← 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 289 DISCUSSION**

Actual exam question from Google's Professional Data Engineer

Question #: 289

Topic #: 1

[All Professional Data Engineer Questions]

You have data located in BigQuery that is used to generate reports for your company. You have noticed some weekly executive report fields do not correspond to format according to company standards. For example, report errors include different telephone formats and different country code identifiers. This is a frequent issue, so you need to create a recurring job to normalize the data. You want a quick solution that requires no coding. What should you do?

A. Use Cloud Data Fusion and Wrangler to normalize the data, and set up a recurring job.

B. Use Dataflow SQL to create a job that normalizes the data, and that after the first run of the job, schedule the pipeline to execute recurrently.

C. Create a Spark job and submit it to Dataproc Serverless.

D. Use BigQuery and GoogleSQL to normalize the data, and schedule recurring queries in BigQuery.

Show Suggested Answer

by 👤 **scaenruy** at *Jan. 4, 2024, 11:12 a.m.*

Comments

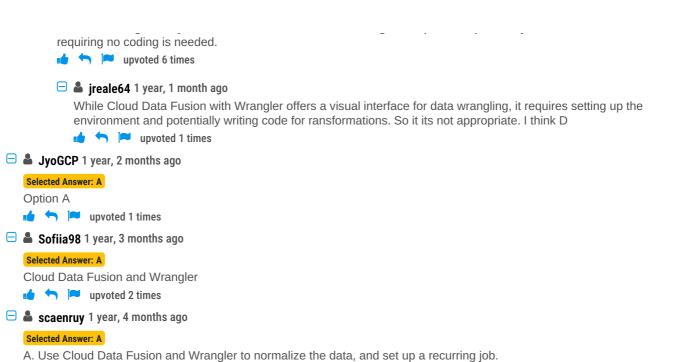| Type your comment... |
| --- |

Submit

⊟ 👤 **Matt_108** `Highly Voted 👍` **1 year, 3 months ago**

`Selected Answer: A`

Definitely A, cloud data fusion and wrangler to setup the clean up pipeline with no coding required

👍 ↩ ⚑ upvoted 8 times

⊟ 👤 **marlon.andrei** `Most Recent ⊘` **3 months, 4 weeks ago**

`Selected Answer: D`

The question say "You want a quick solution that requires no coding.". The data is in BQ, then is most easy normalize the data, and schedule recurring queries in BigQuery.

👍 ↩ ⚑ upvoted 2 times

⊟ 👤 **987af6b** **9 months, 2 weeks ago**

`Selected Answer: A`

A. Use Cloud Data Fusion and Wrangler to normalize the data, and set up a recurring job.

Explanation
No Coding Required: Cloud Data Fusion's Wrangler offers a no-code interface for data transformation tasks. You can visually design data normalization workflows without writing any code.
Recurring Jobs: Cloud Data Fusion allows you to schedule these data normalization tasks to run on a recurring basis, meeting your need for automation.

👍 ↩ ⚑ upvoted 2 times

⊟ 👤 **carmltekai** **9 months, 3 weeks ago**

`Selected Answer: D`

The best solution here is D. Use BigQuery and GoogleSQL to normalize the data, and schedule recurring queries in BigQuery.

Here's why:

* No-code solution: BigQuery's built-in capabilities and GoogleSQL offer a no-code way to transform and standardize data. You can leverage functions like REGEXP_REPLACE to normalize phone numbers and FORMAT to ensure consistent formatting across fields.
* Recurring jobs: BigQuery allows you to schedule queries to run regularly, which is perfect for maintaining data consistency over time.
* Quick and efficient: BigQuery is designed for large-scale data processing, making it fast and efficient for normalization tasks.

👍 ↩ ⚑ upvoted 2 times

  ⊟ 👤 **carmltekai** **9 months, 3 weeks ago**

  Why other options aren't as suitable:

  A. Cloud Data Fusion and Wrangler: While powerful, these tools might be overkill for a simple normalization task and could involve a steeper learning curve.
  B. Dataflow SQL: Dataflow is primarily for stream processing and might not be the most efficient for batch transformations on data already in BigQuery.
  C. Dataproc Serverless: This involves using a Spark job, which requires coding and might be more complex than necessary for this task.

  👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **fitri001** **10 months, 3 weeks ago**

`Selected Answer: A`

https://cloud.google.com/data-fusion/docs

👍 ↩ ⚑ upvoted 2 times

⊟ 👤 **SohiniV** **1 year, 2 months ago**

As per chatGPT, Option D allows you to utilize BigQuery's SQL capabilities to write queries that normalize the data according to company standards.
You can then schedule these queries to run on a recurring basis using BigQuery's scheduled queries feature. This feature allows you to specify a schedule (e.g., weekly) for executing SQL queries automatically.
This approach requires no additional setup or coding outside of BigQuery, making it a quick and straightforward solution to address the issue of data normalization.

👍 ↩ ⚑ upvoted 1 times

  ⊟ 👤 **SohiniV** **1 year, 2 months ago**

  Any views on this ?

  👍 ↩ ⚑ upvoted 1 times

    ⊟ 👤 **RenePetersen** **1 year, 2 months ago**

    Wouldn't writing the SQL transformation be considered coding? The question specifically states that a solution

requiring no coding is needed.

👍 ↩ 🚩 upvoted 6 times

⊟ 👤 **jreale64** 1 year, 1 month ago

While Cloud Data Fusion with Wrangler offers a visual interface for data wrangling, it requires setting up the environment and potentially writing code for ransformations. So it its not appropriate. I think D

👍 ↩ 🚩 upvoted 1 times

⊟ 👤 **JyoGCP** 1 year, 2 months ago

Selected Answer: A

Option A

👍 ↩ 🚩 upvoted 1 times

⊟ 👤 **Sofiia98** 1 year, 3 months ago

Selected Answer: A

Cloud Data Fusion and Wrangler

👍 ↩ 🚩 upvoted 2 times

⊟ 👤 **scaenruy** 1 year, 4 months ago

Selected Answer: A

A. Use Cloud Data Fusion and Wrangler to normalize the data, and set up a recurring job.

👍 ↩ 🚩 upvoted 2 times