



PROFESSIONAL-MACHINE- LEARNING-ENGINEER^{Q&As}

Professional Machine Learning Engineer

**Pass Google PROFESSIONAL-MACHINE-LEARNING-
ENGINEER Exam with 100% Guarantee**

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.geekcert.com/professional-machine-learning-engineer.html>

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Google
Official Exam Center



VCE & PDF

GeekCert.com

<https://www.geekcert.com/professional-machine-learning-engineer.html>

2024 Latest geekcert PROFESSIONAL-MACHINE-LEARNING-ENGINEER

PDF and VCE dumps Download

- ⚙️ **Instant Download** After Purchase
- ⚙️ **100% Money Back** Guarantee
- ⚙️ **365 Days** Free Update
- ⚙️ **800,000+** Satisfied Customers





QUESTION 1

You developed an ML model with AI Platform, and you want to move it to production. You serve a few thousand queries per second and are experiencing latency issues. Incoming requests are served by a load balancer that distributes them across multiple KubeFlow CPU-only pods running on Google Kubernetes Engine (GKE). Your goal is to improve the serving latency without changing the underlying infrastructure. What should you do?

- A. Significantly increase the `max_batch_size` TensorFlow Serving parameter.
- B. Switch to the `tensorflow-model-server-universal` version of TensorFlow Serving.
- C. Significantly increase the `max_enqueued_batches` TensorFlow Serving parameter.
- D. Recompile TensorFlow Serving using the source to support CPU-specific optimizations. Instruct GKE to choose an appropriate baseline minimum CPU platform for serving nodes.

Correct Answer: D

QUESTION 2

You are an ML engineer at a travel company. You have been researching customers' travel behavior for many years, and you have deployed models that predict customers' vacation patterns. You have observed that customers' vacation destinations vary based on seasonality and holidays; however, these seasonal variations are similar across years. You want to quickly and easily store and compare the model versions and performance statistics across years. What should you do?

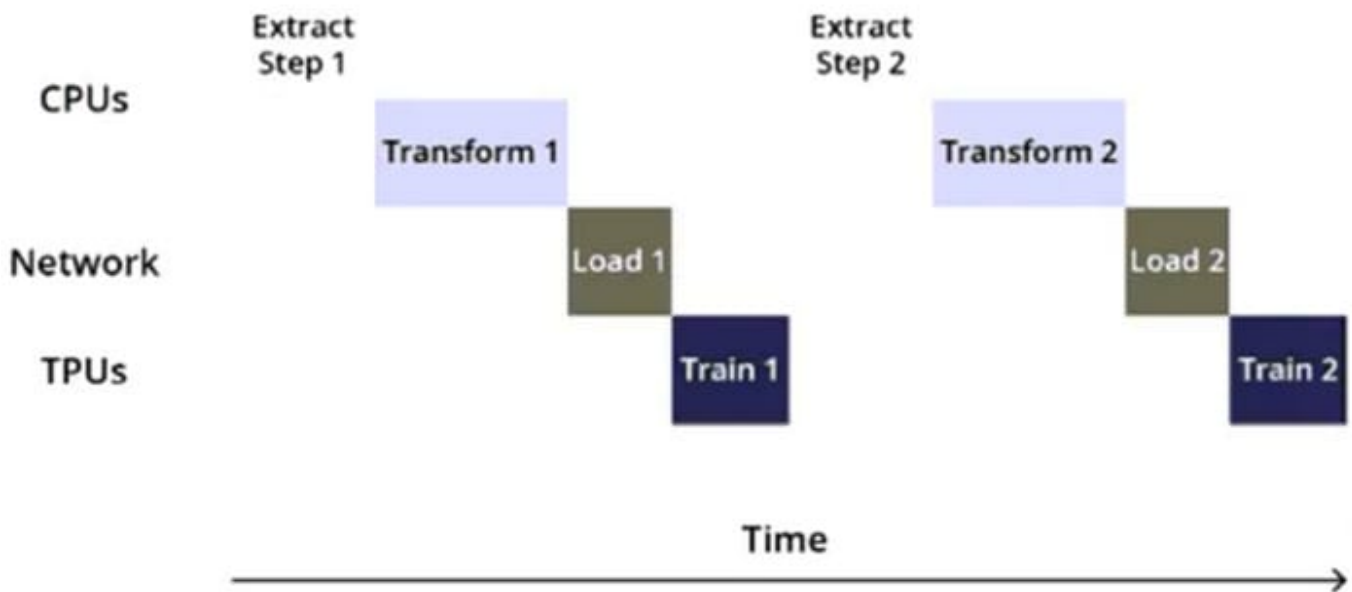
- A. Store the performance statistics in Cloud SQL. Query that database to compare the performance statistics across the model versions.
- B. Create versions of your models for each season per year in Vertex AI. Compare the performance statistics across the models in the Evaluate tab of the Vertex AI UI.
- C. Store the performance statistics of each pipeline run in KubeFlow under an experiment for each season per year. Compare the results across the experiments in the KubeFlow UI.
- D. Store the performance statistics of each version of your models using seasons and years as events in Vertex ML Metadata. Compare the results across the slices.

Correct Answer: D

<https://cloud.google.com/vertex-ai/docs/ml-metadata/analyzing#filtering>

QUESTION 3

You are training an object detection model using a Cloud TPU v2. Training time is taking longer than expected. Based on this simplified trace obtained with a Cloud TPU profile, what action should you take to decrease training time in a cost-efficient way?



- A. Move from Cloud TPU v2 to Cloud TPU v3 and increase batch size.
- B. Move from Cloud TPU v2 to 8 NVIDIA V100 GPUs and increase batch size.
- C. Rewrite your input function to resize and reshape the input images.
- D. Rewrite your input function using parallel reads, parallel processing, and prefetch.

Correct Answer: D

Based on the profile, it appears that the Compute time is relatively low compared to the HostToDevice and DeviceToHost time. This suggests that the data transfer between the host (CPU) and the TPU device is a bottleneck. Therefore, the best action to decrease training time in a cost-efficient way would be to reduce the amount of data transferred between the host and the device.

QUESTION 4

You need to execute a batch prediction on 100 million records in a BigQuery table with a custom TensorFlow DNN regressor model, and then store the predicted results in a BigQuery table. You want to minimize the effort required to build this inference pipeline. What should you do?

- A. Import the TensorFlow model with BigQuery ML, and run the ml.predict function.
- B. Use the TensorFlow BigQuery reader to load the data, and use the BigQuery API to write the results to BigQuery.
- C. Create a Dataflow pipeline to convert the data in BigQuery to TFRecords. Run a batch inference on Vertex AI Prediction, and write the results to BigQuery.
- D. Load the TensorFlow SavedModel in a Dataflow pipeline. Use the BigQuery I/O connector with a custom function to perform the inference within the pipeline, and write the results to BigQuery.

Correct Answer: A

<https://cloud.google.com/bigquery-ml/docs/making-predictions-with-imported-tensorflow-models#api>
<https://towardsdatascience.com/how-to-do-batch-predictions-of-tensorflow-models-directly-in-bigquery-ffa843ebdba6>



QUESTION 5

During batch training of a neural network, you notice that there is an oscillation in the loss. How should you adjust your model to ensure that it converges?

- A. Decrease the size of the training batch.
- B. Decrease the learning rate hyperparameter.
- C. Increase the learning rate hyperparameter.
- D. Increase the size of the training batch.

Correct Answer: B

<https://ai.stackexchange.com/questions/14079/what-could-an-oscillating-training-loss-curve-represent#:~:text=Try%20lowering%20the%20learning%20rate,step%20and%20overshoot%20it%20again.>

[Latest PROFESSIONAL-MACHINE-LEARNING-ENGINEER Dumps](#)

[PROFESSIONAL-MACHINE-LEARNING-ENGINEER Exam Questions](#)

[PROFESSIONAL-MACHINE-LEARNING-ENGINEER Braindumps](#)