← Google Discussions

Exam Professional Data Engineer All Questions
View all questions & answers for the Professional Data Engineer exam

Go to Exam

📄 **EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 256 DISCUSSION**

Actual exam question from Google's Professional Data Engineer
Question #: 256
Topic #: 1
[All Professional Data Engineer Questions]

You are deploying an Apache Airflow directed acyclic graph (DAG) in a Cloud Composer 2 instance. You have incoming files in a Cloud Storage bucket that the DAG processes, one file at a time. The Cloud Composer instance is deployed in a subnetwork with no Internet access. Instead of running the DAG based on a schedule, you want to run the DAG in a reactive way every time a new file is received. What should you do?

A. 1. Enable Private Google Access in the subnetwork, and set up Cloud Storage notifications to a Pub/Sub topic.
2. Create a push subscription that points to the web server URL.

B. 1. Enable the Cloud Composer API, and set up Cloud Storage notifications to trigger a Cloud Function.
2. Write a Cloud Function instance to call the DAG by using the Cloud Composer API and the web server URL.
3. Use VPC Serverless Access to reach the web server URL.

C. 1. Enable the Airflow REST API, and set up Cloud Storage notifications to trigger a Cloud Function instance.
2. Create a Private Service Connect (PSC) endpoint.
3. Write a Cloud Function that connects to the Cloud Composer cluster through the PSC endpoint.

D. 1. Enable the Airflow REST API, and set up Cloud Storage notifications to trigger a Cloud Function instance.
2. Write a Cloud Function instance to call the DAG by using the Airflow REST API and the web server URL.
3. Use VPC Serverless Access to reach the web server URL.

Show Suggested Answer

by 👤 scaenruy at *Jan. 3, 2024, 4:56 p.m.*

Comments

Type your comment...

Submit

○ 👤 **raaad** `Highly Voted 👍` 1 year, 4 months ago

`Selected Answer: C`

- Enable Airflow REST API: In Cloud Composer, enable the "Airflow web server" option.
- Set Up Cloud Storage Notifications: Create a notification for new files, routing to a Cloud Function.
- Create PSC Endpoint: Establish a PSC endpoint for Cloud Composer.
- Write Cloud Function: Code the function to use the Airflow REST API (via PSC endpoint) to trigger the DAG.

========
Why not Option D
- Using the web server URL directly wouldn't work without internet access or a direct path to the web server.

👍 ↩ 🚩 upvoted 12 times

  ○ 👤 **AllenChen123** 1 year, 3 months ago

  Why not B, use Cloud Composer API

  👍 ↩ 🚩 upvoted 4 times

○ 👤 **STEVE_PEGLEG** `Highly Voted 👍` 9 months ago

`Selected Answer: A`

This is the guidance how to use method in A:
https://cloud.google.com/composer/docs/composer-2/triggering-gcf-pubsub
"In this specific example, you create a Cloud Function and deploy two DAGs. The first DAG pulls Pub/Sub messages and triggers the second DAG according to the Pub/Sub message content."

For C & D, this guidance says it can't be done when you have Private or VPS Service Controls set up:
https://cloud.google.com/composer/docs/composer-2/triggering-with-gcf#check_your_environments_networking_configuration
"This solution does not work in Private IP and VPC Service Controls configurations because it is not possible to configure connectivity from Cloud Functions to the Airflow web server in these configurations."

👍 ↩ 🚩 upvoted 6 times

○ 👤 **aditya_ali** `Most Recent ⊘` 1 day, 7 hours ago

`Selected Answer: C`

PSC is the only secure way to reach the Airflow REST API privately from a serverless service in a VPC-restricted Cloud Composer environment.

Therefore, Option C provides the most secure and functional architecture.

👍 ↩ 🚩 upvoted 1 times

○ 👤 **Augustax** 3 months ago

`Selected Answer: B`

Option B is the only viable solution because:

It uses the Cloud Composer API, which is compatible with Private IP configurations.

It leverages VPC Serverless Access to allow Cloud Functions to securely access the Airflow web server within the subnetwork.

It avoids the limitations of the Airflow REST API in Private IP environments.

👍 ↩ 🚩 upvoted 1 times

○ 👤 **Pime13** 3 months, 3 weeks ago

`Selected Answer: D`

Why Option D is the Best Choice:
Airflow REST API: Enabling the Airflow REST API allows you to programmatically trigger DAG runs, which is essential for a reactive setup.
Cloud Storage Notifications: Setting up notifications ensures that your DAG is triggered every time a new file is received in the Cloud Storage bucket.
VPC Serverless Access: This allows your Cloud Function to securely access the Cloud Composer web server URL without needing external IP addresses, complying with your subnetwork's no Internet access constraint.

👍 ↩ 🚩 upvoted 3 times

○ 👤 b i m h h 4 k

baimus 6 months, 4 weeks ago

This is A, as steve_pegleg says, there is no way to connect the cloud function to the Airflow instance, without first enabling private access. The pubsub pattern makes sense in this context.

👍 ↩ 🏳 upvoted 2 times

---

josech 11 months, 3 weeks ago

C is not correct because "this solution does not work in Private IP and VPC Service Controls configurations because it is not possible to configure connectivity from Cloud Functions to the Airflow web server in these configurations".
https://cloud.google.com/composer/docs/how-to/using/triggering-with-gcf
The correct answer is A using Pub/Sub https://cloud.google.com/composer/docs/composer-2/triggering-gcf-pubsub

👍 ↩ 🏳 upvoted 4 times

---

chrissamharris 1 year, 1 month ago

Why not Option C? C involves creating a Private Service Connect (PSC) endpoint, which, while viable for creating private connections to Google services, adds complexity and might not be required when simpler solutions like VPC Serverless Access (as in Option D) can suffice.

👍 ↩ 🏳 upvoted 3 times

> **chrissamharris** 1 year, 1 month ago
>
> https://cloud.google.com/vpc/docs/serverless-vpc-access: Serverless VPC Access makes it possible for you to connect directly to your Virtual Private Cloud (VPC) network from serverless environments such as Cloud Run, App Engine, or Cloud Functions
>
> 👍 ↩ 🏳 upvoted 2 times

---

d11379b 1 year, 1 month ago

The answer should be D
Serverless VPC Access makes it possible for you to connect directly to your Virtual Private Cloud (VPC) network from serverless environments such as Cloud Run, App Engine, or Cloud Functions. Configuring Serverless VPC Access allows your serverless environment to send requests to your VPC network by using internal DNS and internal IP addresses (as defined by RFC 1918 and RFC 6598). The responses to these requests also use your internal network.
You can use Serverless VPC Access to access Compute Engine VM instances, Memorystore instances, and any other resources with internal DNS or internal IP address.
(Reference: https://cloud.google.com/vpc/docs/serverless-vpc-access)
When you use Airflow Rest API to tigger the job, the url is based on the private IP address of Cloud Composer Instance, so you need to use Serverless VPC Access for it.

👍 ↩ 🏳 upvoted 4 times

> **d11379b** 1 year, 1 month ago
>
> Why not C:
> The reference here (https://cloud.google.com/vpc/docs/private-service-connect#published-services) limits the available use cases:
> Private Service Connect supports access to the following types of managed services:
> Published VPC-hosted services, which include the following:
> Google published services, such as Apigee or the GKE control plane
> Third-party published services provided by Private Service Connect partners
> Intra-organization published services, where the consumer and producer might be two different VPC networks within the same company
> Google APIs, such as Cloud Storage or BigQuery
>
> Unfortunately your airflow Rest API is not published as a service in the list, so you can not use it
> This is also one of the reasons why you should reject A
>
> 👍 ↩ 🏳 upvoted 2 times

> > **d11379b** 1 year, 1 month ago
> >
> > B is not appropriate while Cloud Composer API can really execute Airflow command☐but It's not via web server Url to run a DAG in this case, and I doubt if it is really possible
> >
> > 👍 ↩ 🏳 upvoted 2 times

Matt_108 1 year, 3 months ago

Option C, raaad explained well why

👍 ↩ 🏳 upvoted 1 times

---

scaenruy 1 year, 4 months ago

C.
1. Enable the Airflow REST API, and set up Cloud Storage notifications to trigger a Cloud Function instance.
2. Create a Private Service Connect (PSC) endpoint.
3. Write a Cloud Function that connects to the Cloud Composer cluster through the PSC endpoint.

👍 ↩ 🏳 **upvoted 1 times**