Choose the Right Hardware

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

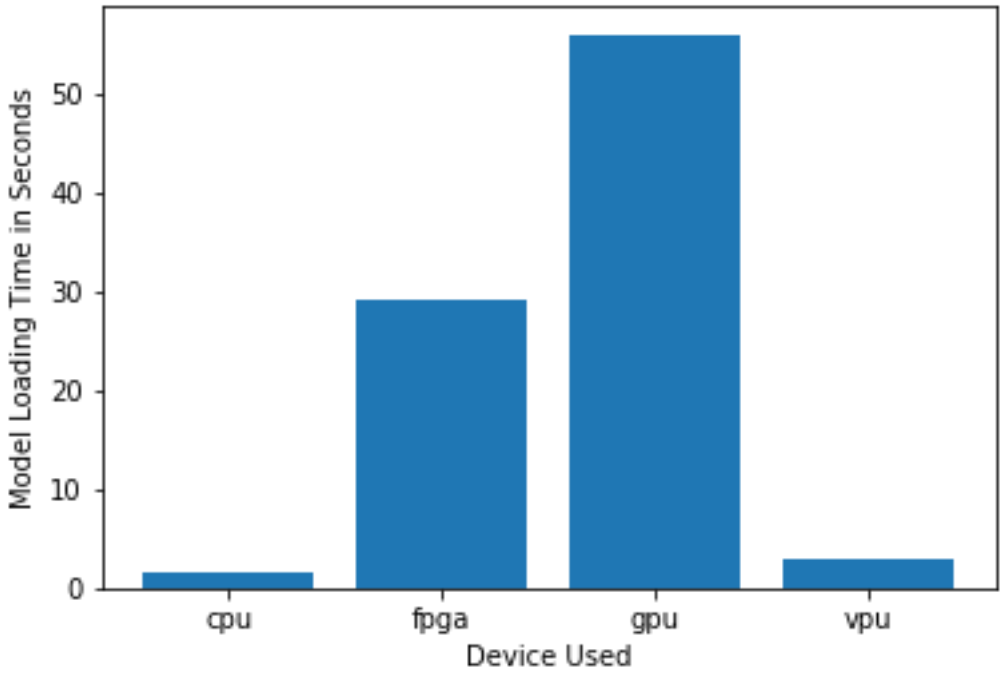
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Client wants to flexible design that can be reprogrammed	FPGAs can easily reprogrammed to adopt new design
System should last for 5-10 years	FPGAs are designed to work for long period of time
Client want to make inference 5 times in a second	FPGAs can make inference very fast

Queue Monitoring Requirements

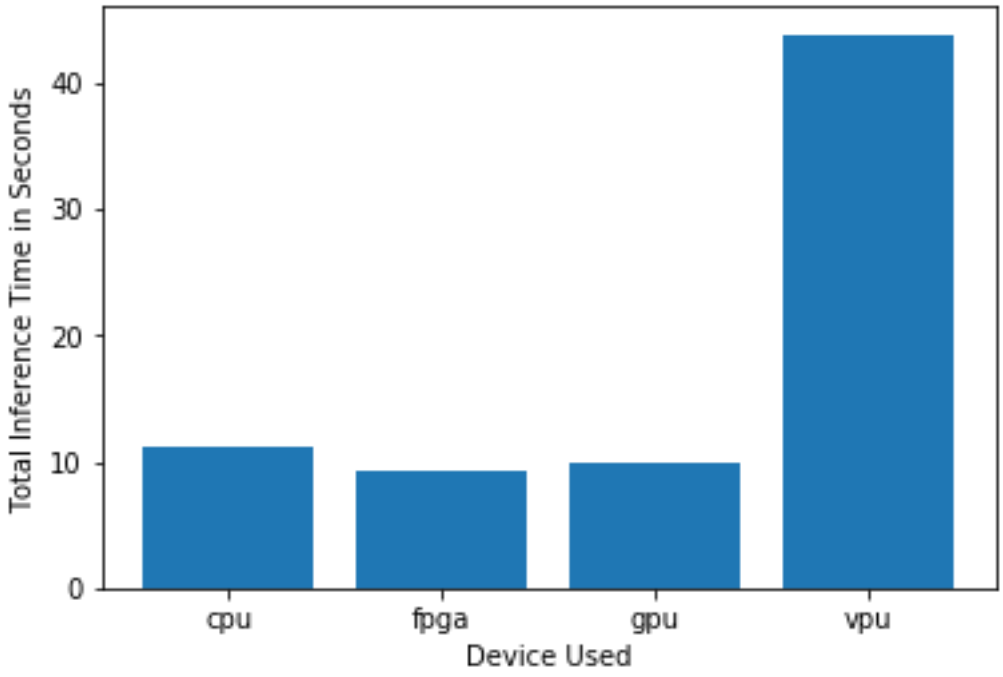
Maximum number of people in the queue	2
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

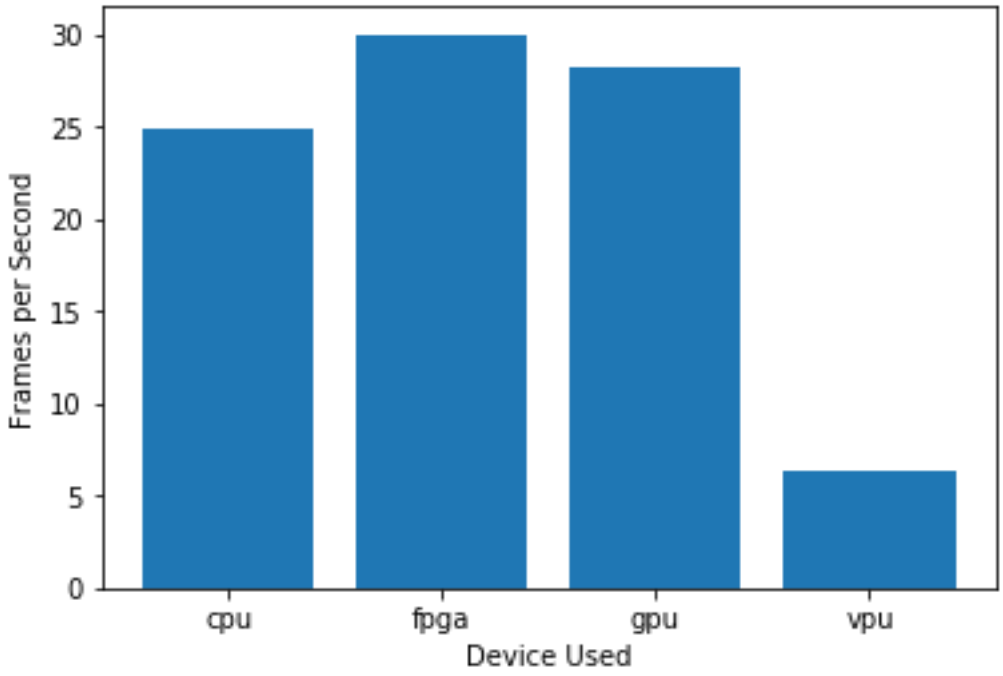
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



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
Keeping in mind client requirements and output results got from running inference on CPU, GPU, VPU and FPGA on Intel Dev Cloud, FPGA satisfy client requirements better than other devices. *FPGA have long life span, so they exactly fit user need for 5-10 years. It can be reprogrammed when a new design came out. From output result, we can also find FGPA has highest Frame per Second compared to other devices*

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

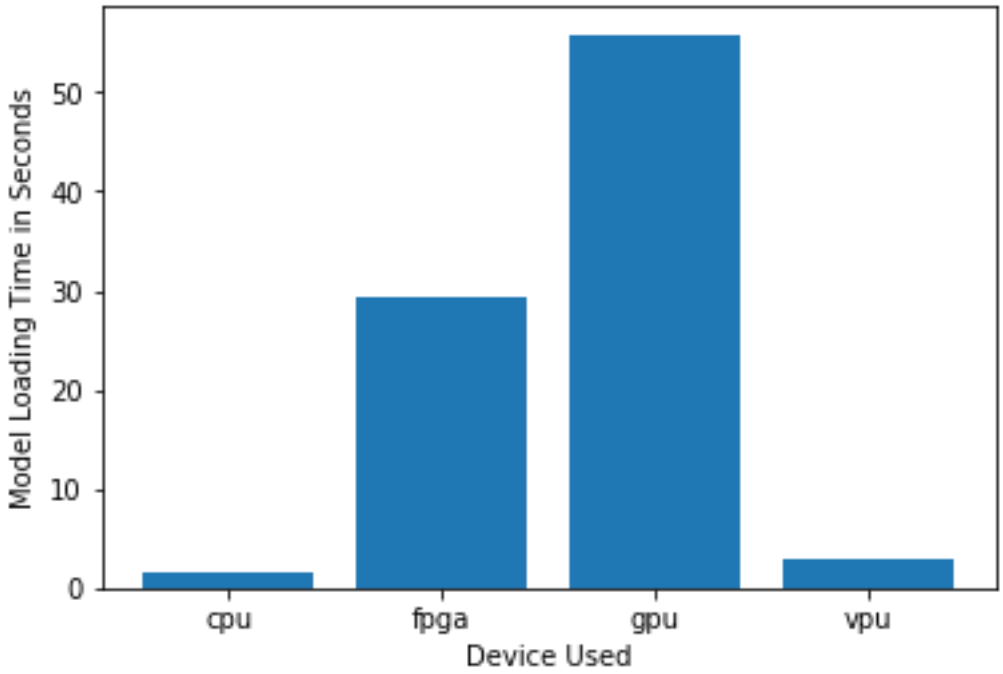
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Client does not have too much amount to invest	As Core i7 processors are not performing heavy computation task, they can be used for inference because they also have IGPU built in
Limited power	As not additional device is being used, so no more energy require. IGPU can perform task with little energy consumption

Queue Monitoring Requirements

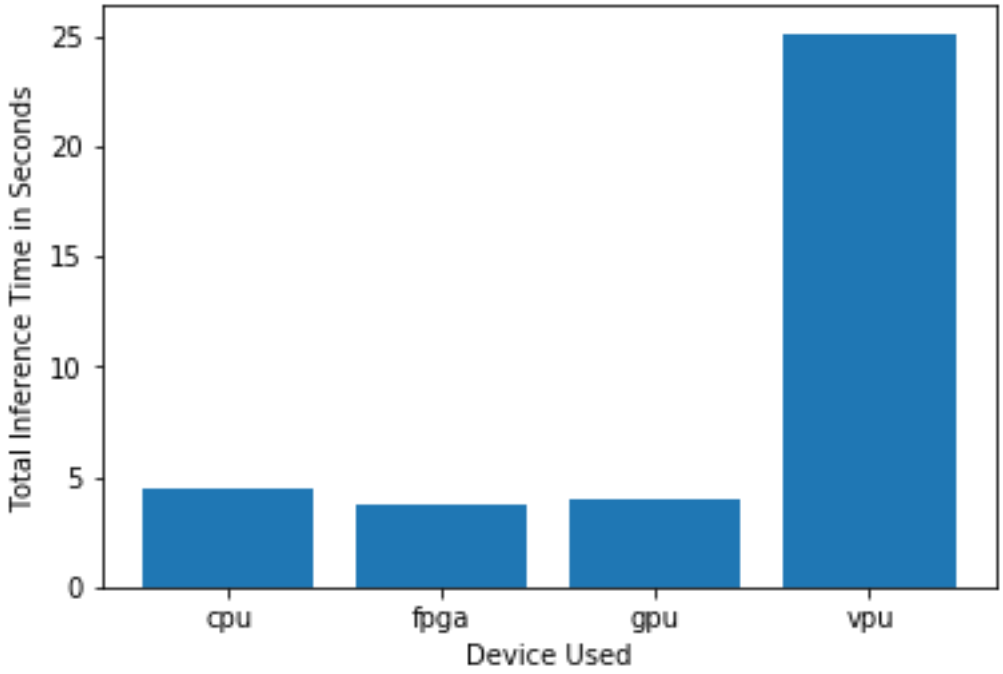
Maximum number of people in the queue	*2*
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

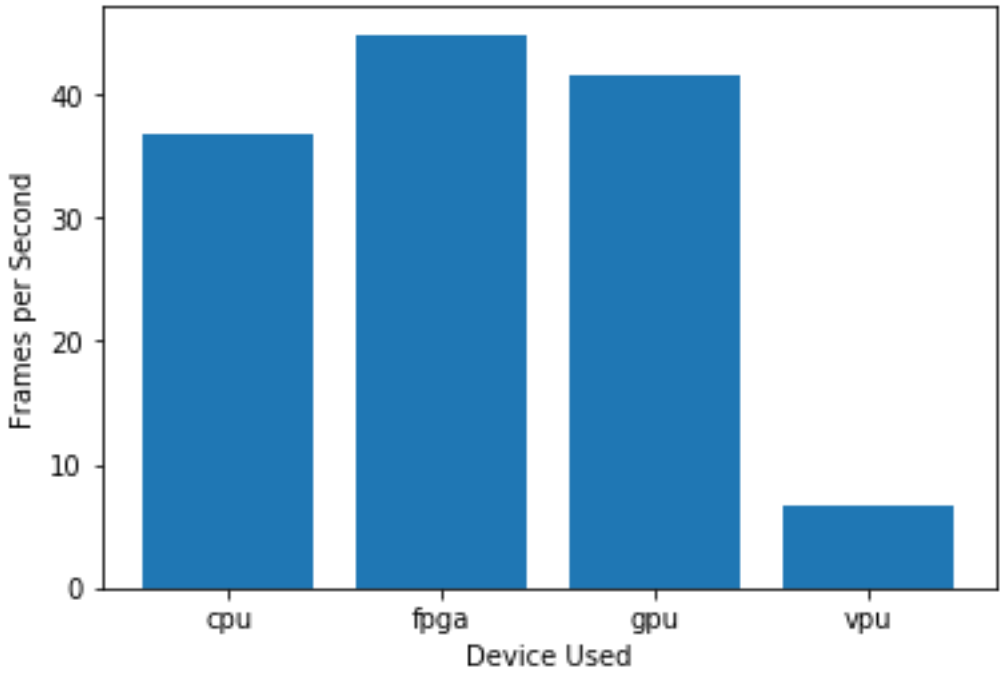
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



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
Comparing result got from running inference on all 4 devices on Intel Cloud Dev, we can find although IGPU have high loading time, but has low inference time and high Frame per Second. So it can be used for making inference but it does not add additional hardware and energy cost.

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

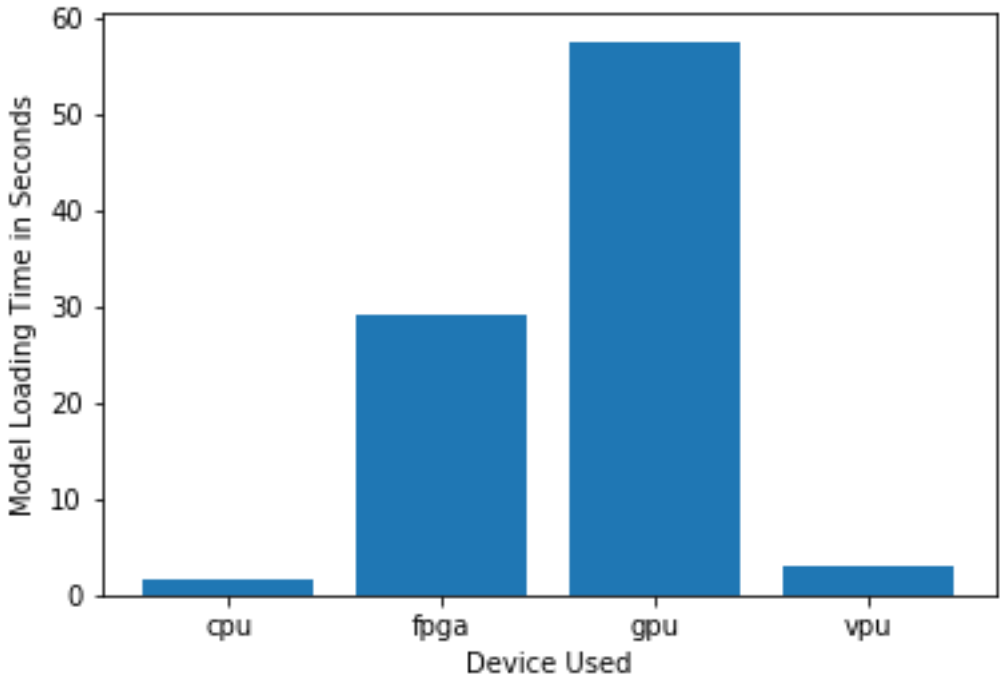
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Only 300$ for each device	Intel Compute stick cost less than $100
Power requirement	It need very little power ~1W

Queue Monitoring Requirements

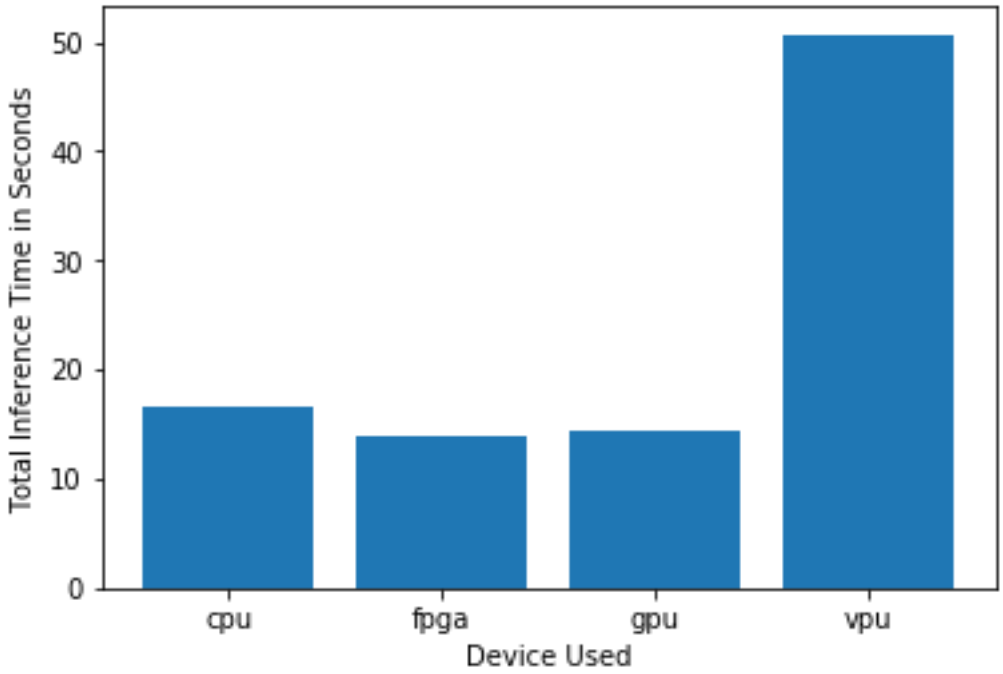
Maximum number of people in the queue	*7*
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

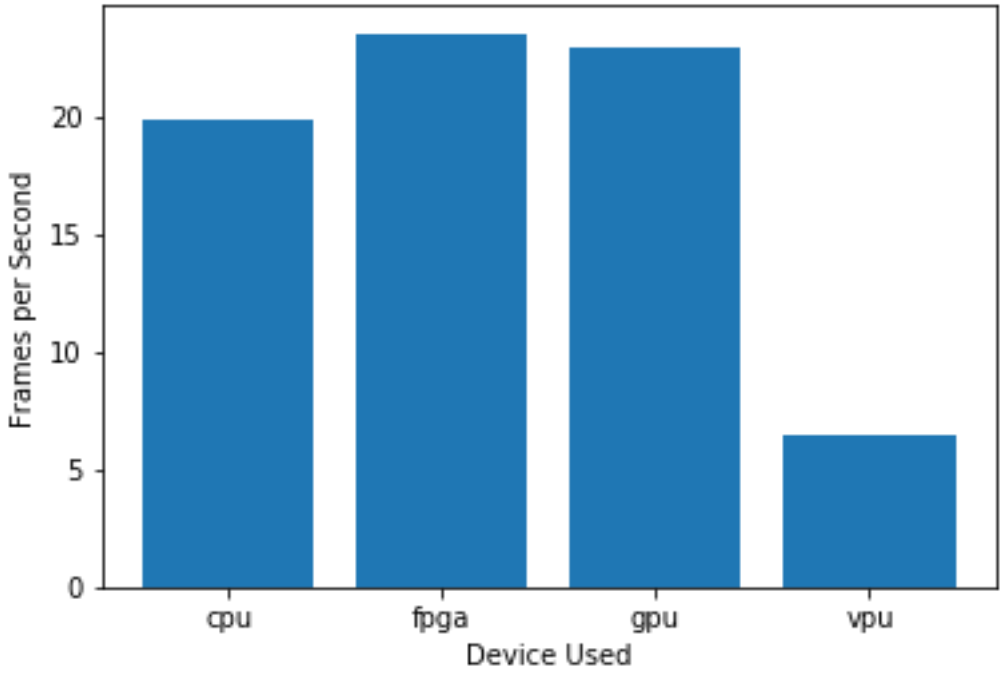
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
Comparing graphs we can find that VPU has highest inference time and lowest frame per second. But keeping in view client requirement it is best available option to meet client budget and power requirement. Upgrading CPU + IGPU seem to better option than VPU, but it need more budget and IGPU will consume more power than Intel NCS2.