Choose the Right Hardware

Proposal

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)
Field Programmable Gate Arrays (FPGAs)

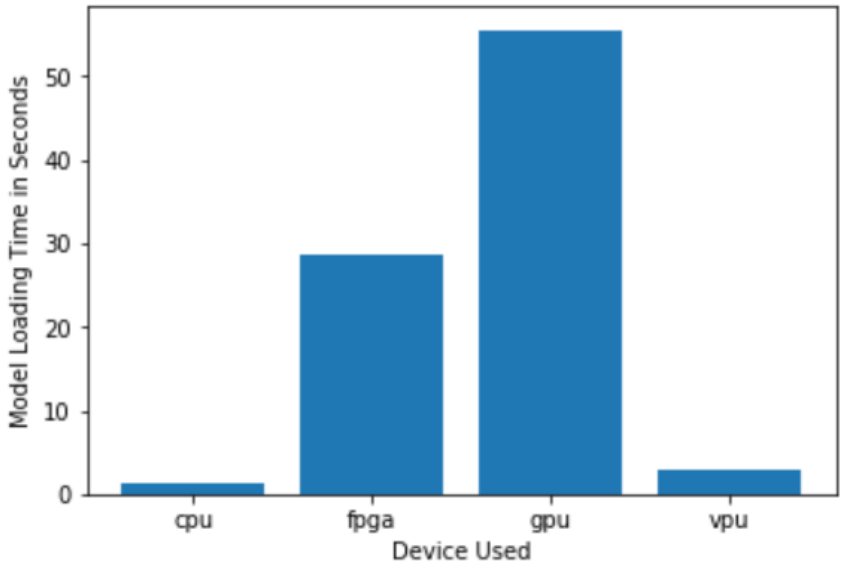
Requirement Observed (Include at least two.)	**How does the chosen hardware meet this requirement?**
Client equipment (camera) can record video at 30-35 FPS (Frames Per Second). Client wants for the *image processing task to be completed five times per sec.*	*FPGA is a flexible, robust, with high-performance and low latency. It can handle very large network. I will be able to meet customer requirements.*
System would need to be flexible so that it can be reprogrammed and optimized to quickly detect flaws in different chip designs.	*Since FPGAs are field-programmable, therefore they can be reprogrammed to adapt to new, evolving, and custom networks.*
It's a significant investment for a client but client expects to last at least 5-10 years.	*Has a long lifespan. Intel's FPGAs have a guaranteed availability of 10 years (from start to production).*
The system would need to be able to run inference on the video stream very quickly, to be able to detect chip flows without slowing down the packaging process.	*FPGAs can be used as a hardware accelerator.*
No budget limitations	*Expensive to develop.*
Customer wants to keep the floor running 24 hours a day.	*Robust with100% on-time performance.*

Queue Monitoring Requirements

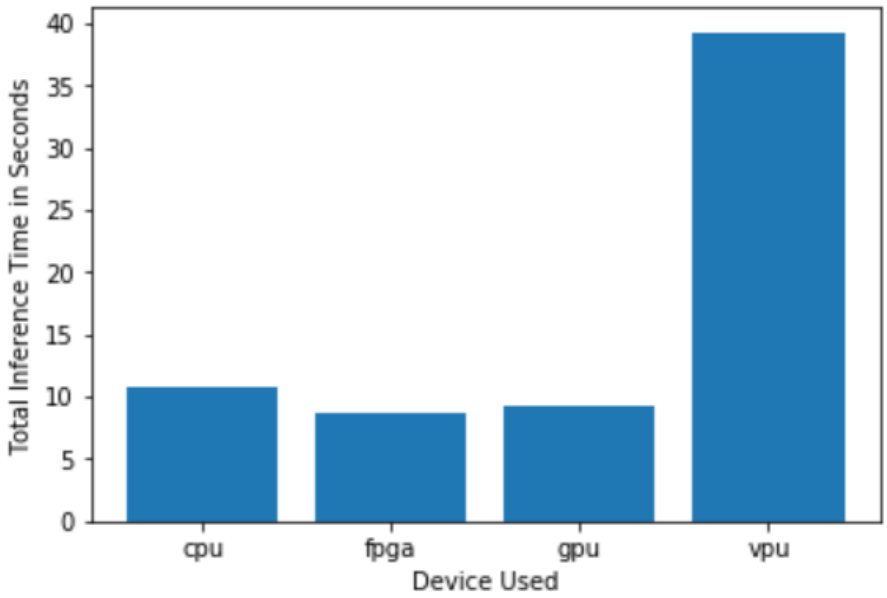
Maximum number of people in the queue	*2*
Model precision chosen (FP32, FP16, or Int8)	*Due to various precision options supports (FP16, 11, and 9) , FPGAs allow developers a balance between speed and accuracy. Here best suited is FP16.*

Test Results

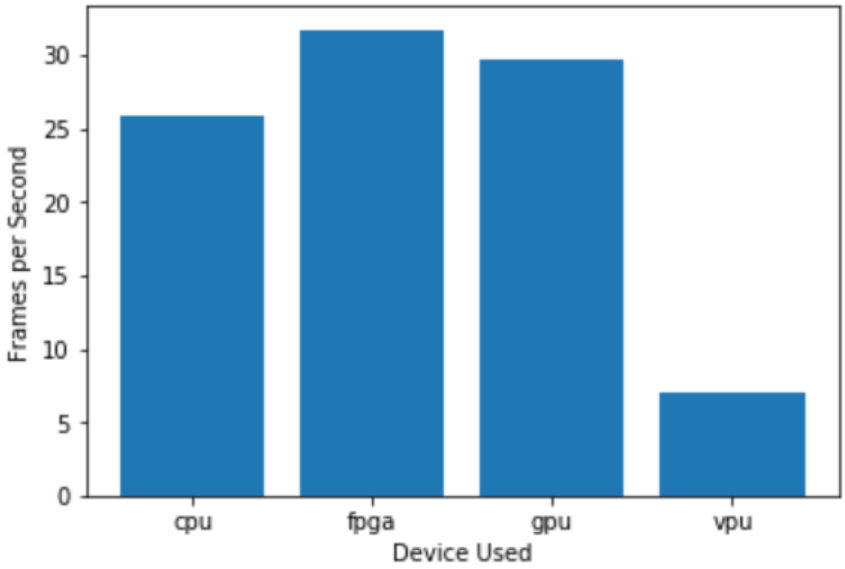
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

Overall, based on provided finings above FPGA is still the best option for the customer. And here is why: *With respect to performance, FPGA showed the best results comparing to CPU / IGPU and VPU. Even though it has a slightly higher than CPU and VPU model loading time, it has the highest number of FPS, and lowest inference time. It meets client 30-35 FPS with 5 image per second process and it's within customer budget as well.*

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)
Integrated GPU (IGPU)

Requirement Observed (Include at least two.)	**How does the chosen hardware meet this requirement?**
Client is already equipped with a modern computer with an Intel i7 core processor within it. These processors are not using they full compute power as they are mainly used to carry out some minimal tasks.	*Client can utilize existing hardware. Additional advantage of using IGPU is the fact that IGPU is located on the same die as CPU which reduce memory latency by speeding up data transfer between two devices.*
Limited budget.	*Using existing Intel i7 core processor reduce cost of investing in additional hardware.*
Client would like to save as much as possible on the electric bill.	*Due to configurable power consumption (controlled clock rate allows unused sections in a GPU to be powered down to reduce power consumption).*

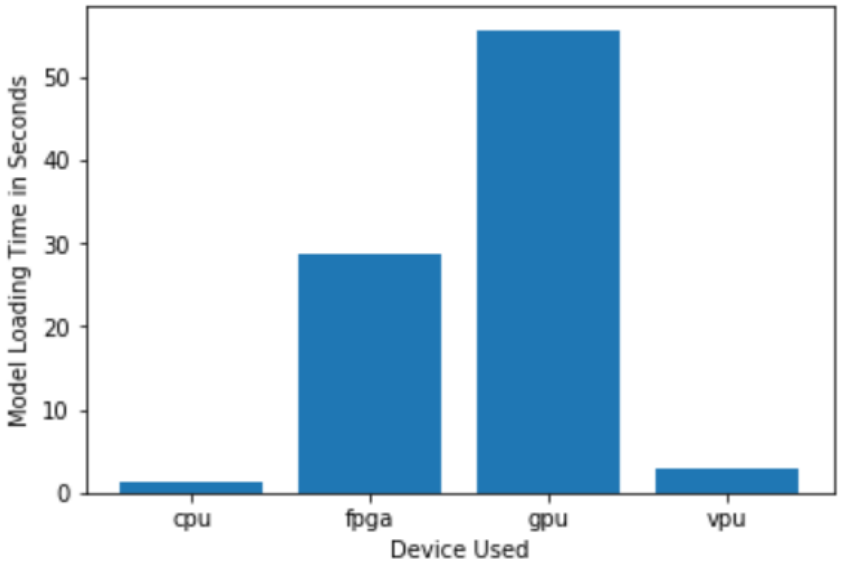
Queue Monitoring Requirements

Maximum number of people in the queue	*2-5. Total number of people waiting in a queue will depend how busy store is. During normal daily hours 2 is max number of people waiting in queue. When day is busy 5 is max number of people waiting in queue.*

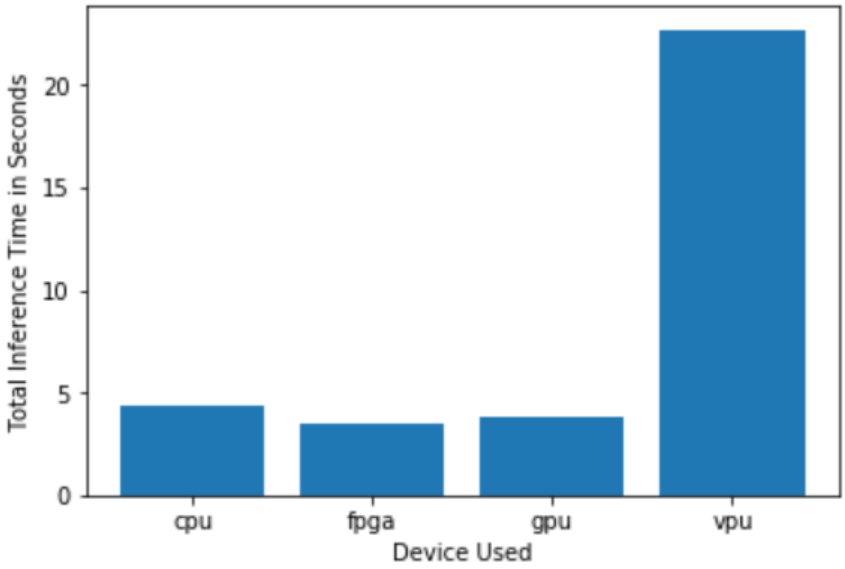
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

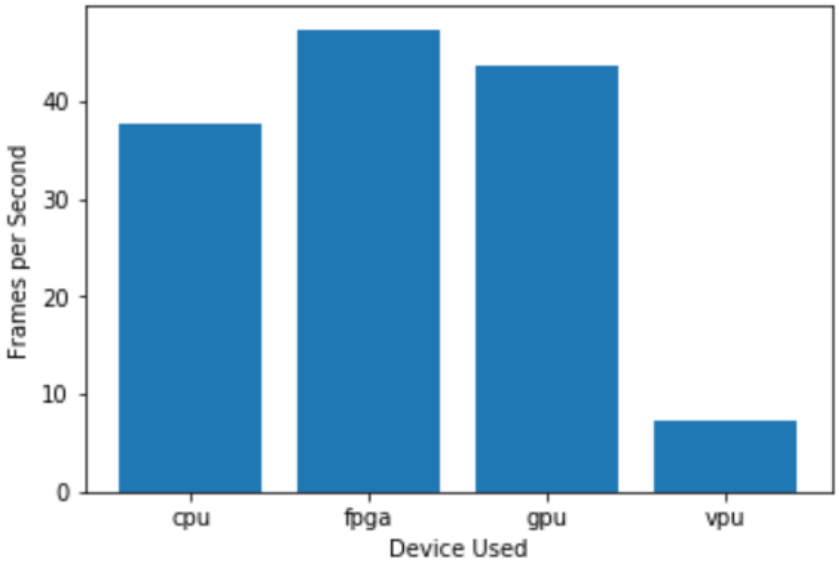
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
Based on the above findings, inference time and FTP are comparable with FPGA. Given the fact that IGPUS are optimized for 16bit data types, it gives better inference speed (as shown above) and can process twice as many operands per clock cycle. Due to processing the data in large batch size, IGPU might give a performance boost. Although model load time takes significantly longer than by using other hardware, it's a single transaction that is taking place during launching time. Thanks to ability to control a clock rate, unused sections can be shut down to reduce power consumption. Given the above explanation, IGPU meets the client expectations and is a final hardware recommendation.

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)
Vision Processing Units (VPUs)

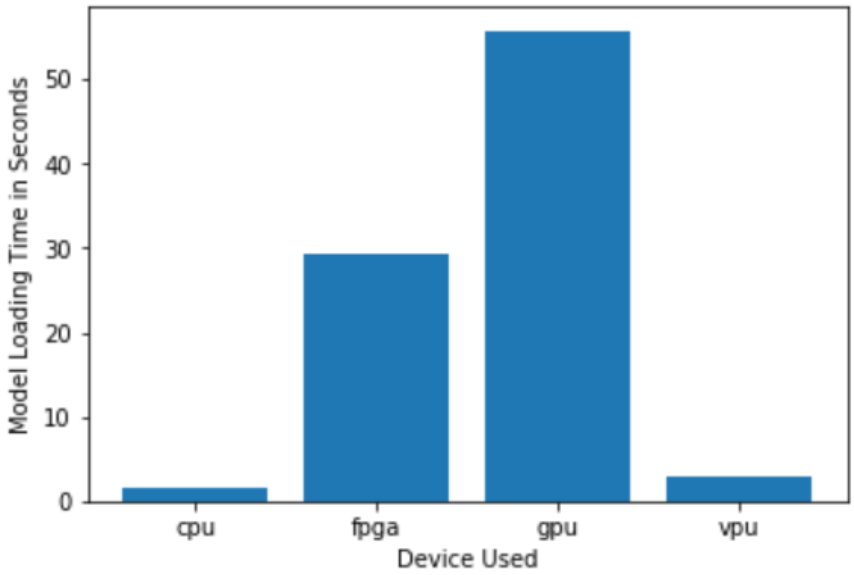
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Cline hardware supplies include 7 CCTV cameras on the platform that are connected to closed All-In-One PCs.	*Portable CPU can be easily plugged into USB port that are available on any PCs. Since VCPU is a specialized accelerator for improving image processing performance, it won't do actual calculations. VPU is an inexpensive solution to boost performance.*
The CPUs in mentioned PCs are being used to process and view CCTV footage for security purposes and no significant additional processing power is available to run inference.	*Low-power (1-2 Watts) device.*
Client has a fix budget of up to $300 per machine. In addition, a client would like to save as much as possible both on hardware and future power requirements.	*Low-cost device due to on-chip memory. Average cost of VPU is below $100, which is within client budget.*

Queue Monitoring Requirements

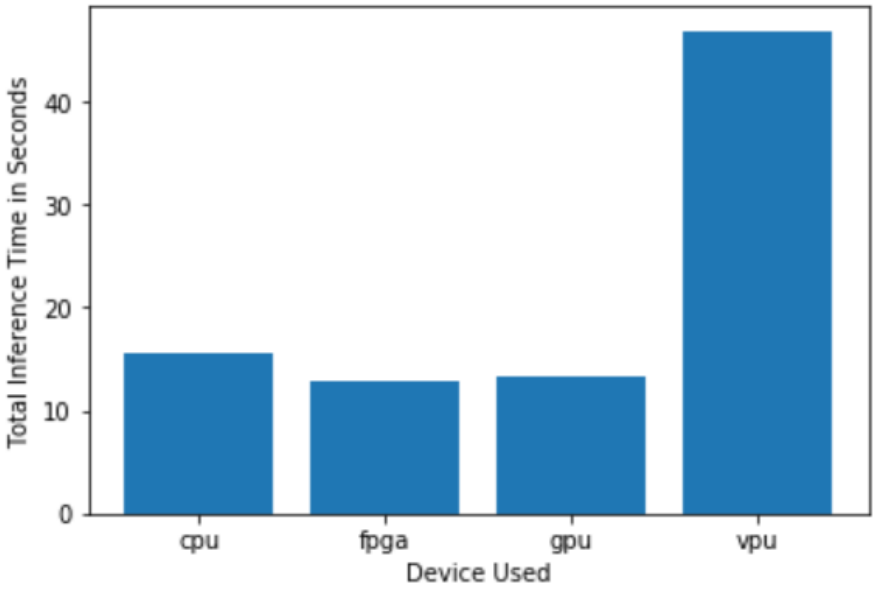
Maximum number of people in the queue	*7-15. (Based on provided video) total number of people waiting in a queue will depend how busy metro is. During normal daily hours 7 is max number of people waiting in queue outside every door. When day is busy 15 is max number of people waiting in queue outside every door.*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

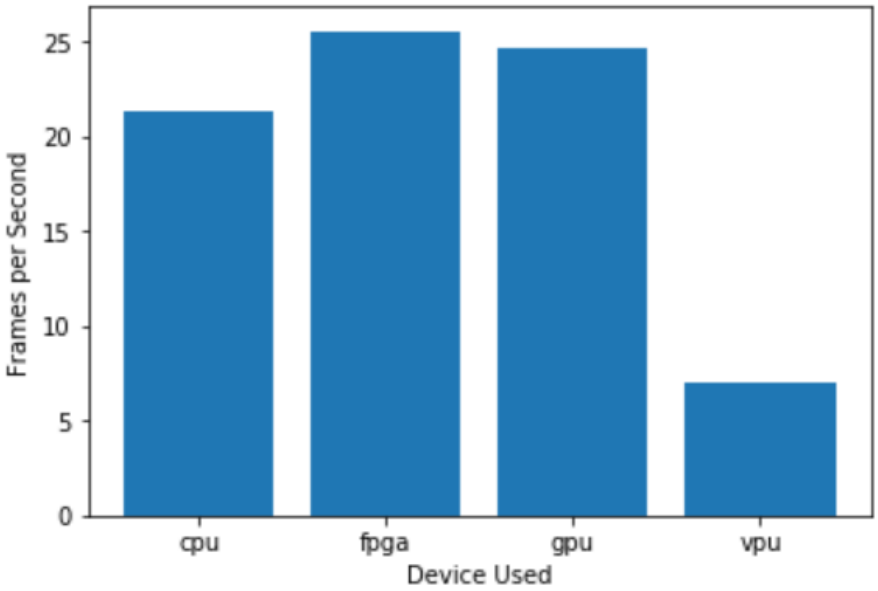
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
High VPU inference time contrast with low FTP (when comparing against other hardware resources). With that in mind VPU still meets customer conditions. With low-power and low-cost advantage, VPU will still meet up to 15 people in the queue requirement and it makes an ideal candidate for our customer needs.