



PR OFESSIIONAL-DATA-ENGINEER^{Q&As}

Professional Data Engineer on Google Cloud Platform

**Pass Google PROFESSIONAL-DATA-ENGINEER
Exam with 100% Guarantee**

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.passapply.com/professional-data-engineer.html>

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Google
Official Exam Center



VCE & PDF

PassApply.com

<https://www.passapply.com/professional-data-engineer.html>
2024 Latest passapply PROFESSIONAL-DATA-ENGINEER PDF and VCE
dumps Download

- ⚙️ **Instant Download** After Purchase
- ⚙️ **100% Money Back** Guarantee
- ⚙️ **365 Days** Free Update
- ⚙️ **800,000+** Satisfied Customers





QUESTION 1

You have a job that you want to cancel. It is a streaming pipeline, and you want to ensure that any data that is in-flight is processed and written to the output. Which of the following commands can you use on the Dataflow monitoring console to stop the pipeline job?

- A. Cancel
- B. Drain
- C. Stop
- D. Finish

Correct Answer: B

Using the Drain option to stop your job tells the Dataflow service to finish your job in its current state. Your job will immediately stop ingesting new data from input sources, but the Dataflow service will preserve any existing resources (such as worker instances) to finish processing and writing any buffered data in your pipeline.

Reference: <https://cloud.google.com/dataflow/pipelines/stopping-a-pipeline>

QUESTION 2

You work for a large fast food restaurant chain with over 400,000 employees. You store employee information in Google BigQuery in a Users table consisting of a FirstName field and a LastName field. A member of IT is building an application and asks you to modify the schema and data in BigQuery so the application can query a FullName field consisting of the value of the FirstName field concatenated with a space, followed by the value of the LastName field for each employee. How can you make that data available while minimizing cost?

- A. Create a view in BigQuery that concatenates the FirstName and LastName field values to produce the FullName.
- B. Add a new column called FullName to the Users table. Run an UPDATE statement that updates the FullName column for each user with the concatenation of the FirstName and LastName values.
- C. Create a Google Cloud Dataflow job that queries BigQuery for the entire Users table, concatenates the FirstName value and LastName value for each user, and loads the proper values for FirstName, LastName, and FullName into a new table in BigQuery.
- D. Use BigQuery to export the data for the table to a CSV file. Create a Google Cloud Dataproc job to process the CSV file and output a new CSV file containing the proper values for FirstName, LastName and FullName. Run a BigQuery load job to load the new CSV file into BigQuery.

Correct Answer: B

BigQuery has no quota on the DML statements. (Search Google - does bigquery have quota for update).

Why not C: This is a one time activity and SQL is the easiest way to program it. DataFlow is way overkill for this. You will need to find an engineer who can develop DataFlow pipelines. Whereas, SQL is so much more widely known and easier. One of the great features about BigQuery is its SQL interface. Even for BigQueryML services.



QUESTION 3

An organization maintains a Google BigQuery dataset that contains tables with user-level data. They want to expose aggregates of this data to other Google Cloud projects, while still controlling access to the user-level data. Additionally, they need to minimize their overall storage cost and ensure the analysis cost for other projects is assigned to those projects. What should they do?

- A. Create and share an authorized view that provides the aggregate results.
- B. Create and share a new dataset and view that provides the aggregate results.
- C. Create and share a new dataset and table that contains the aggregate results.
- D. Create dataViewer Identity and Access Management (IAM) roles on the dataset to enable sharing.

Correct Answer: A

Reference: <https://cloud.google.com/bigquery/docs/access-control>

QUESTION 4

You are collecting IoT sensor data from millions of devices across the world and storing the data in BigQuery. Your access pattern is based on recent data filtered by location_id and device_version with the following query:

```
SELECT
  MAX(temperature)
FROM
  acme_iot_data.sensors
WHERE
  create_date > DATE_SUB(CURRENT_DATE(), INTERVAL 7 day)
  AND location_id = "SW1W9TQ"
  AND device_version = "202007r3"
```

You want to optimize your queries for cost and performance. How should you structure your data?

- A. Partition table data by create_date, location_id and device_version
- B. Partition table data by create_date cluster table data by location_id and device_version
- C. Cluster table data by create_date location_id and device_version
- D. Cluster table data by create_date, partition by location and device_version

Correct Answer: C

QUESTION 5

You are designing storage for two relational tables that are part of a 10-TB database on Google Cloud. You want to support transactions that scale horizontally. You also want to optimize data for range queries on nonkey columns. What



should you do?

- A. Use Cloud SQL for storage. Add secondary indexes to support query patterns.
- B. Use Cloud SQL for storage. Use Cloud Dataflow to transform data to support query patterns.
- C. Use Cloud Spanner for storage. Add secondary indexes to support query patterns.
- D. Use Cloud Spanner for storage. Use Cloud Dataflow to transform data to support query patterns.

Correct Answer: C

Reference: <https://cloud.google.com/solutions/data-lifecycle-cloud-platform>

[Latest PROFESSIONAL-DATA-ENGINEER Dumps](#)

[PROFESSIONAL-DATA-ENGINEER VCE Dumps](#)

[PROFESSIONAL-DATA-ENGINEER Practice Test](#)