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?**
To be able to detect chip flaws without slowing down the packaging process, the system would need to be able to run inference on the video stream very quickly.	*FPGA* High performance, low latency
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.	**Flexibility.** FPGAs are flexible in a few different ways: ● They are field-programmable; they can be reprogrammed to adapt to new, evolving, and custom networks ● Various precision options (FP16, 11 and 9 bit) are supported—allowing developers a balance between speed and accuracy. ● The bitstreams being used can be updated without changing the hardware. This allows you to improve the

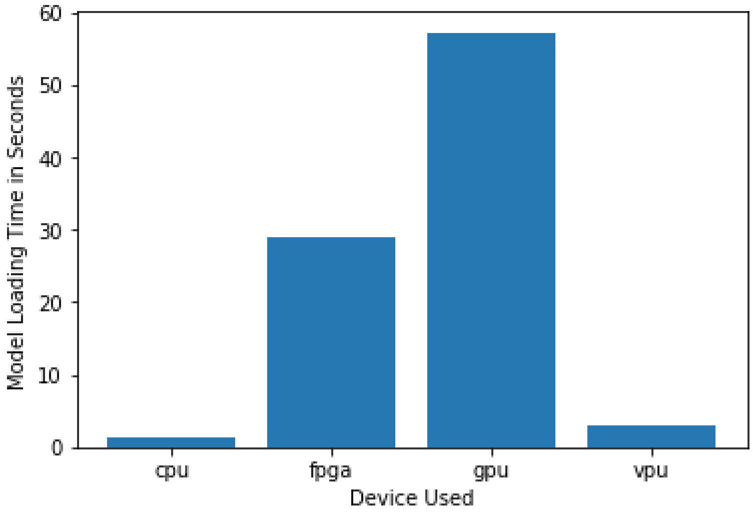
	performance of your system without replacing the FPGA.
Naomi Semiconductors has plenty of revenue to install a quality system, this is still a significant investment and they would ideally like it to last for at least 5-10 years.	**Robust.** FPGAs are designed to have 100% on-time performance, meaning they can be continuously running 24 hours a day, 7 days a week, 365 days a year. **Long Lifespan.** FPGAs have a long lifespan. For example, FPGAs that use devices from Intel's Internet of Things Group have a guaranteed availability of 10 years, from start of production.

Queue Monitoring Requirements

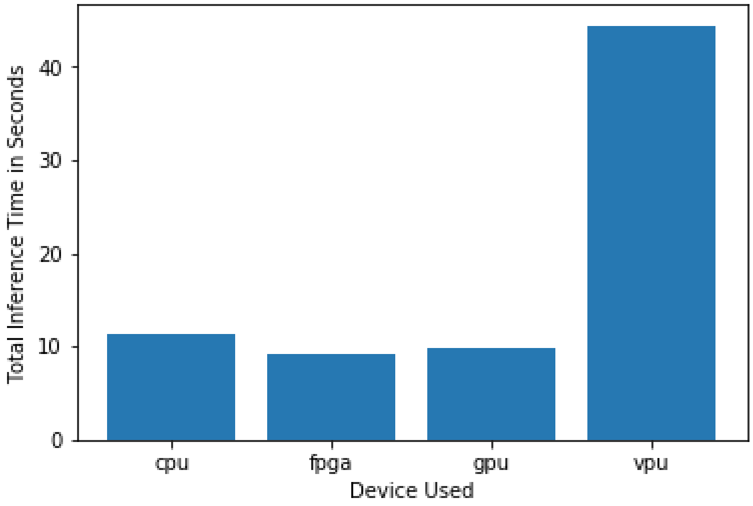
Maximum number of people in the queue	*8*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

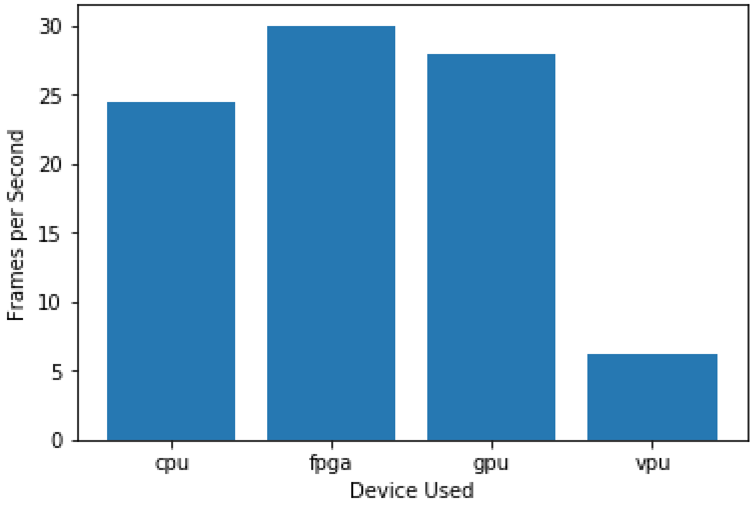
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

Naomi Semiconductors has plenty of revenue to install a quality system, this is still a significant investment and they would ideally like it to last for at least 5-10 years. To be able to detect chip flaws without slowing down the packaging process, 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.

***FPGA**:*
High performance, low latency as the test results.

Flexibility. FPGAs are flexible in a few different ways:

- They are field-programmable; they can be reprogrammed to adapt to new, evolving, and custom networks
- Various precision options (FP16, 11 and 9 bit) are supported—allowing developers a balance between speed and accuracy.
- The bitstreams being used can be updated without changing the hardware. This allows you to improve the performance of your system without replacing the FPGA.

Robust. FPGAs are designed to have 100% on-time performance, meaning they can be continuously running 24 hours a day, 7 days a week, 365 days a year.

Long Lifespan. FPGAs have a long lifespan. For example, FPGAs that use devices from Intel’s Internet of Things Group have a guaranteed availability of 10 years, from start of production.

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)
VPU

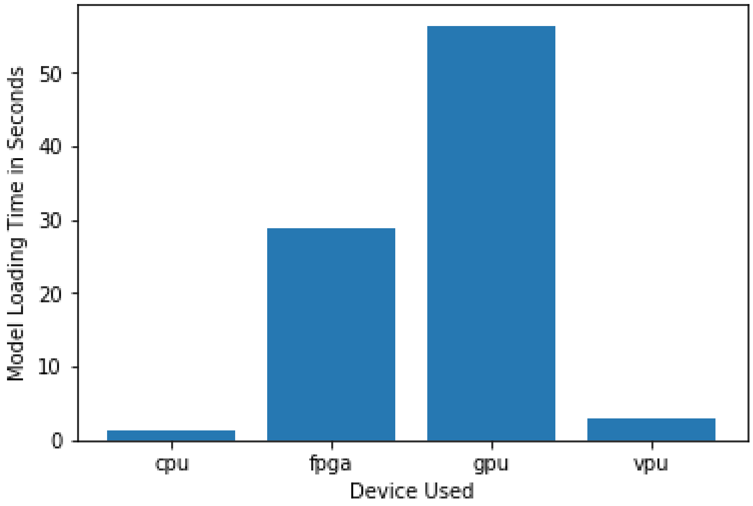
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Most of the store's checkout counters already have a modern computer, each of which has an Intel i7 core processor. Currently these processors are only used to carry out some minimal tasks that are not computationally expensive.	VPU is an *accelerator* which can *accelerate* the performance of the pre-existing CPU.
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.	*VPU is a low cost device can be use to accelerate the performance of a pre-existing system.*

Queue Monitoring Requirements

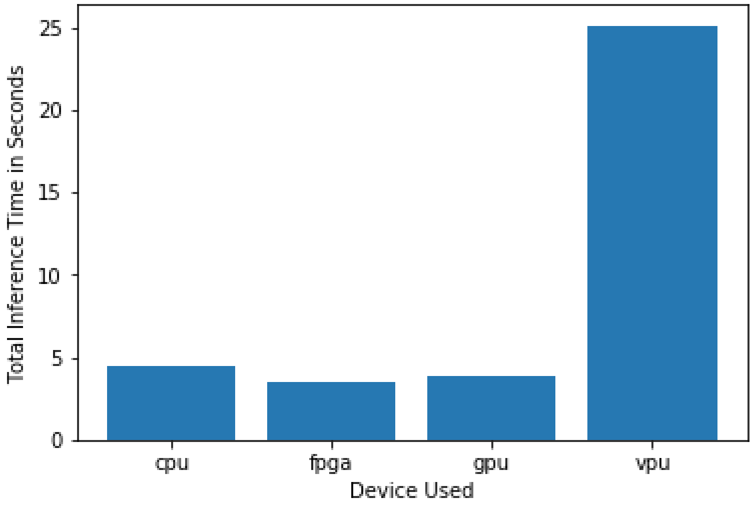
Maximum number of people in the queue	*6*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

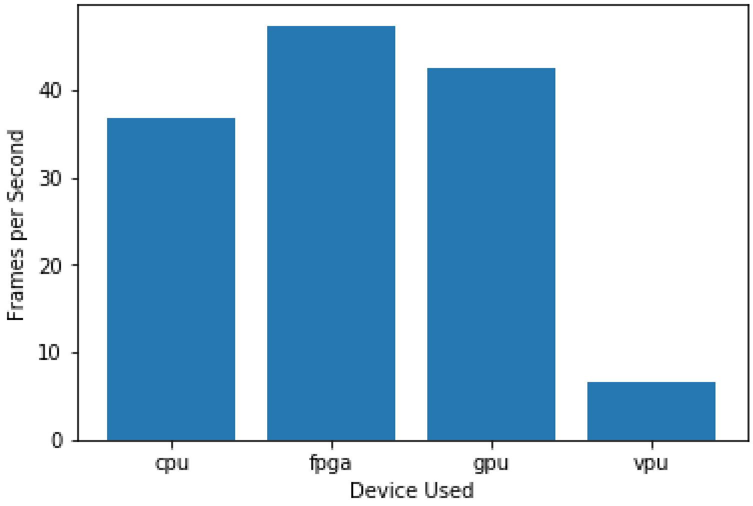
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
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. 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. ***VPU***: VPU is an *accelerator* which can *accelerate* the performance of the pre-existing CPU. *VPU is a low cost device that can be used to accelerate the performance of a pre-existing system.*

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

Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
Ms. Leah would like to automate this using an Edge AI system that would monitor the queues in real-time and quickly direct the crowd in the right manner.	**Model Precision and Speed.** On IGPUs, the Execution Unit instruction set and hardware are optimized for 16bit floating point data types. This improves inference speed, as we can process twice as many 16bit operands per clock cycle as we can when using 32 bit operands.
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	**Configurable Power Consumption.** The clock rate for the slice and unslice can be controlled separately. This means that unused sections in

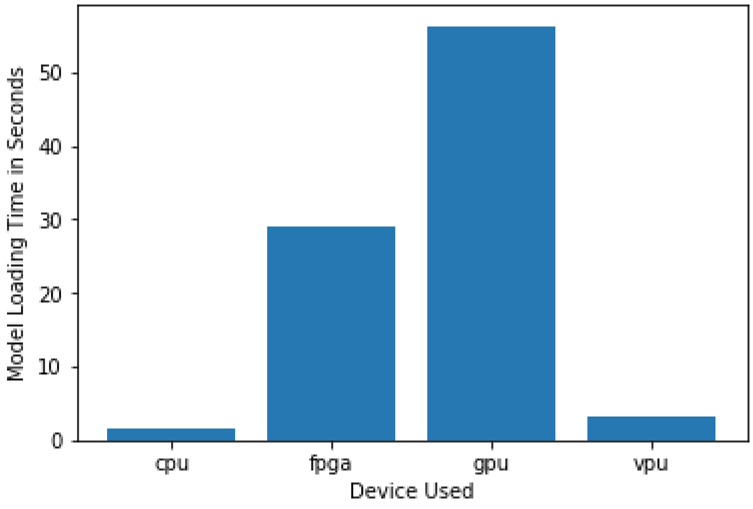
machines are currently being used to process and view CCTV footage for security purposes and no significant additional processing power is available to run inference.	a GPU can be powered down to reduce power consumption.

Queue Monitoring Requirements

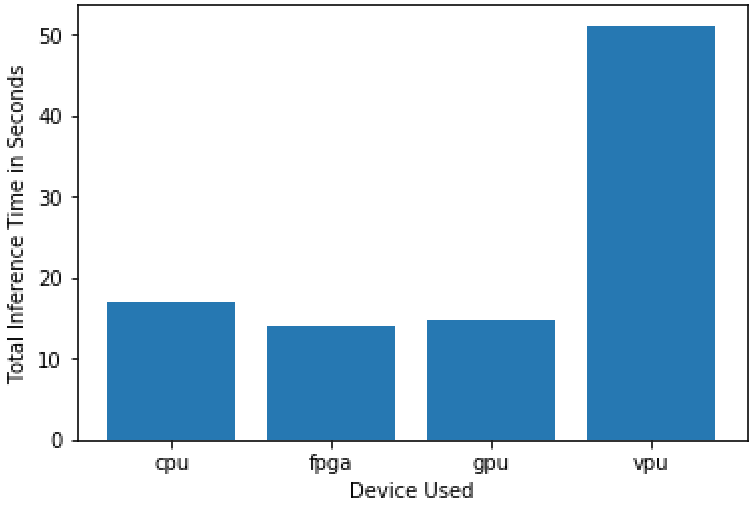
Maximum number of people in the queue	*12*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

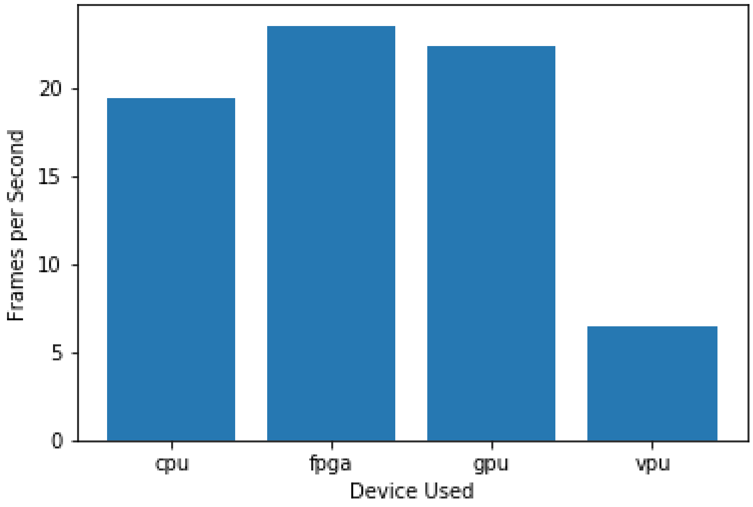
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

Ms. Leah would like to automate this using an Edge AI system that would monitor the queues in real-time and quickly direct the crowd in the right manner. 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 for security purposes and no significant additional processing power is available to run inference. 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.

IGPU:
Model Precision and Speed. On IGPUs, the Execution Unit instruction set and hardware are optimized for 16bit floating point data types. This improves inference speed, as we can process twice as many 16bit operands per clock cycle as we can when using 32 bit operands as the test results.
Configurable Power Consumption. The clock rate for the slice and unslice can be controlled separately. This means that unused sections in a GPU can be powered down to reduce power consumption.