Exam Professional Machine Learning Engineer All Questions

View all questions & answers for the Professional Machine Learning Engineer exam

Go to Exam

EXAM PROFESSIONAL MACHINE LEARNING ENGINEER TOPIC 1 QUESTION 177 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer
Question #: 177
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You have created a Vertex AI pipeline that includes two steps. The first step preprocesses 10 TB data completes in about 1 hour, and saves the result in a Cloud Storage bucket. The second step uses the processed data to train a model. You need to update the model's code to allow you to test different algorithms. You want to reduce pipeline execution time and cost while also minimizing pipeline changes. What should you do?

A. Add a pipeline parameter and an additional pipeline step. Depending on the parameter value, the pipeline step conducts or skips data preprocessing, and starts model training.

B. Create another pipeline without the preprocessing step, and hardcode the preprocessed Cloud Storage file location for model training.

C. Configure a machine with more CPU and RAM from the compute-optimized machine family for the data preprocessing step.

D. Enable caching for the pipeline job, and disable caching for the model training step.

Show Suggested Answer

by pikachu007 at *Jan. 11, 2024, 11:24 a.m.*

Comments

Type your comment...

Submit

fitri001 6 months, 2 weeks ago

Selected Answer: D

Caching Preprocessed Data: Since the preprocessed data (10 TB) is the same for different model training runs, enabling caching allows Vertex AI to reuse it for subsequent pipeline executions. This significantly reduces execution time and cost, especially for large datasets.
Disabling Model Training Cache: Model training is typically non-deterministic due to factors like random initialization. Caching the model training step could lead to stale models and inaccurate results. Disabling caching ensures the model is re-trained each time with potentially updated code for different algorithms.

upvoted 1 times

gscharly 6 months, 2 weeks ago

Selected Answer: D

agree with guilhermebutzke

upvoted 1 times

guilhermebutzke 8 months, 4 weeks ago

Selected Answer: D

According to this documentation cited: https://cloud.google.com/vertex-ai/docs/pipelines/configure-caching

it is possible to write a pipeline setting True or False for each task component, like this:

```
# Model training step with caching disabled
train_model_task = train_model_op()
train_model_task.set_caching_options(False) # Disable caching for this step

# Model training step depends on the preprocessing step
train_model_task.after(preprocess_task)
```

So, with this, letter D is the best option. Furthermore, letter A and, Adding a pipeline parameter and an additional pipeline step introduces unnecessary complexity when caching can handle conditional execution efficiently and in letter C, configuring a machine with more CPU and RAM for preprocessing does not address the goal of minimizing pipeline changes and reducing execution time/cost effectively.

upvoted 3 times

b1a8fae 9 months, 4 weeks ago

Selected Answer: D

Not A. Adding a pipeline parameter and new pipeline steps does not minimise pipeline changes.

Not C. The idea is not to re-run the preprocessing step at all.

Not B. Creating a whole new pipeline implies a significant investment of effort.

I opt for D: Enabling caching only for preprocessing job (although it says “pipeline job” in the option, I think that is a typo). Quoting Vertex AI docs: “If there is a matching execution in Vertex ML Metadata, the outputs of that execution are used and the step is skipped. This helps to reduce costs by skipping computations that were completed in a previous pipeline run.” https://cloud.google.com/vertex-ai/docs/pipelines/configure-caching

upvoted 4 times

pikachu007 9 months, 4 weeks ago

Selected Answer: A

The pipeline already generates the preprocessed dataset and stores, there's no need to preprocess again for another model

upvoted 1 times

pikachu007 9 months, 3 weeks ago

rereading the question, I agree with b1a8fae that its D

upvoted 1 times

Start Learning for free

Social Media

Facebook , Twitter
YouTube , Reddit
Pinterest

We are the biggest and most updated IT certification exam material website.
Using our own resources, we strive to strengthen the IT professionals community for free.