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)
Based on the information provided by the customer (Naomi Semiconductors) and the list of requirements identified, the hardware that looks more appropriated to deploy the solution is an FPGA.

Requirement Observed (Include at least two.)	**How does the chosen hardware meet this requirement?**
Naomi Semiconductors has **plenty of revenue** to install a quality system	The company has plenty o revenue to invest in the solution, and there is no claim about the hardware cost. Therefore, an FPGA fits well into this requirement. The cost of an FPGA in the market is around thousands of dollars.
Workers alternate shifts to keep the floor **running 24 hours a day** so that packaging continues nonstop	FPGAs are designed to have 100% on-time performance, meaning they can be continuously running 24 hours a day, seven days a week, 365 days a year. They can also function over a wide range of temperatures, from 0° C to 60° C. This means that FPGAs can be deployed in harsh environments like factory floors and still perform optimally. Engineering from Intel has deployed solutions with guaranteed availability over ten years from the start of production.

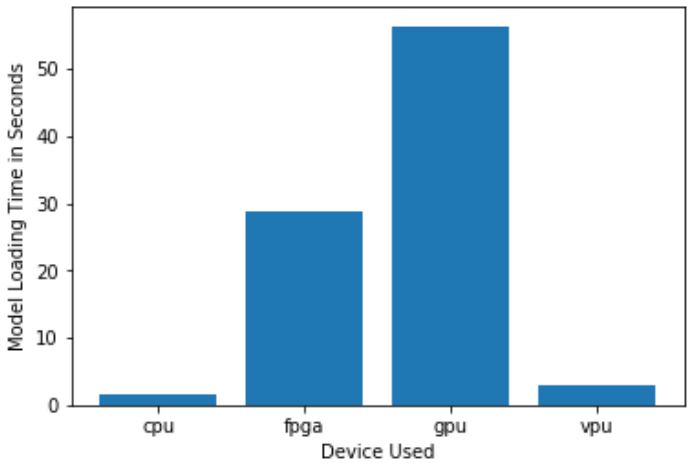
Mr. Vishwas wants a system to monitor the number of people in the factory line. The factory has a vision camera installed at every belt. **Each camera records video at 30-35 FPS (Frames Per Second)**	FPGAs can execute neural networks with high performance and very little latency. The high performance comes from the ability to run many sections of the FPGA in parallel.
The system would need to be able to run inference on the video stream very quickly. Additionally, also need to be flexible so that it can **be reprogrammed and optimized to quickly detect flaws in different chip designs.**	One of the main features of FPGAs is that they are field-programmable, which means that they can be reprogrammed to adapt to new, evolving, and custom networks. It can be achieved either using a high or low-level API, using languages such as python or c++ with the OpenVINO plugin (using OpenCL) or with RTL(Register transfer language) directly (To mention some alternatives).

Queue Monitoring Requirements

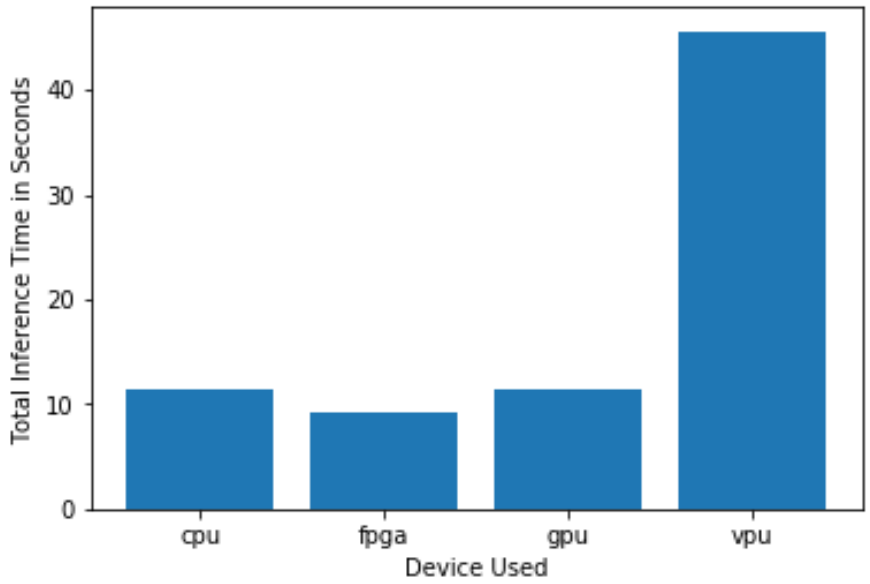
Maximum number of people in the queue	*Depends of the model and field view of the camera*
Model precision chosen (FP32, FP16, or Int8)	FP16, 11 and 9 bit are supported

Test Results

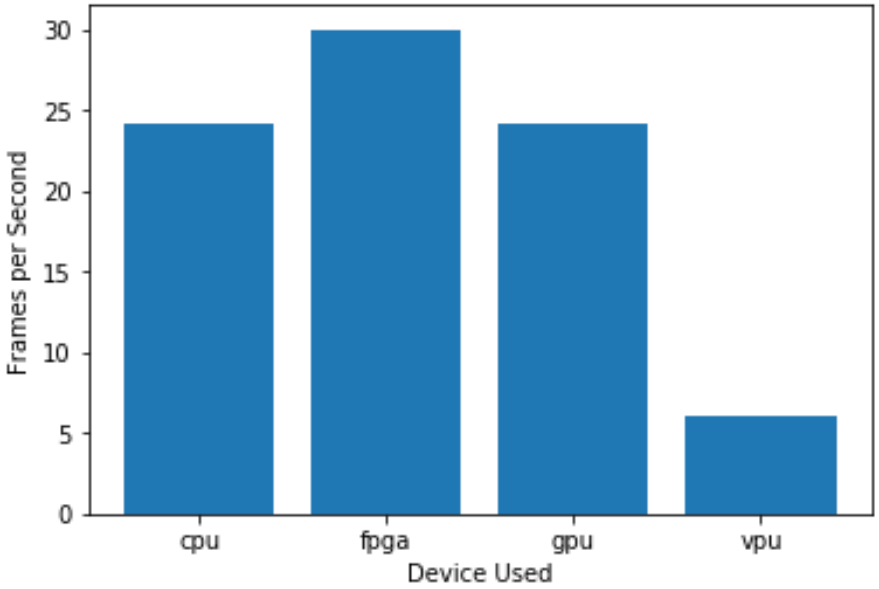
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

For this problem, before the evaluation of the devices, the hardware proposed was an FPGA, mainly because the user emphasized in the requirements that hardware should be reprogrammed and optimized to detect flaws in different chip designs quickly. Another feature considered was the FPS, the user mentioned that each camera records video at 30-35 FPS, based on the visualization generated after the evaluation in dev cloud we can infer that the FPGA is the solution that fits better in this scenario.

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

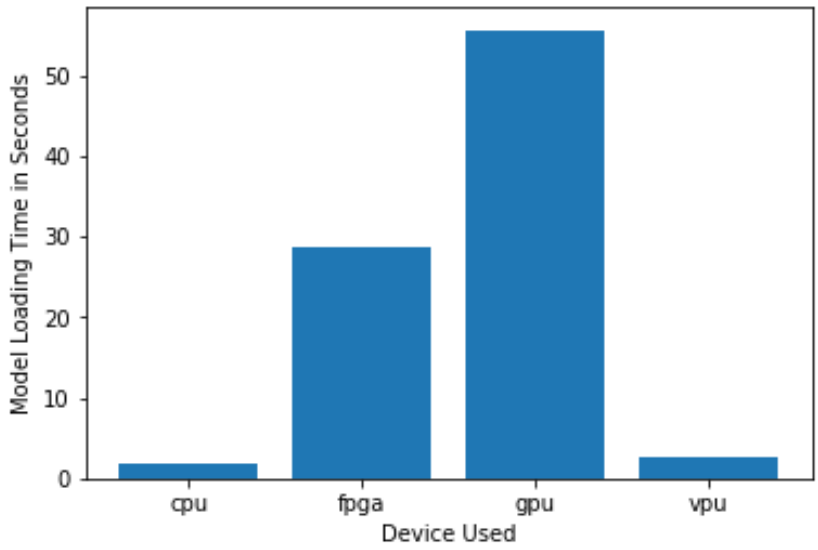
Requirement Observed (Include at least two.)	**How does the chosen hardware meet this requirement?**
Mr. Lin **does not have much money to invest** in additional hardware, and also would like to save as much as possible on his electric bill.	Using the CPU available on the existent PC for running the edge solution, the client doesn't need to incur extra expenses acquiring new hardware. Plus, there is no increase in power consumption.
Most of the store's checkout counters already have a modern computer, has an **Intel i7 core processor**.	Based on the client's assumption that these processors are only used to carry out some minimal tasks, the CPU can be used for making the inference.

Queue Monitoring Requirements

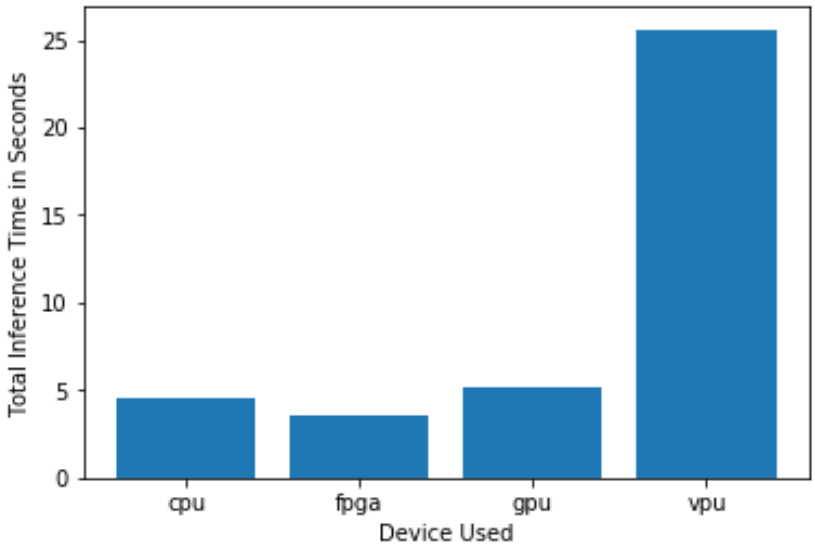
Maximum number of people in the queue	The total number of people in the checkout queue ranges from an average of 2 per queue (during normal daily hours) to 5 per queue (during rush hours).
Model precision chosen (FP32, FP16, or Int8)	*FP16, FP32, or Int8*

Test Results

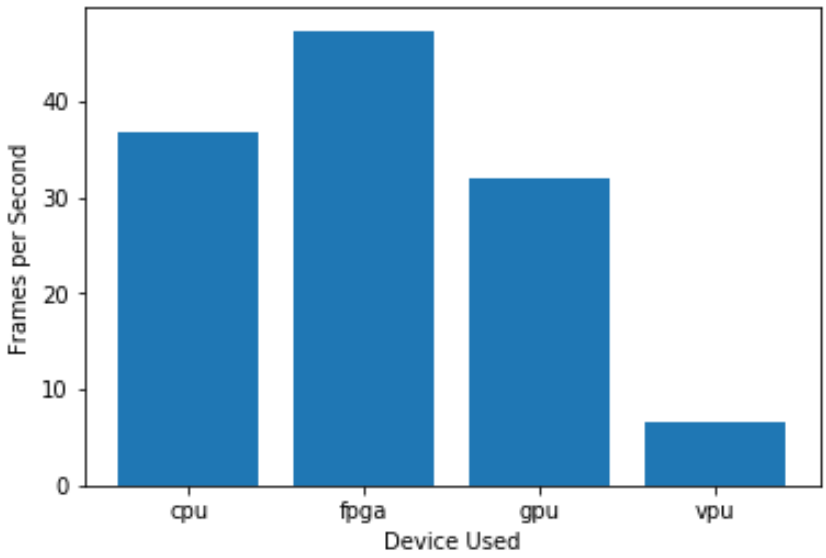
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
In this scenario, the client stated that he does not have much money to invest in the solution and would like to save as much as possible on his electric bill. Additionally, he mentioned that most of the store's checkout counters already have a modern computer, so using the CPU, he doesn't need to expend money buying new hardware and doesn't require an external power supply. Another alternative to the CPU can be a VPU; however, the CPU gives a better performance based on the comparison result.

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)
Since the budget is limited, in this scenario, the most appropriate hardware to use is a VPU

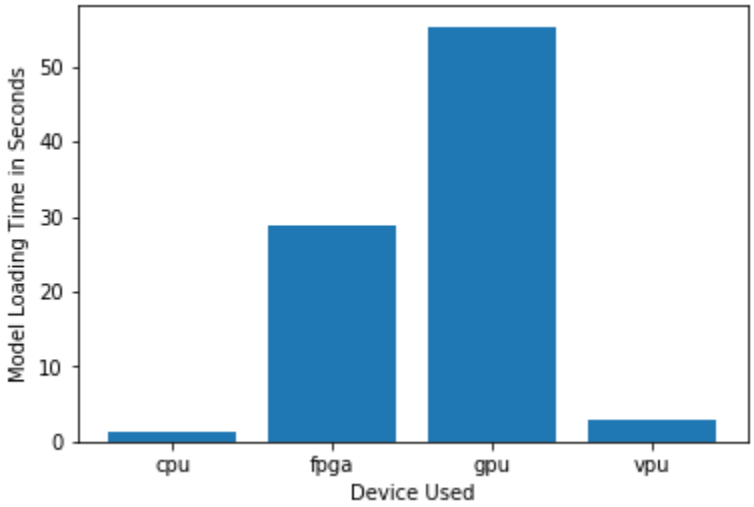
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Ms. Leah's **budget allows for a maximum of $300** per machine, and she would like to save as much as possible both on hardware and future power requirements.	VPUs are low-cost devices, for instance, Compared to other AI accelerators, the NCS2 is an inexpensive option, typically costing around $70 to $100.
They monitor the entire situation with 7 CCTV cameras on the platform. These are connected to closed All-In-One PCs that are located in a nearby security booth**. The CPUs in these machines are currently being used to process and view CCTV footage.**	VPUS can be used to accelerate the performance of a pre-existing system configuration. So, it can be attached to the existing All-In-One PCs.
For security purposes, and **no significant additional processing power is available to run inference.**	VPUS as low power consumption (1W), and they just require an usb connection available in the host PC. NO external power supply is required.

Queue Monitoring Requirements

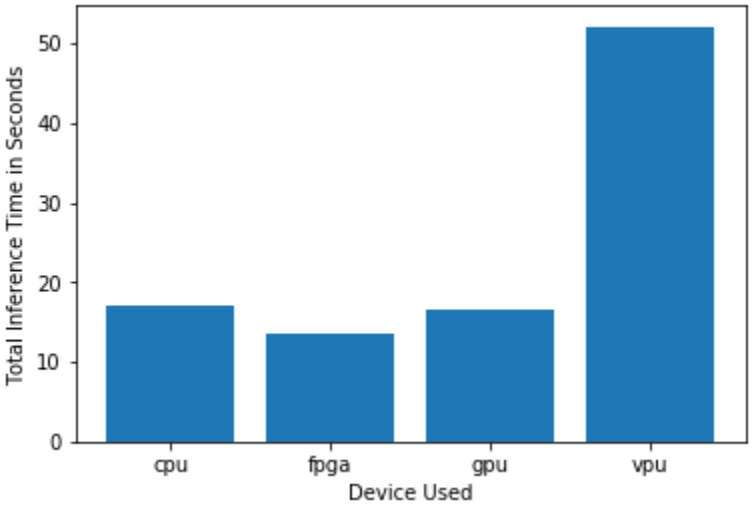
Maximum number of people in the queue	In peak hours they currently have over 15 people on average in a single queue outside every door in the Metro Rail. But during non-peak hours, the number of people reduces to 7 people in a single queue.
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

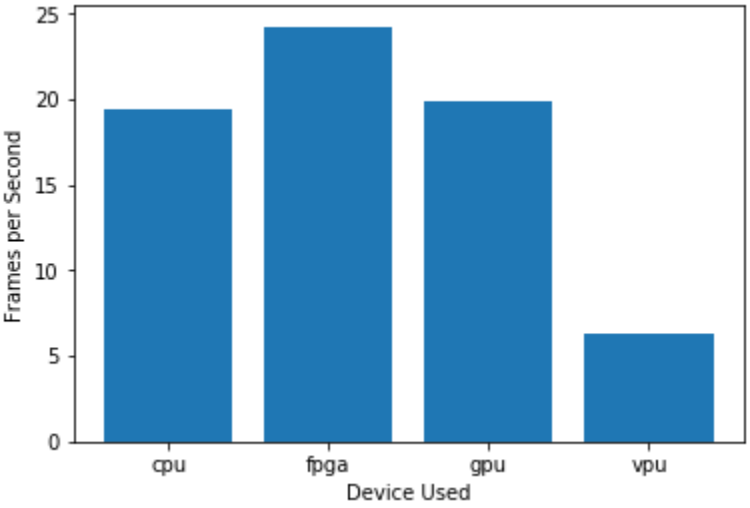
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
For the transportation solution, we can observe that the VPU performed pretty well, considering the inference time; however, the latency was low in comparison with the other alternatives. To sum up, we can conclude that the customer should consider going either for an FPGA, or CPU in order to get better results.