



- Expert Verified, Online, Free.

MENU



Google Discussions



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 201 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 201

Topic #: 1

[\[All Professional Machine Learning Engineer Questions\]](#)

You developed a Vertex AI pipeline that trains a classification model on data stored in a large BigQuery table. The pipeline has four steps, where each step is created by a Python function that uses the KubeFlow v2 API. The components have the following names:

```
dt=datetime.now().strftime("%Y%m%d%H%M%S")
f"export-{dt}.yaml", f"preprocess-{dt}.yaml", f"train-{dt}.yaml",
f"calibrate-{dt}.yaml"
```

You launch your Vertex AI pipeline as the following:

```
job = aip.PipelineJob(
    display_name="my-awesome-pipeline",
    template_path="pipeline.json",
    job_id=f"my-awesome-pipeline-{dt}",
    parameter_values=params,
    enable_caching=True,
    location="europe-west1"
)
```

You perform many model iterations by adjusting the code and parameters of the training step. You observe high costs associated with the development, particularly the data export and preprocessing steps. You need to reduce model development costs. What should you do?

- A. Change the components' YAML filenames to export.yaml, preprocess.yaml, f "train-{dt}.yaml", f "calibrate-{dt}.yaml".
- B. Add the {"kubeflow.v1.caching": True} parameter to the set of params provided to your PipelineJob.
- C. Move the first step of your pipeline to a separate step, and provide a cached path to Cloud Storage as an input to the main pipeline.
- D. Change the name of the pipeline to f"my-awesome-pipeline-{dt}".

Show Suggested Answer

by  pikachu007 at Jan. 13, 2024, 4:21 a.m.

Comments

Type your comment...

Submit

  **Foxy2021** 1 week, 3 days ago

I select C:

By leveraging a Dataproc cluster, you can maintain compatibility with your existing PySpark jobs, minimize management overhead, and create a scalable proof of concept quickly and efficiently.

   upvoted 1 times

  **Foxy2021** 1 week, 3 days ago

I select B.

A:

Changing the YAML filenames does not affect caching behavior or cost reduction. The pipeline's efficiency and cost effectiveness are primarily governed by how it handles inputs and outputs rather than the filenames of the components.
C:

Moving the first step to a separate pipeline may help with organization but doesn't directly address the cost incurred by repeated data exports and preprocessing. Also, simply providing a cached path does not guarantee that the preprocessing step itself won't be executed multiple times.

D:

Changing the name of the pipeline to include a timestamp or other identifier does not influence caching or resource usage. It merely alters the identification of the pipeline runs without any impact on the efficiency of the operations being performed.

   upvoted 1 times

  **gscharly** 6 months, 1 week ago

Selected Answer: A

see guilhermebutzke

   upvoted 1 times

  **pinimichele01** 6 months, 1 week ago

Selected Answer: A

see guilhermebutzke

   upvoted 1 times

  **Yan_X** 7 months, 2 weeks ago

Selected Answer: C

C

Caching should be enabled for all steps, e.g., export, preprocessing and training.

   upvoted 1 times

  **guilhermebutzke** 8 months, 1 week ago

Selected Answer: A

My Answer: A

From what I understood, it's about optimizing the process of adjusting code while utilizing previously processed results from the pipeline. Kubeflow inherently caches these steps, eliminating the need to explicitly store results in a designated path.

However, the original filenames include a timestamp (**`-dt`**), suggesting that by removing this timestamp, the pipeline steps might not rerun as expected.

Option C could be an approach, but it would require more effort to implement (since Kubeflow handles it automatically). Additionally, the beginning of the option only mentions moving the first step, which is the export, and doesn't say anything about preprocessing (which could be one of the more expensive steps).

So, considering all of these factors, I think A is the best choice."

👍 ↩ 🚩 upvoted 4 times

🗨️ 👤 **shadz10** 9 months, 1 week ago

Selected Answer: C

Not A - Changing file names does not help with reducing costs

Not B - you cannot directly use kubeflow.v1.caching on a pipeline that uses the KubeFlow v2 API. Version Incompatibility: The kubeflow.v1.caching module is specifically designed for KubeFlow Pipelines v1, and its structure and functionality are not directly compatible with KubeFlow Pipelines v2.

so best option here is C

👍 ↩ 🚩 upvoted 2 times

🗨️ 👤 **b1a8fae** 9 months, 1 week ago

Selected Answer: C

I considered B but a search of "kubeflow.v1.caching" on Google only produces 1 result, which is this very question on this very website. Thus, I rule it out as non-existent (please share a resource if there is any that proves it exists) and opt for C.

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **BlehMaks** 9 months, 1 week ago

Selected Answer: A

i think it's A.

1)if we want to use the same results several times we shouldn't rename them. so we need to delete {dt} from the first two components names.

2)we already have this option enable_caching = True, why do we need kubeflow.v1.caching then?

3)i'm not sure but may be it does metter

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **BlehMaks** 9 months, 1 week ago

3)i'm not sure but may be it does matter that KubeFlow v2 API and kubeflow.v1.caching have different versions (v1 and v2)

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **pikachu007** 9 months, 1 week ago

Selected Answer: B

Enables caching: Setting this parameter instructs Vertex AI Pipelines to cache the outputs of pipeline steps that have successfully completed. This means that if a step's inputs haven't changed, its execution can be skipped, reusing the cached output instead.

Targets costly steps: The prompt highlights that data export and preprocessing steps are particularly expensive. Caching these steps can significantly reduce costs during model iterations.

👍 ↩ 🚩 upvoted 2 times



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.



© 2024 ExamTopics

ExamTopics doesn't offer Real Microsoft Exam Questions. ExamTopics doesn't offer Real Amazon Exam Questions. ExamTopics Materials do not contain actual questions and answers from Cisco's Certification Exams.

CFA Institute does not endorse, promote or warrant the accuracy or quality of ExamTopics. CFA® and Chartered Financial Analyst® are registered trademarks owned by CFA Institute.