

Choose the Right Hardware

Proposed By Ashish Karn

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

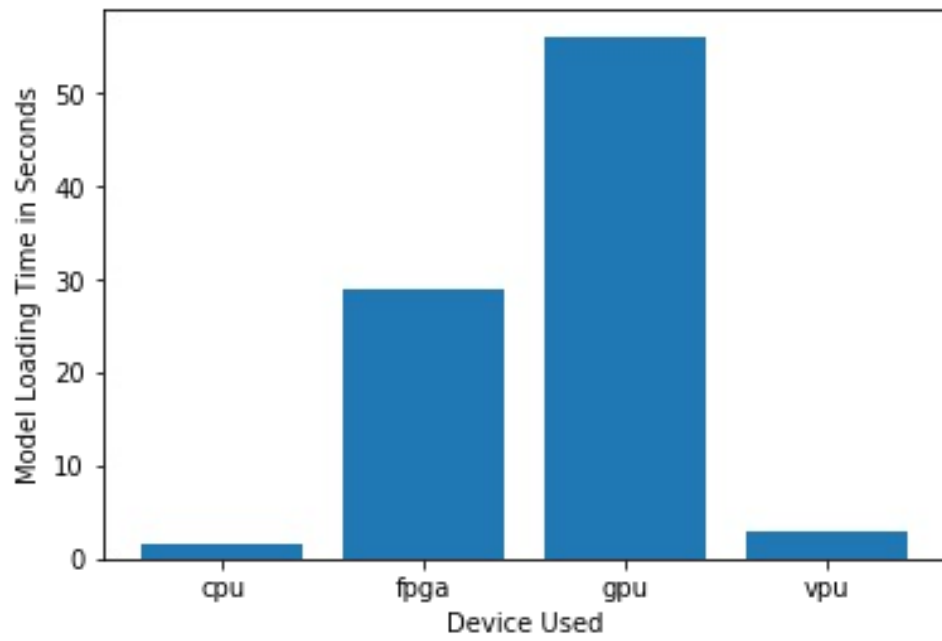
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Good Durability (5-10 years)	FPGA's last that long, we can trust quality and durability with it.
30-35 FPS with minimum 5 images per second processing speed	FPGA's outperform other devices and will be able to meet these requirements
Cost is not an issue and want permanent device	FPGA's even if expensive, won't be a problem in this scenario
There are multiple chip design and new designs are created regularly. Flexibility of the system is needed in present as well as future.	FPGA fulfills the requirement of flexibility as it can be designed as required.

Queue Monitoring Requirements

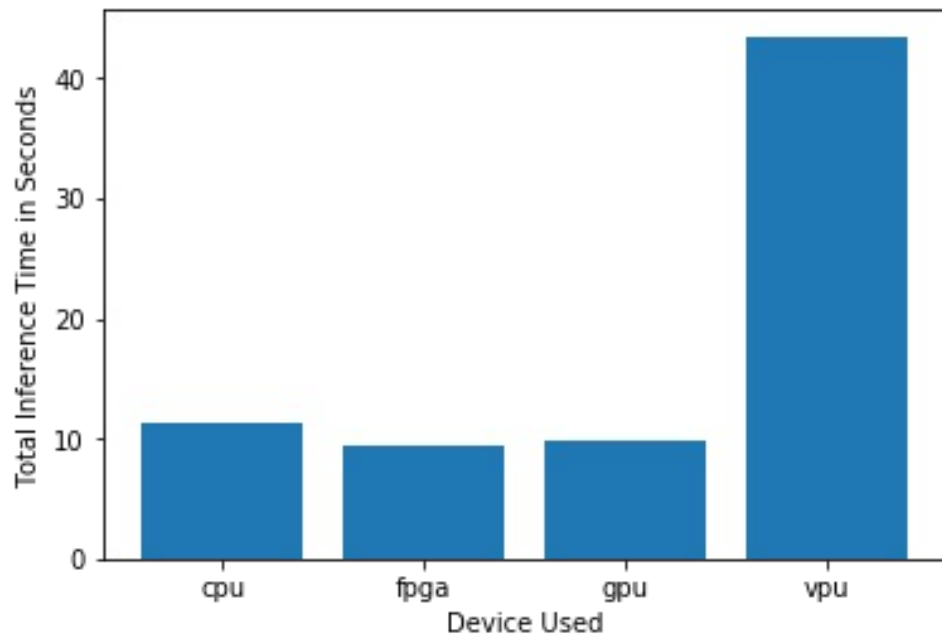
Maximum number of people in the queue	2
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

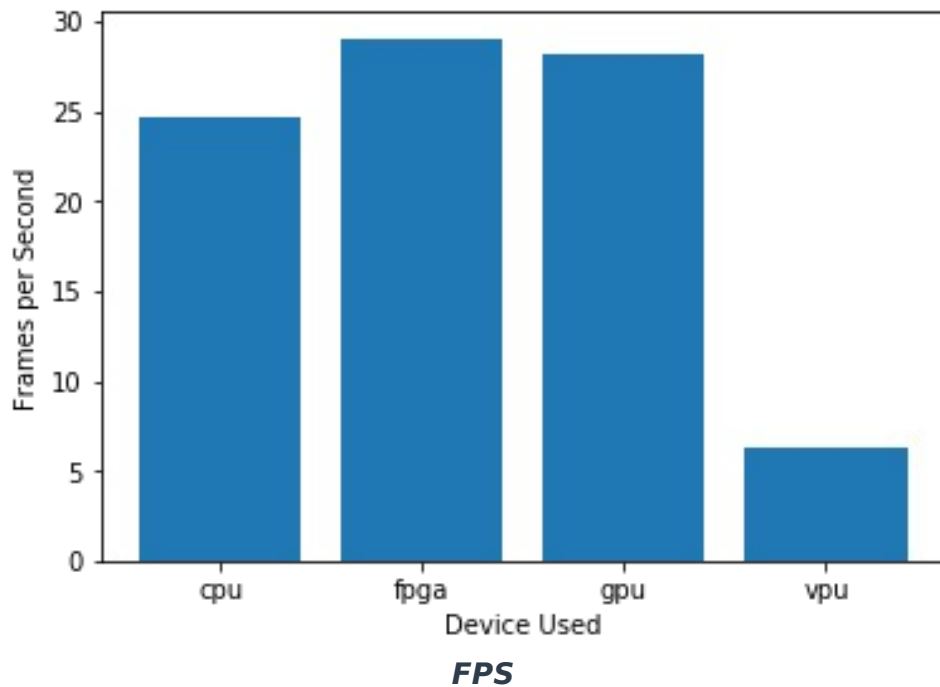
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

As seen in the graphs above, FPGA shows the best performance. It meets the client's 30-35 FPS requirements. Even if it is a little slow in loading the model, but least inference time and highest FPS cover for the loading time. The client has a 24-hour prediction running requirement which is easily satisfied by FPGA. Thus **FPGA is the good choice** for this client.

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

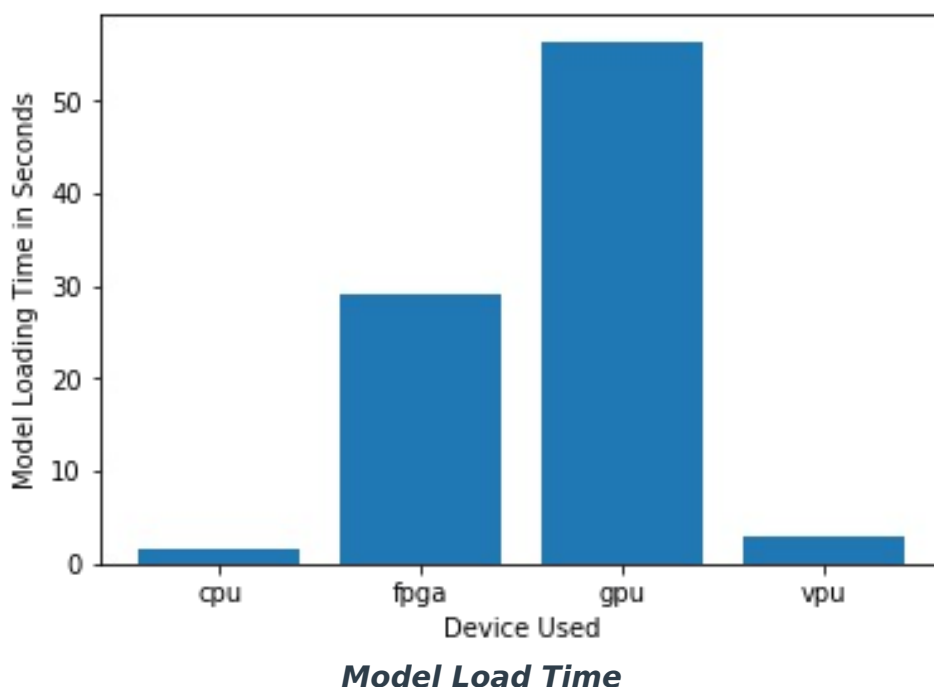
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
The client requires a tiny device to be connected to their CPU and their budget is less and don't invest in Hardware.	A good configuration of cpu i.e. i7 core processors available.
Need during rush hours, good FPS and processing needed	Since computers are already used, running the application on that itself at checkout would be cost and compute efficient.
The client would like to save as much as possible on electric bill.	<i>CPU's consume less power in compare to other. Hence is a good way to save electric bill.</i>

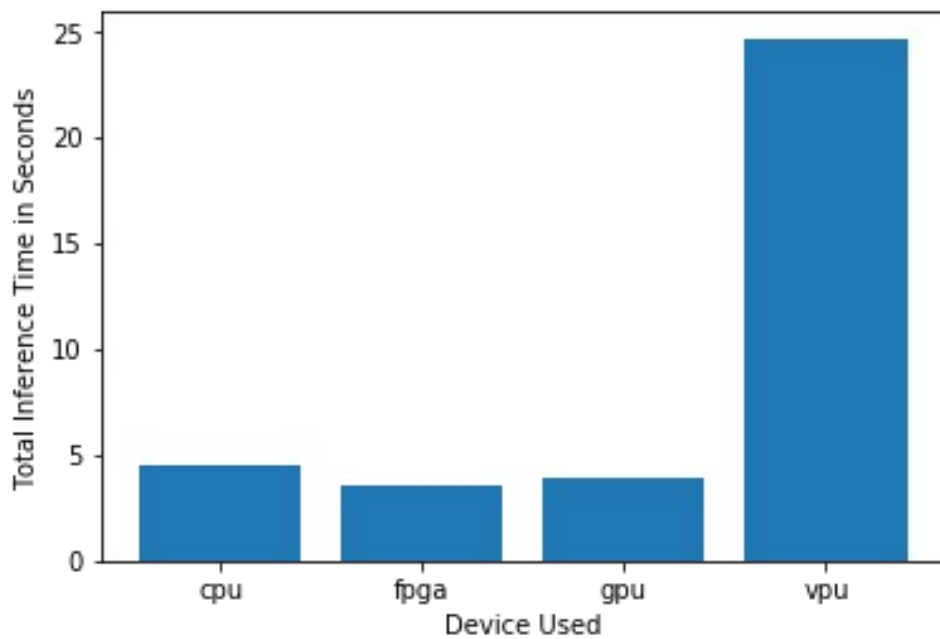
Queue Monitoring Requirements

Maximum number of people in the queue	2-5
Model precision chosen (FP32, FP16, or Int8)	FP16

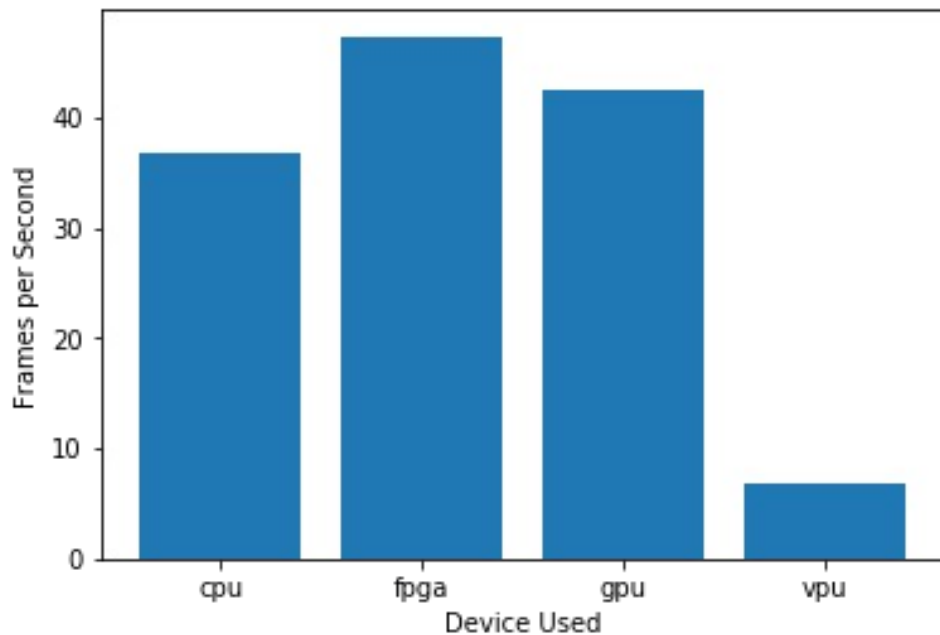
Test Results

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).





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 you can see, the CPU does give a good performance and even model load time is least, even the inference time is not bad. No extra hardware needed so this makes a good case. I would also like to mention that IGPUs are a good option too. Though there will be some initial hardware expenses and model load time is highest but that is only a one-time thing, after that GPUs perform really well and even their FPS is quite high, also they have configurable power consumption so electric bills could be saved. So it's a tough choice between CPUs and GPUs if a little initial investment is affordable, I'd recommend GPUs as they can save money on electric bills which are a monthly affair in most places and compared to an initial investment, in the long run, GPUs will be profitable. If the client is not ready to spend even a little on hardware then CPUs would work too. Thus **CPU is the best choice** for the client.

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

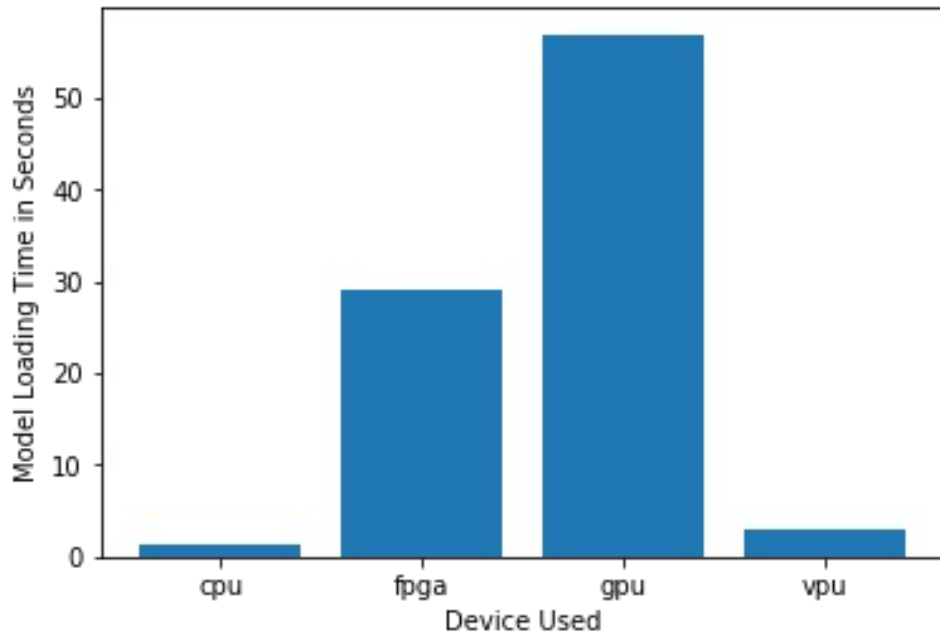
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
The CPUs in these machines are currently being used to process and view CCTV footage for security purposes and no significant additional processing power is available to run inference.	EA VPU or NCS 2 can be plugged in a USB port and has a convenient plug and play kind of performance. This particularly is favorable as no more additional processing power is available to run inference.
The client's budget allows for a maximum of \$300 per machine.	A VPU or an NCS 2 stick costs almost \$100 as opposed to the comparatively higher costs of FPGA, CPU, or GPU.
The client would like to save as much as possible both on hardware and future power requirements.	A VPU or NCS 2 stick is designed to run on very low power. The NCS 2 can run on just 1 W of power adhering to the client's needs

Queue Monitoring Requirements

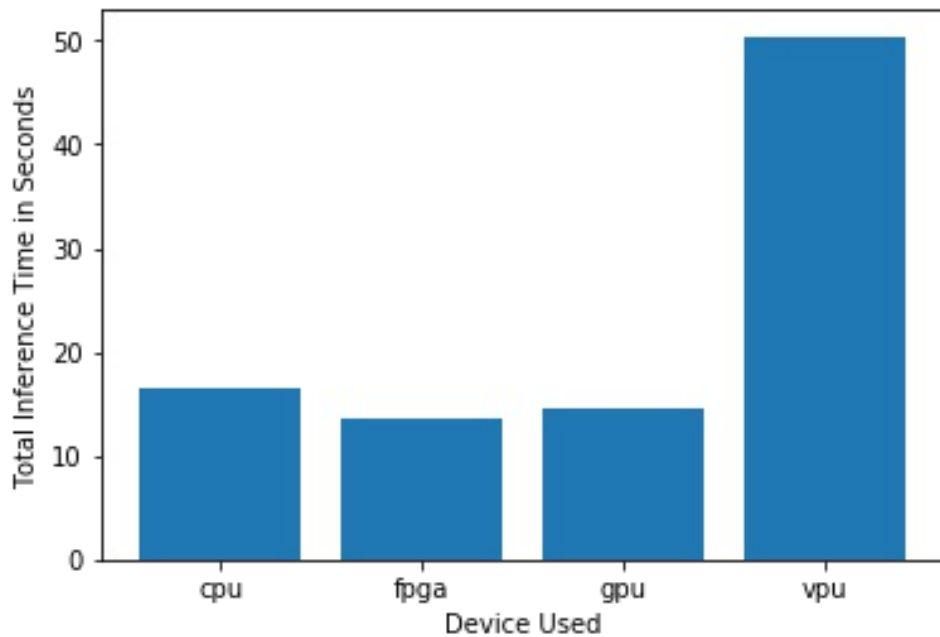
Maximum number of people in the queue	7-15
Model precision chosen (FP32, FP16, or	FP16

Test Results

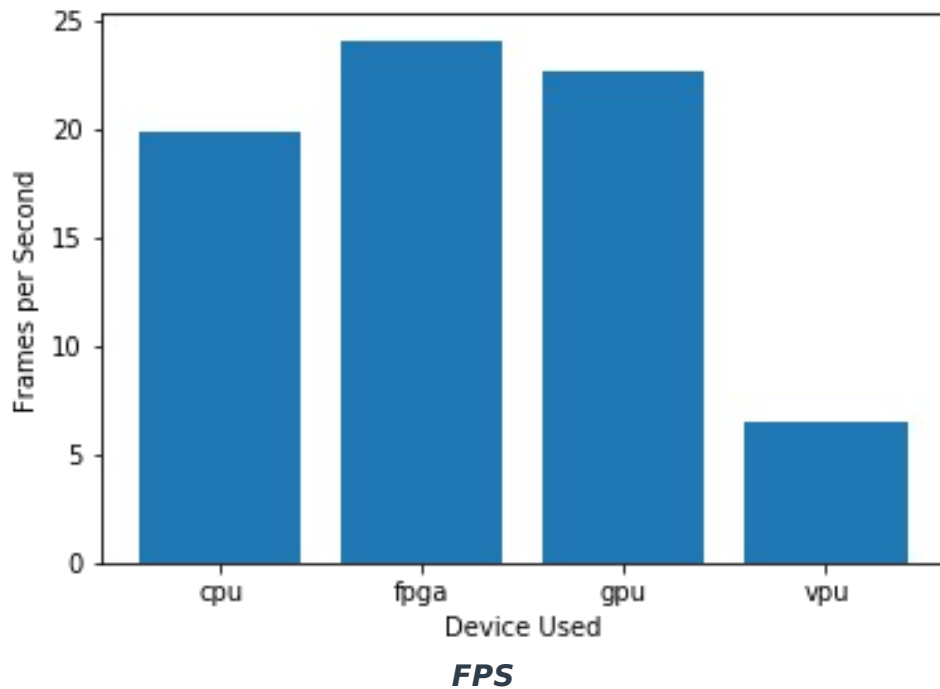
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

The inference time of VPU is high and fps are comparatively lesser than other devices but It will still meet the requirements of the client to handle 7 to 15 people in a queue without spending money on the new hardware for each computer. A simple plugin VPU is the best choice for this scenario. Additional VPUs, that is more than 1VPU can be purchased in the given budget. For example, Intel NCS2 costs around \$70 to \$100 and at least 3 can be bought in the given budget range. Thus **VPU is the good choice** for the client.