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.
Image processing to be completed 5x/sec	*fast*
Flexible, wants to repurpose for chip quality	*reprogrammable*
Last 5-10 years	*can have additional fast operations added*

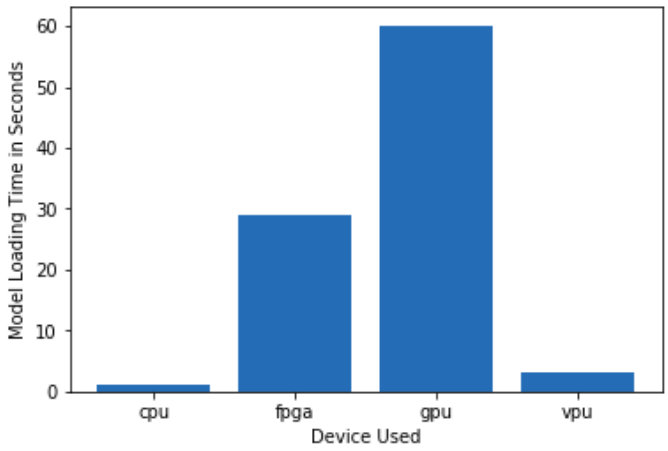
Queue Monitoring Requirements

Maximum number of people in the queue	*Unspecified, Thinking worst case*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

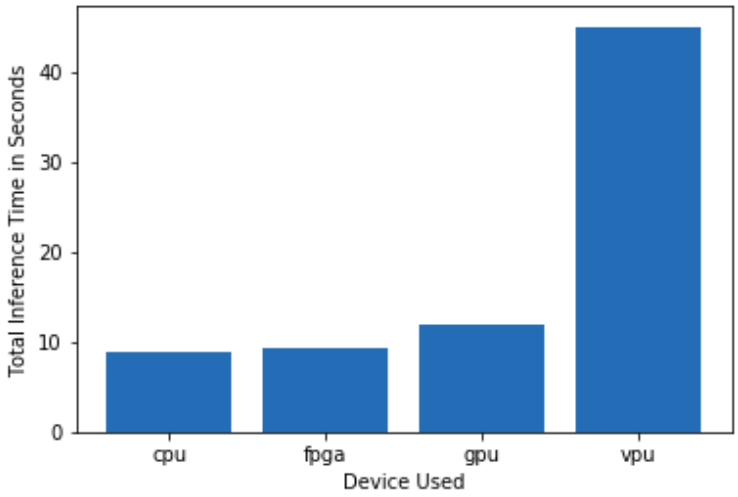
Test Results

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).

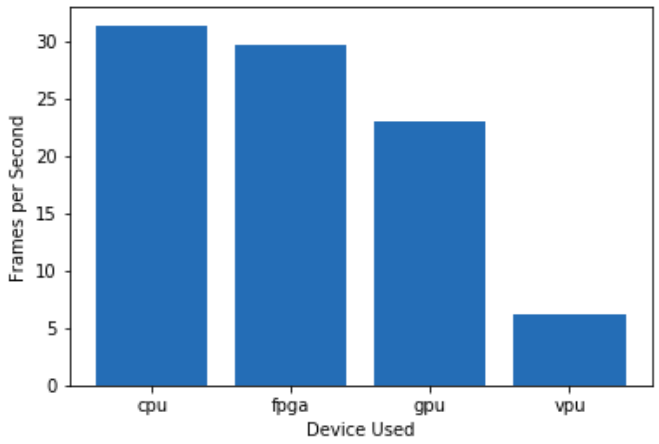
FP32



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
In my experiments FP16 wasn't much faster than FP32, I would expect a 2x speedup. Also the accelerators didn't have the amount of performance improvement I would expect. I think this would improve if executing in parallel or running batches. *My final recommendation remains an FPGA:* ** They have a long life-span* ** meets speed requirements (mostly thanks to the CPU)* ** reprogrammable*

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

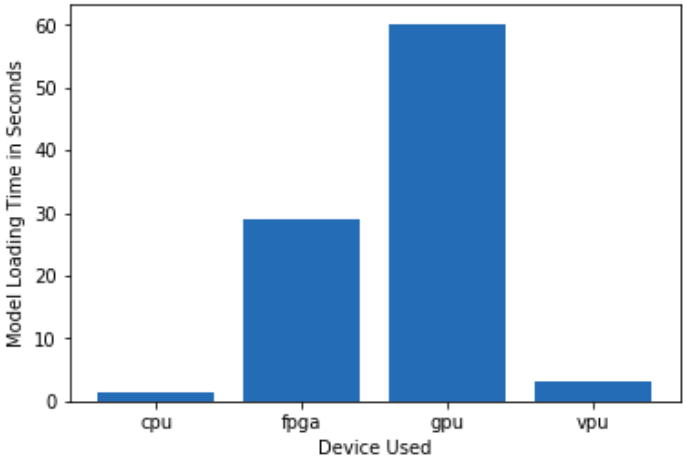
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
Low power (save on electric bill)	*Already powered up for cash registers*
Cheap – avoid additional hardware, has i7 CPUs already	*Already has the hardware*
Doesn't need to be fast - Average wait time of 230-400 secs (2-5 per queue), <20 seconds	*Doesn't need to run several fps for improvement*

Queue Monitoring Requirements

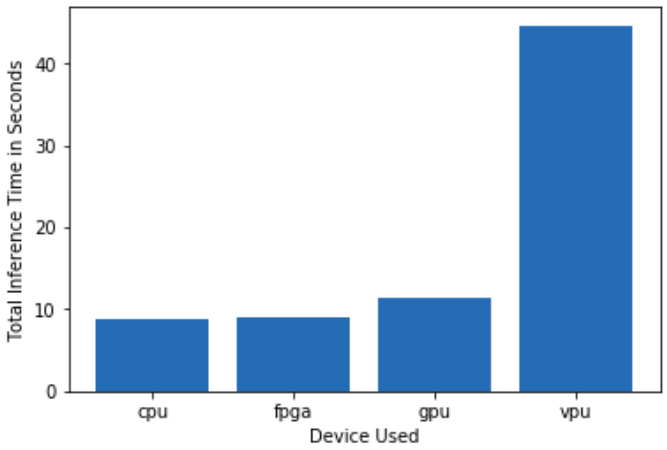
Maximum number of people in the queue	*2 on weekdays 5 on weekends*
Model precision chosen (FP32, FP16, or Int8)	*FP32*

Test Results

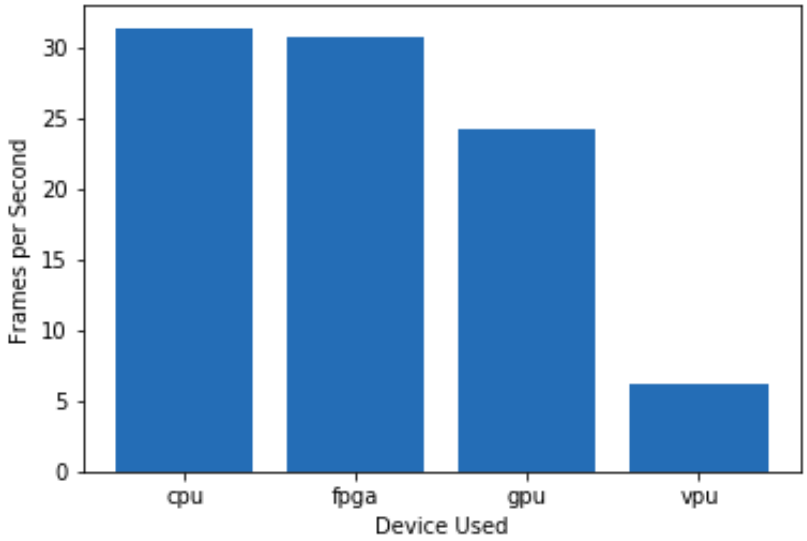
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
The model didn’t change so my results look about the same. *CPU is my final recommendation:* ** Doesn’t need to buy new hardware* ** Doesn’t need to be crazy fast, CPU is plenty fast in my tests* ** Low additional power - power is already being consumed by 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? (CPU / IGPU / VPU / FPGA)

VPU / NCS2

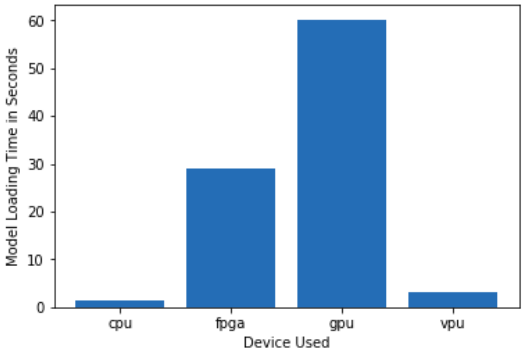
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Example requirement: The client requires a tiny device to be connected to their CPU—and their budget is only about $100 for each device.	*Example explanation:* VPU or NCS2 is only about 27.40 mm in size and would fit in the price range.
$300/machine	*NCS2 would fit in the price range*
Save on power	*VPUs are low power*
Can't use current cpus	*Pairs well with current setting*
Train every 2 minutes	*Fast enough to do several frames between trains*

Queue Monitoring Requirements

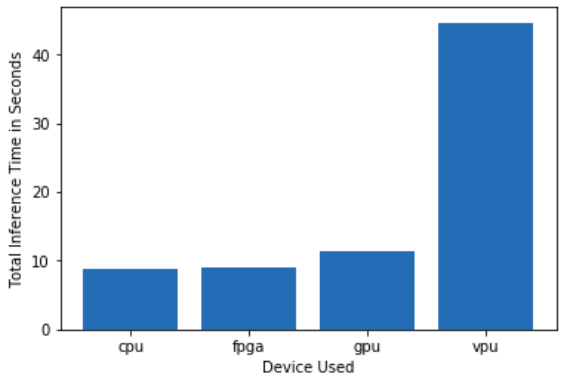
Maximum number of people in the queue	*7-15*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

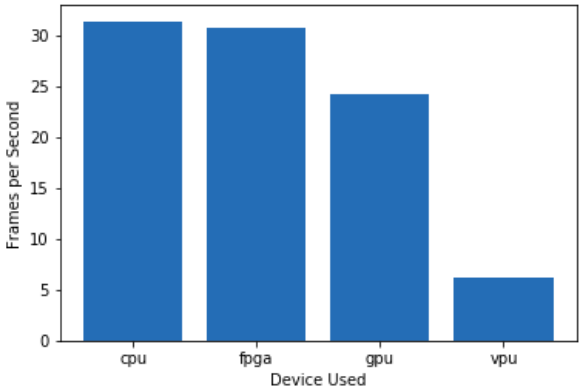
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
Still going with an NCS2, because: ** it is low power* ** inexpensive (less than $300)* ** Doesn’t use current CPU* ** fast enough for use case*