- Expert Verified, Online, Free.

**■** MENU

C

**G** 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 39 DISCUSSIO..**

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 39

Topic #: 1

[All Professional Machine Learning Engineer Questions]

You work with a data engineering team that has developed a pipeline to clean your dataset and save it in a Cloud Storage bucket. You have created an ML model and want to use the data to refresh your model as soon as new data is available. As part of your CI/CD workflow, you want to automatically run a Kubeflow

Pipelines training job on Google Kubernetes Engine (GKE). How should you architect this workflow?

- A. Configure your pipeline with Dataflow, which saves the files in Cloud Storage. After the file is saved, start the training job on a GKE cluster.
- B. Use App Engine to create a lightweight python client that continuously polls Cloud Storage for new files. As soon as a file arrives, initiate the training job.
- C. Configure a Cloud Storage trigger to send a message to a Pub/Sub topic when a new file is available in a storage bucket. Use a Pub/Sub-triggered Cloud Function to start the training job on a GKE cluster.
- D. Use Cloud Scheduler to schedule jobs at a regular interval. For the first step of the job, check the timestamp of objects in your Cloud Storage bucket. If there are no new files since the last run, abort the job.

**Show Suggested Answer** 

by Paul\_Dirac at June 26, 2021, 10:56 a.m.

#### Comments

Submit  Paul_Dirac Highly Voted 1 3 years, 4 months ago C https://cloud.google.com/architecture/architecture-for-mlops-using-tfx-kubeflow-pipelines-and-cloud-build#triggering-an	
Paul_Dirac Highly Voted 1 3 years, 4 months ago	
C	
scheduling-kubeflow-pipelines  upvoted 15 times	<b>i</b> -
Paul_Dirac Highly Voted 10 3 years, 4 months ago	1-
C https://cloud.google.com/architecture/architecture-for-mlops-using-tfx-kubeflow-pipelines-and-cloud-build#triggering-an scheduling-kubeflow-pipelines  upvoted 7 times	
<ul> <li>□  ori5225 3 years, 2 months ago</li> <li>On a schedule, using Cloud Scheduler.</li> <li>Responding to an event, using Pub/Sub and Cloud Functions. For example, the event can be the availability of new files in a Cloud Storage bucket.</li> <li>□  upvoted 1 times</li> </ul>	data
<ul> <li>tavva_prudhvi 1 year, 4 months ago</li> <li>Option D requires the job to be scheduled at regular intervals, even if there are no new files. This can waste reso and lead to unnecessary delays in the training process.</li> <li>upvoted 1 times</li> </ul>	urces
□ ♣ PhilipKoku Most Recent ② 5 months ago	
Selected Answer: C	
C) PUB/sub trigger from Cloud Storage & Cloud Function  i pupoted 1 times	
= 🏝 fragkris 11 months ago	
Selected Answer: C  C - This is the google reccomended method.   □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □	
□ 🏝 Sum_Sum 11 months, 3 weeks ago	
Selected Answer: C  C- because you don't want to re-engineer the pipeline  upvoted 1 times	
M25 1 year, 6 months ago	
Selected Answer: C  Went with C	
Fatiy 1 year, 8 months ago	
Selected Answer: C	
The scenario involves automatically running a Kubeflow Pipelines training job on GKE as soon as new data becomes available. To achieve this, we can use Cloud Storage to store the cleaned dataset, and then configure a Cloud Storage trigger that sends a message to a Pub/Sub topic whenever a new file is added to the storage bucket. We can then crea Pub/Sub-triggered Cloud Function that starts the training job on a GKE cluster.	te a

E hehzadsw 1 year, 10 months ago

#### Selected Answer: A

The question says: As part of your CI/CD workflow, you want to automatically run a Kubeflow..

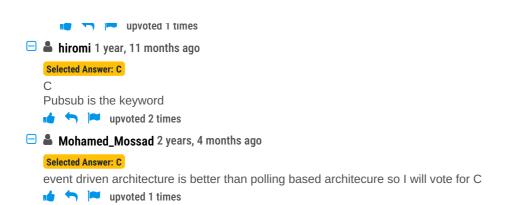
C is also an option but it seems more cumbersome.

One thing hat could be against A is that the data engineering team is separate team so they might not access your CI/CD if any changes from their side is needed..

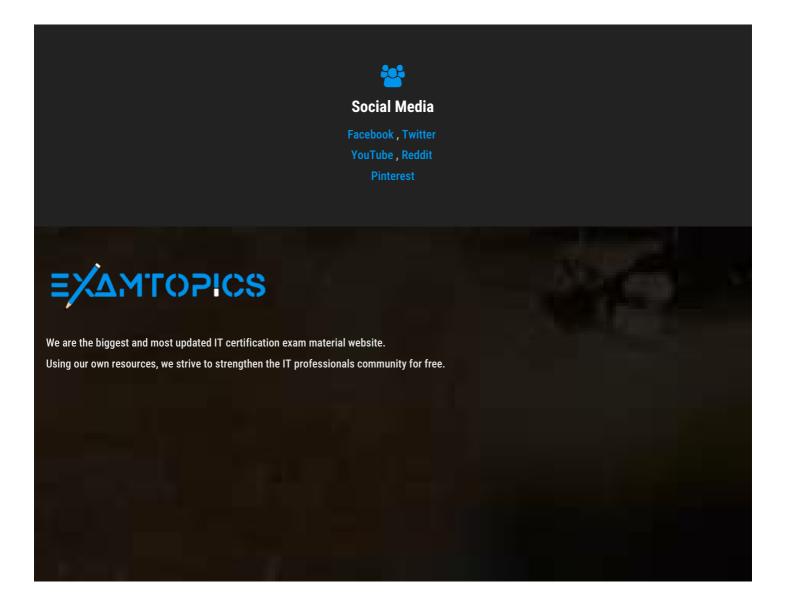
upvoted 1 times

# 🖃 🏜 tavva\_prudhvi 1 year, 4 months ago

Option A requires the data engineering team to modify the pipeline, which can be time-consuming and error-prone.



# Start Learning 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.