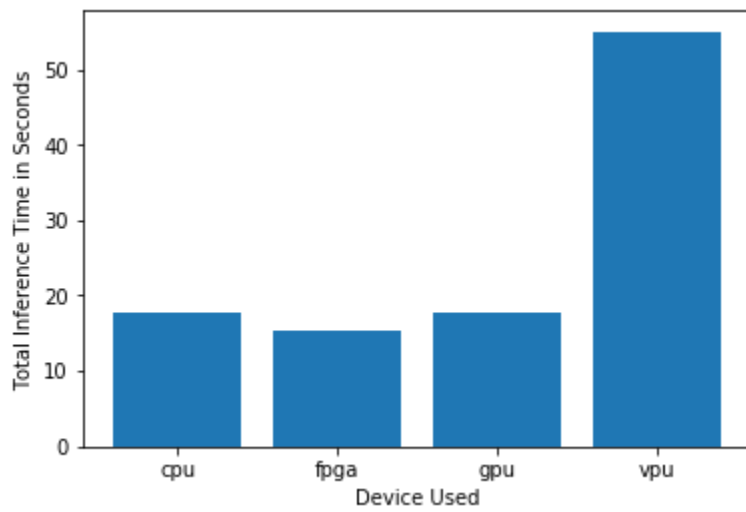
Maximum number of people in the queue	3
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

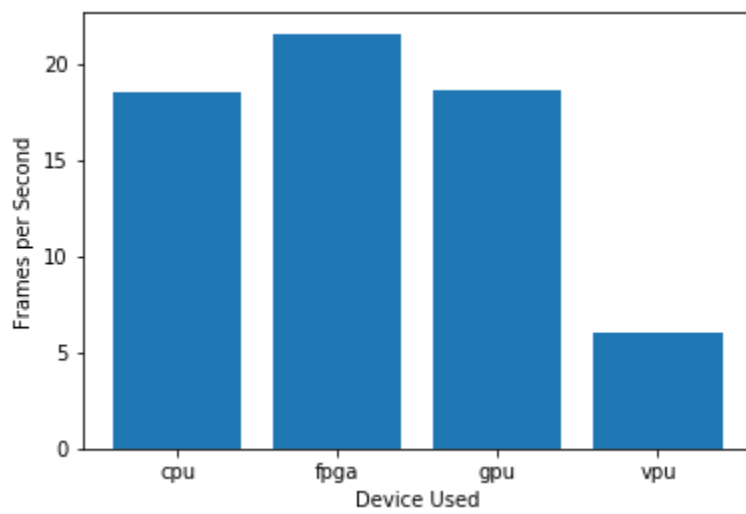
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

*VPU provides a very high inference time and inference on very low Frames per Second. A good alternative if client agrees to purchase would be a CPU/GPU but since power is another factor **VPU** is a better choice.*