← 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 270 DISCUSSION**

Actual exam question from Google's Professional Data Engineer

Question #: 270

Topic #: 1

[All Professional Data Engineer Questions]

You need to create a SQL pipeline. The pipeline runs an aggregate SQL transformation on a BigQuery table every two hours and appends the result to another existing BigQuery table. You need to configure the pipeline to retry if errors occur. You want the pipeline to send an email notification after three consecutive failures. What should you do?

A. Use the BigQueryUpsertTableOperator in Cloud Composer, set the retry parameter to three, and set the email_on_failure parameter to true.

B. Use the BigQueryInsertJobOperator in Cloud Composer, set the retry parameter to three, and set the email_on_failure parameter to true.

C. Create a BigQuery scheduled query to run the SQL transformation with schedule options that repeats every two hours, and enable email notifications.

D. Create a BigQuery scheduled query to run the SQL transformation with schedule options that repeats every two hours, and enable notification to Pub/Sub topic. Use Pub/Sub and Cloud Functions to send an email after three failed executions.

Show Suggested Answer

by 👤 **scaenruy** at *Jan. 3, 2024, 6:43 p.m.*

Comments

Type your comment...

raaad `Highly Voted 👍` 10 months ago

`Selected Answer: B`

- It provides a direct and controlled way to manage the SQL pipeline using Cloud Composer (Apache Airflow).
- The BigQueryInsertJobOperator is well-suited for running SQL jobs in BigQuery, including aggregate transformations and handling of results.
- The retry and email_on_failure parameters align with the requirements for error handling and notifications.
- Cloud Composer requires more setup than using BigQuery's scheduled queries directly, but it offers robust workflow management, retry logic, and notification capabilities, making it suitable for more complex and controlled data pipeline requirements.

👍 ↩ ⚑ upvoted 6 times

> **SuperVan** 8 months, 2 weeks ago
>
> The prompt wants an email notification sent after three failed attempts. Is there any concern that the retry parameter is set to 3, wouldn't this mean that the email is sent after 4 failed attempts (1 original + 3 retries)?
>
> 👍 ↩ ⚑ upvoted 6 times

gabbferreira `Most Recent ⊙` 1 week, 6 days ago

`Selected Answer: D`

It´s D
"Notifications on operator failure
Set email_on_failure to True to send an email notification when an operator in the DAG fails. To send email notifications from a Cloud Composer environment, you must configure your environment to use SendGrid."

Setting email_on_failure = True will send the email after the FIRST failure

https://cloud.google.com/composer/docs/composer-2/write-dags#notifications_on_operator_failure

👍 ↩ ⚑ upvoted 1 times

gabbferreira 1 week, 6 days ago

`Selected Answer: D`

chatpgt and gemini said it is D

👍 ↩ ⚑ upvoted 1 times

MarcoPellegrino 2 months, 2 weeks ago

`Selected Answer: D`

A) Wrong, Upsert is not for appending
B) Wrong, doesn't mention the 2 hours scheduling
C) Wrong, doesn't mention the emailing
D) Correct

👍 ↩ ⚑ upvoted 2 times

Augustax 3 months ago

`Selected Answer: D`

The retry times in B and clearly mentioned 2 hours in D make me think D is the better option..

👍 ↩ ⚑ upvoted 1 times

plum21 3 months ago

`Selected Answer: D`

"You want the pipeline to send an email notification after three consecutive failures" - it is not about retries which are configurable via Composer operator - it is about 3 consecutive executions which could be for different hours.

👍 ↩ ⚑ upvoted 2 times

b3e59c2 3 months, 4 weeks ago

`Selected Answer: D`

Terrible options as usual. Whilst B is the most elegant, it doesn't explicitly address the 2 hour scheduling (you can schedule within Composer, but the answer doesn't mention it).

If we take these answers on the surface level, D is the only option that actually achieves our goal.

👍 ↩ ⚑ upvoted 1 times

Pime13 3 months, 4 weeks ago

`Selected Answer: B`

B. Use the BigQueryInsertJobOperator in Cloud Composer, set the retry parameter to three, and set the email_on_failure parameter to true.

This approach leverages Cloud Composer (Apache Airflow) to manage the SQL pipeline, providing robust workflow

This approach leverages Cloud Composer (Apache Airflow) to manage the SQL pipeline, providing robust workflow management, retry logic, and notification capabilities. By setting the retry parameter to three and enabling email notifications, you ensure that the pipeline will retry on errors and notify you after three consecutive failures.

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **e593506** 4 months, 2 weeks ago

Selected Answer: D

The prompt wants an email notification sent after three failed attempts
Option B does not meet that condition

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **josech** 5 months, 2 weeks ago

Selected Answer: B

https://airflow.apache.org/docs/apache-airflow-providers-
google/stable/_api/airflow/providers/google/cloud/operators/bigquery/index.html#airflow.providers.google.cloud.operators.bigquer
y.BigQueryInsertJobOperator
https://cloud.google.com/composer/docs/composer-2/write-dags#notifications_on_operator_failure

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **joao_01** 7 months ago

It s B (however for me its a incomplete answers cause it does not address the schedule of every 2 hours).

Its not C or D because BigQuery scheduled queries by default does not retries the queries when error occurs. Link:
https://cloud.google.com/bigquery/docs/scheduling-queries

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **JyoGCP** 8 months, 2 weeks ago

Selected Answer: B

Option B

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **datapassionate** 9 months, 1 week ago

Selected Answer: D

D. Create a BigQuery scheduled query to run the SQL transformation with schedule options that repeats every two hours, and enable notification to Pub/Sub topic. Use Pub/Sub and Cloud Functions to send an email after three failed executions

This method utilizes BigQuery's native scheduling capabilities for running the SQL job and leverages Pub/Sub and Cloud Functions for customized notification handling, including the specific requirement of sending an email after three consecutive failures.

👍 ↩ 🏳 upvoted 2 times

☐ 👤 **RenePetersen** 8 months, 2 weeks ago

Option D mentions nothing about how the job retrying is put in place, so for that reason I don't think this is the correct option.

👍 ↩ 🏳 upvoted 3 times

☐ 👤 **scaenruy** 10 months ago

Selected Answer: B

B. Use the BigQueryInsertJobOperator in Cloud Composer, set the retry parameter to three, and set the email_on_failure parameter to true.

👍 ↩ 🏳 upvoted 1 times