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

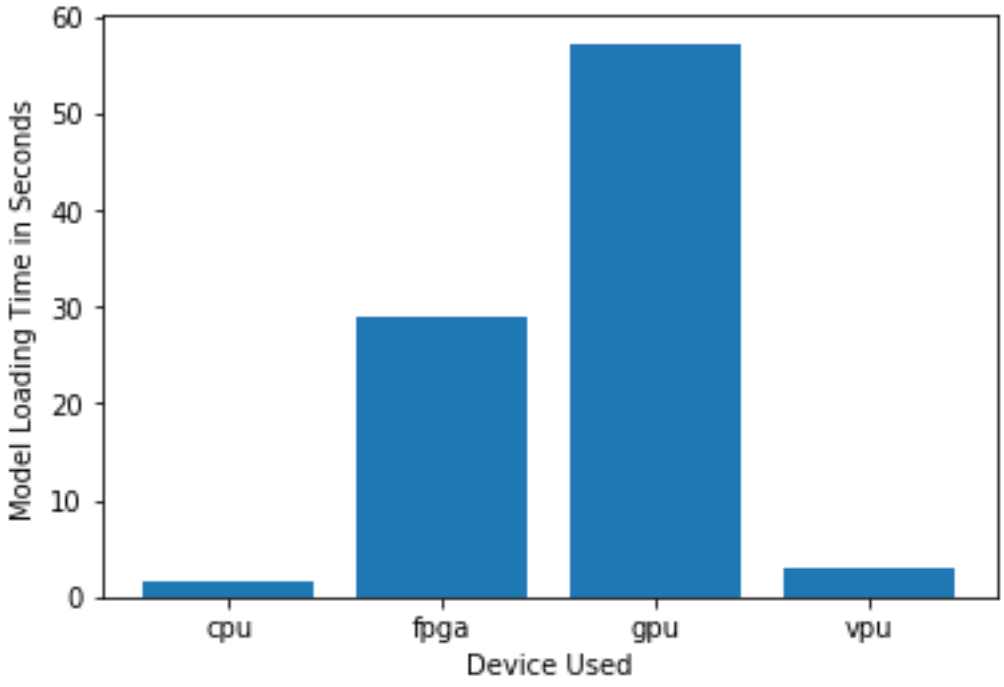
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Once this productivity problem has been addressed, Mr. Vishwas would like to be able to repurpose the system to address a second issue	FPGA's are reprogrammable and are able to easily handle a 5 FPS inference speed
has plenty of revenue to install a quality system	Has the ability to make a proper investment
the system would need to be able to run inference on the video stream very quickly. Additionally, because there are multiple chip designs—and new designs are created regularly—the system would also need to be flexible so that it can be reprogrammed and optimized to quickly detect flaws in different chip designs.	FPGAs can perform inference fast. Also are flexible, as they can be reprogrammed to newer networks
significant investment and they would ideally like it to last for at least 5-10 years.	FPGA's lifespan can handle that requirement

Queue Monitoring Requirements

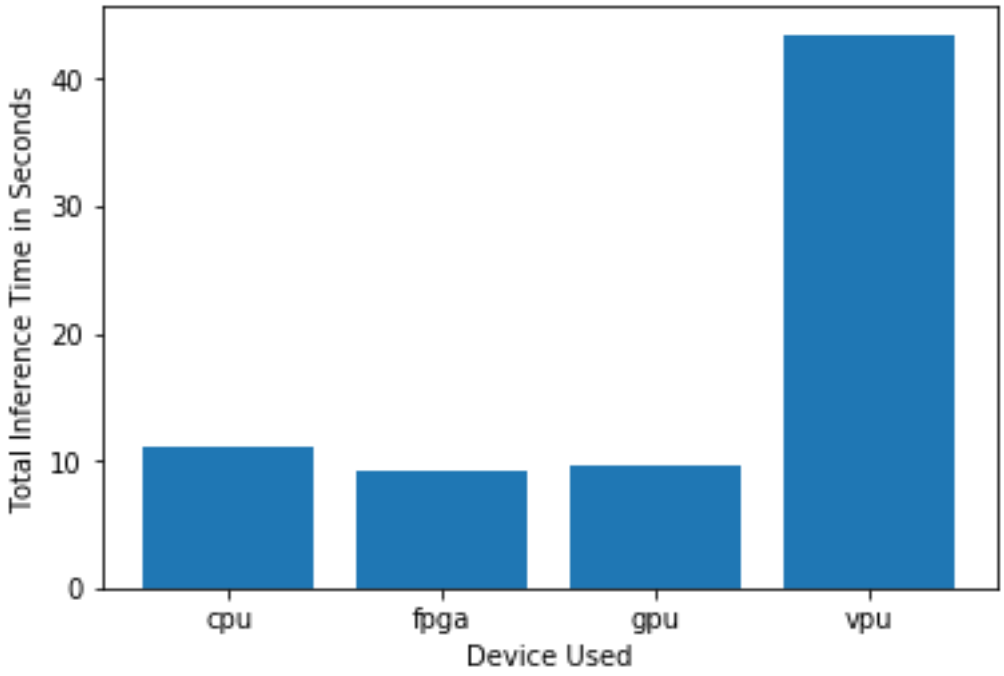
Maximum number of people in the queue	2 [shift transition]
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

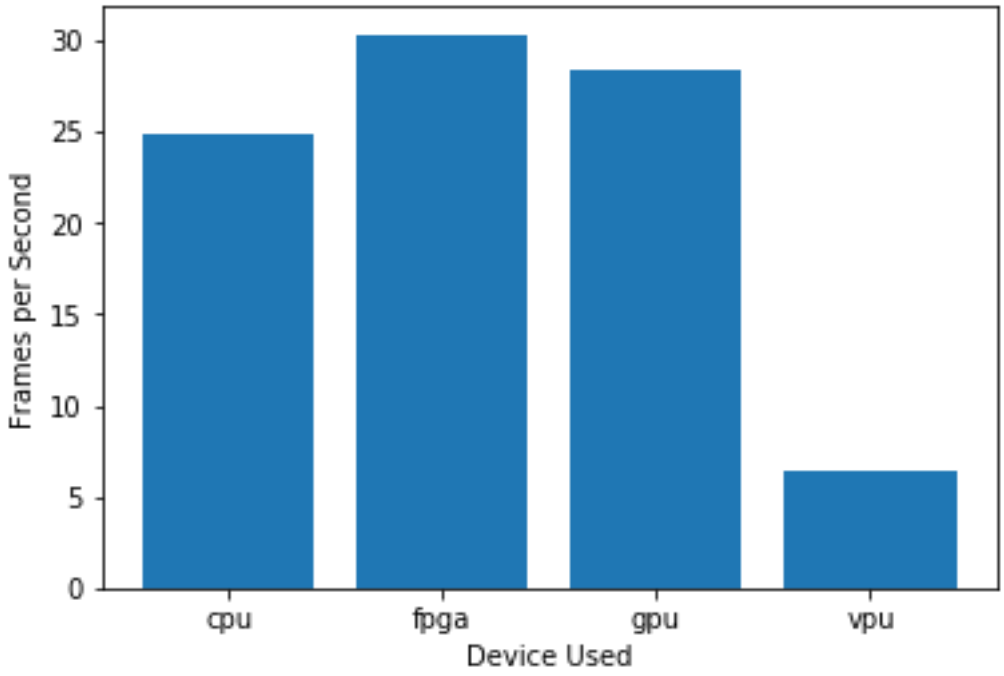
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
FPGAs (with the necessary CPU support) seem to be the best choice. Except the one-time cost of loading the model, the have the lowest inference time and the highest FPS.

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 with IGPU support if needed

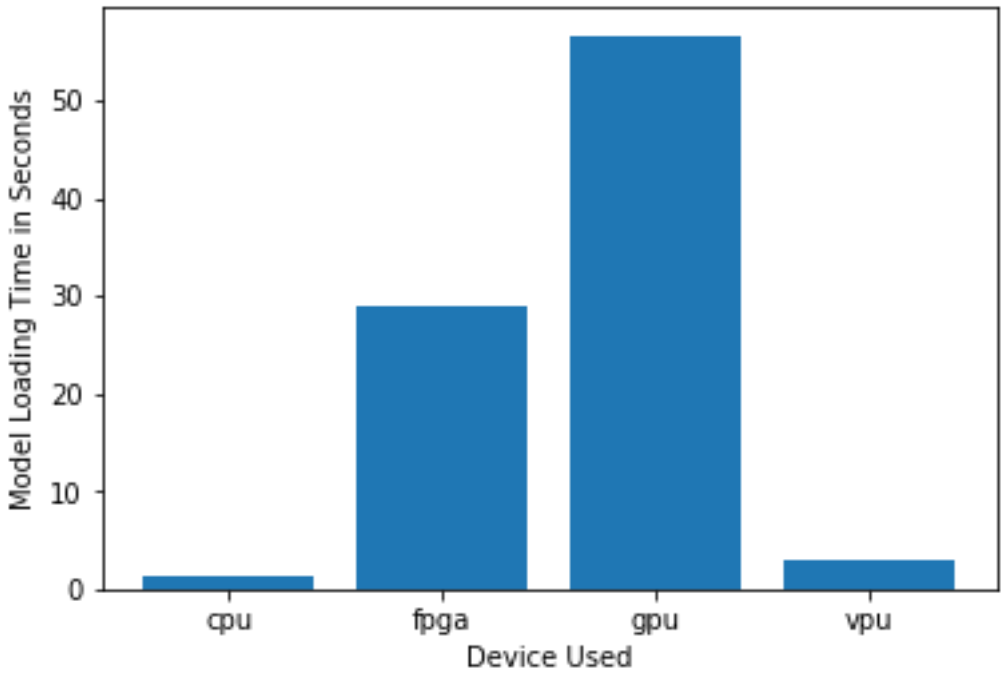
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Does not need to be fast (average wait time of 230 seconds at the checkout counter)	No need of high FPS
Low cost in terms of extra infrastructure / equipment	Using what Mr. Lin already possesses does not add any extra cost
Save on other extra costs (such as electric bill)	Not adding any new equipment will not add to the current costs (apart the extra power the CPU - with or without the IGPU - will require dues to its utilization)

Queue Monitoring Requirements

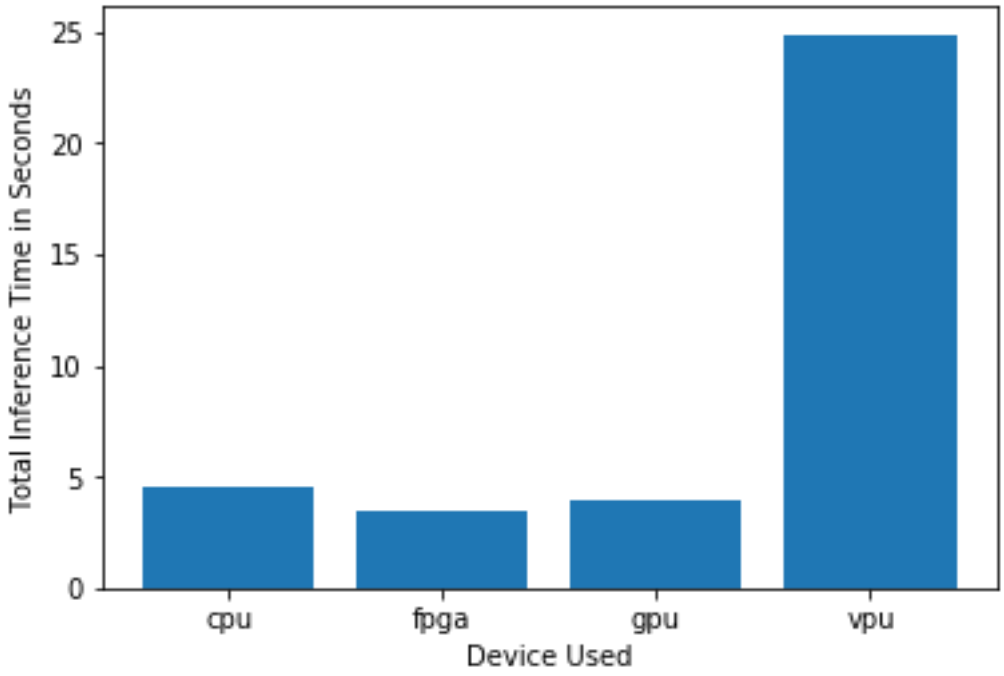
Maximum number of people in the queue	Over 5 at peak hours - ideal 3
Model precision chosen (FP32, FP16, or Int8)	FP32 in case of CPU only, FP16 if we use the IGPU

Test Results

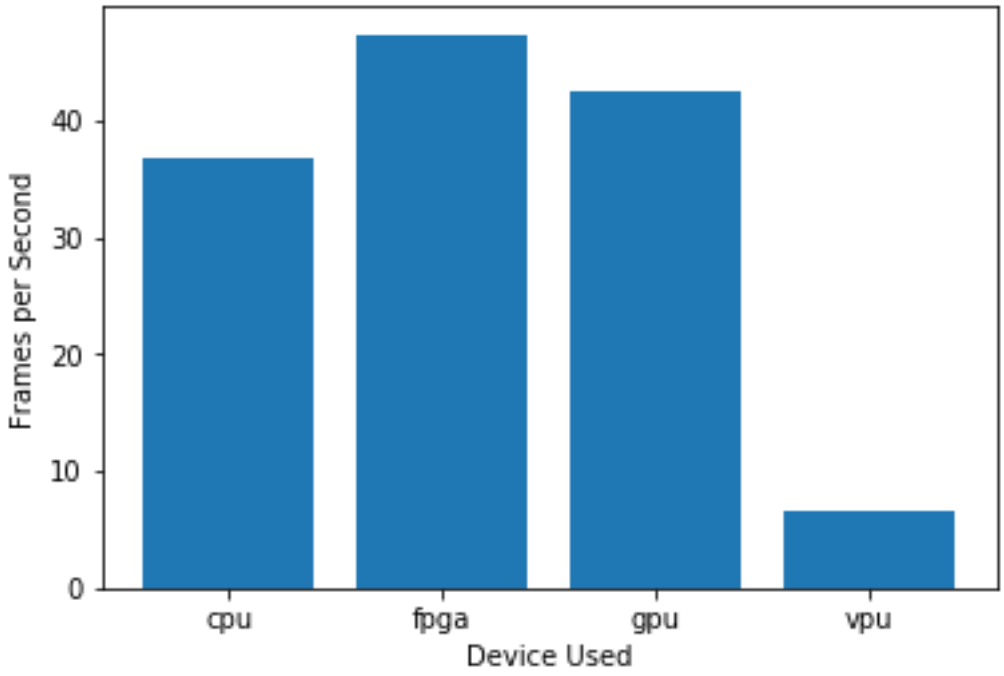
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
GPU, except the one time cost of loading the model, performs better in terms of FPS and inference time, and we do not add any 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)
IGPU – VPU [NCS2]

Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Current PCs are highly utilized	Adding a VPU to run the inference will not add up to current CPU workload
Each PC is responsible for 7 CCTVs	The NCS2 stick has USB3.1 which could handle the I/O bandwidth. In case of USB2.0, the inference will be slower (I/O throttling), but adding two VPUs could hide the extra usb cost
Keep cost and power requirements low (max budget of $300)	Adding a NCS2 is in between the client's budget, and their power requirements are low

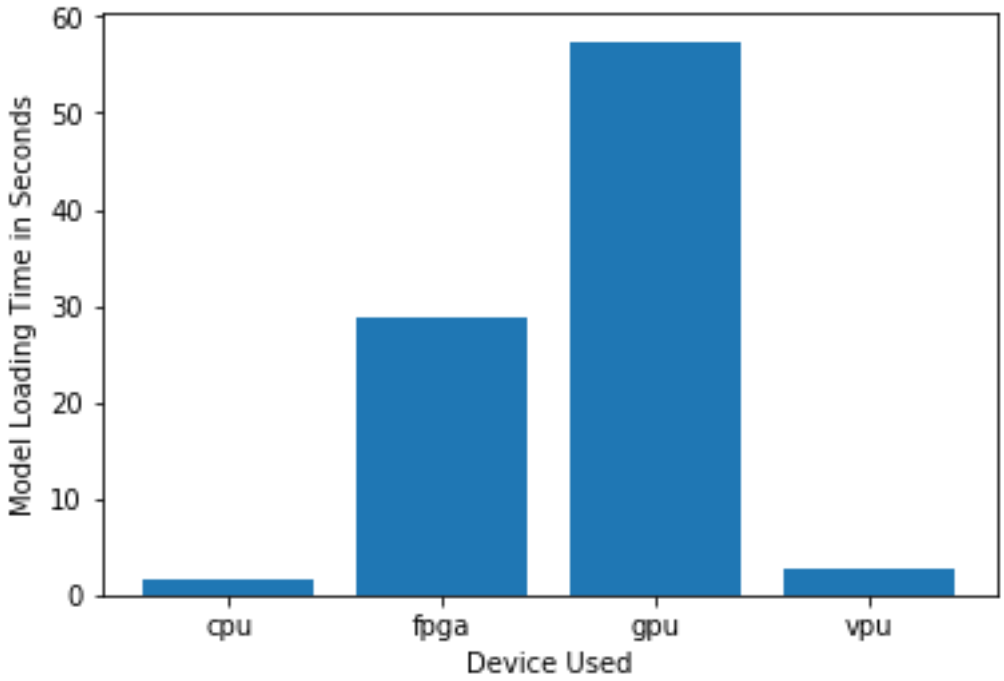
Queue Monitoring Requirements

Maximum number of people in the queue	Over 15 at peak hours - ideal 7

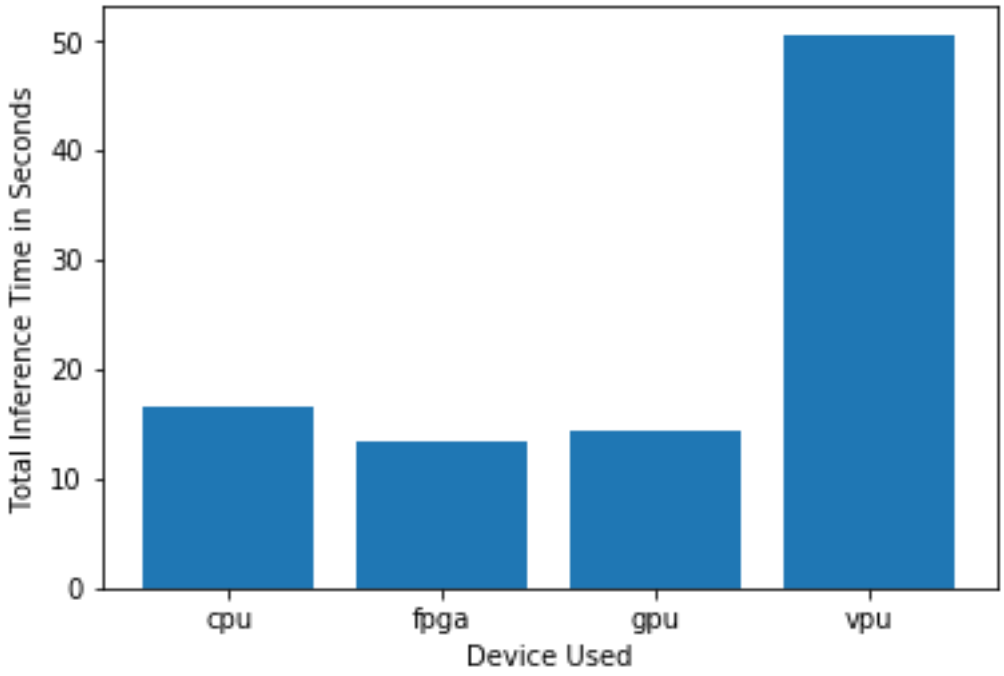
Model precision chosen (FP32, FP16, or Int8)	FP16

Test Results

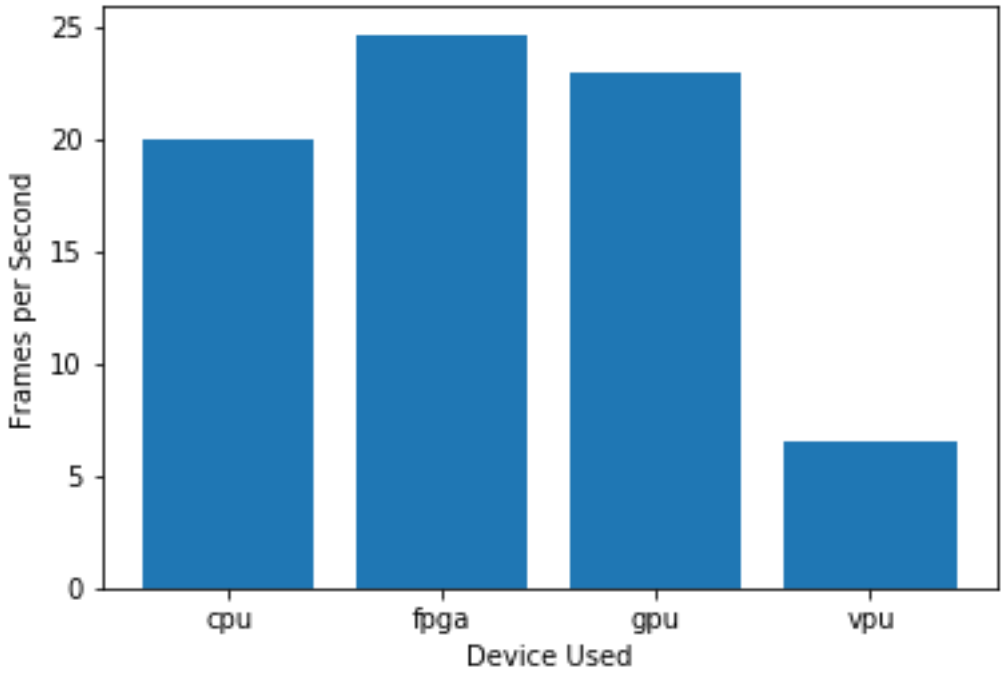
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
It seems that VPU is fast enough for such a case scenario, is between the client's budget, it does not add up to current CPU usage (which is high) and has low power consumption.