



- 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 279 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 279

Topic #: 1

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

You are creating an ML pipeline for data processing, model training, and model deployment that uses different Google Cloud services. You have developed code for each individual task, and you expect a high frequency of new files. You now need to create an orchestration layer on top of these tasks. You only want this orchestration pipeline to run if new files are present in your dataset in a Cloud Storage bucket. You also want to minimize the compute node costs. What should you do?


- A. Create a pipeline in Vertex AI Pipelines. Configure the first step to compare the contents of the bucket to the last time the pipeline was run. Use the scheduler API to run the pipeline periodically.
- B. Create a Cloud Function that uses a Cloud Storage trigger and deploys a Cloud Composer directed acyclic graph (DAG).
- C. Create a pipeline in Vertex AI Pipelines. Create a Cloud Function that uses a Cloud Storage trigger and deploys the pipeline.
- D. Deploy a Cloud Composer directed acyclic graph (DAG) with a GCSObjectUpdateSensor class that detects when a new file is added to the Cloud Storage bucket.

Show Suggested Answer

by [guilhermebutzke](#) at Feb. 19, 2024, 3:15 p.m.

## Comments

Type your comment...

  **Foxy2021** 1 week, 1 day ago

My answer is D: While C (Cloud Function + Vertex AI Pipelines) is a viable approach for triggering ML pipelines, D (Cloud Composer DAG with GCSObjectUpdateSensor) is the more appropriate and scalable solution when your orchestration spans multiple Google Cloud services and you want to minimize costs by only triggering the pipeline when new files appear.



   upvoted 1 times

  **tardigradum** 2 months, 1 week ago

**Selected Answer: D**

The key here is "that uses different Google Cloud services". Taking this into account, Cloud Composer is the correct answer (for instance, Vertex AI pipelines is not integrated with classic Dataproc or Cloud Composer DAGs). Moreover, GCSObjectUpdateSensor is more efficient than a Cloud Function.


   upvoted 1 times

  **fitri001** 4 months, 4 weeks ago

**Selected Answer: C**

Option C appears to be the best choice for balancing the requirements of efficient orchestration, cost minimization, and ensuring the pipeline only runs when new files are present. By using a Cloud Function triggered by Cloud Storage events to deploy a Vertex AI Pipeline, you can leverage the event-driven model of Cloud Functions to minimize unnecessary runs and associated costs, while still using the powerful orchestration capabilities of Vertex AI Pipelines.

   upvoted 4 times

  **fitri001** 4 months, 4 weeks ago

why not D?

Pros:

Cloud Composer provides a powerful orchestration framework that can handle complex dependencies and workflows. GCSObjectUpdateSensor can efficiently detect new files in the bucket and trigger the pipeline.

Cons:

Cloud Composer can be relatively costly due to the continuous operation of its environment. Overhead of maintaining Cloud Composer for potentially simple file-triggered tasks.

   upvoted 1 times

  **tardigradum** 2 months, 1 week ago

I think we should use Cloud Composer here because of "that uses different Google Cloud services". Vertex AI is less integrated with the rest of services than Cloud Composer, which was designed exactly for that.



   upvoted 1 times

  **Kili1** 5 months ago

**Selected Answer: D**

"Different Google Cloud services" and GCSObjectUpdateSensor: This sensor class specifically checks for updates to Cloud Storage objects. This ensures the DAG only triggers when there's a new file in the bucket, minimizing unnecessary executions.

   upvoted 1 times

  **CHARLIE2108** 7 months ago

Why not D?

   upvoted 1 times

  **Yan\_X** 7 months, 2 weeks ago

**Selected Answer: C**

C

Cloud Function to be triggered by Cloud storage trigger, and then deploy the Vertex AI pipeline.

   upvoted 3 times

  **JG123** 7 months, 2 weeks ago

Its C. Vertex pipelines are recommended to run ML pipeline!




   upvoted 1 times

  **guilhermebutzke** 8 months ago

**Selected Answer: B**

My Answer: B

Cloud Function that uses a Cloud Storage trigger ("run if new files are present in your dataset in a Cloud Storage bucket") and Cloud Composer directed acyclic graph (DAG) ("model deployment that uses different Google Cloud services", "orchestration layer on top of these tasks",)

   upvoted 1 times

 **tavva\_prudhvi** 8 months ago

Cloud Composer already provides a way to orchestrate tasks, and creating a Cloud Function to deploy a DAG is not a common practice. The Cloud Function with a Cloud Storage trigger would be redundant since the GCSObjectUpdateSensor within the DAG itself can handle the file detection.

   upvoted 5 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.



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.