



# PR PROFESSIONAL-DATA-ENGINEER<sup>Q&As</sup>

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.geekcert.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



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





### QUESTION 1

What are two of the benefits of using denormalized data structures in BigQuery?

- A. Reduces the amount of data processed, reduces the amount of storage required
- B. Increases query speed, makes queries simpler
- C. Reduces the amount of storage required, increases query speed
- D. Reduces the amount of data processed, increases query speed

Correct Answer: B

Denormalization increases query speed for tables with billions of rows because BigQuery's performance degrades when doing JOINS on large tables, but with a denormalized data structure, you don't have to use JOINS, since all of the data has been combined into one table. Denormalization also makes queries simpler because you do not have to use JOIN clauses.

Denormalization increases the amount of data processed and the amount of storage required because it creates redundant data.

Reference: [https://cloud.google.com/solutions/bigquery-data-warehouse#denormalizing\\_data](https://cloud.google.com/solutions/bigquery-data-warehouse#denormalizing_data)

---

### QUESTION 2

Your company built a TensorFlow neural-network model with a large number of neurons and layers. The model fits well for the training data. However, when tested against new data, it performs poorly. What method can you employ to address this?

- A. Threading
- B. Serialization
- C. Dropout Methods
- D. Dimensionality Reduction

Correct Answer: C

Reference <https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877>

---

### QUESTION 3

Does Dataflow process batch data pipelines or streaming data pipelines?

- A. Only Batch Data Pipelines
- B. Both Batch and Streaming Data Pipelines



- C. Only Streaming Data Pipelines
- D. None of the above

Correct Answer: B

Dataflow is a unified processing model, and can execute both streaming and batch data pipelines Reference:  
<https://cloud.google.com/dataflow/>

#### QUESTION 4

You are migrating a table to BigQuery and are deciding on the data model. Your table stores information related to purchases made across several store locations and includes information like the time of the transaction, items purchased, the store ID and the city and state in which the store is located You frequently query this table to see how many of each item were sold over the past 30 days and to look at purchasing trends by state city and individual store. You want to model this table to minimize query time and cost. What should you do?

- A. Partition by transaction time; cluster by state first, then city then store ID
- B. Partition by transaction time cluster by store ID first, then city, then state
- C. Top-level cluster by state first, then city then store
- D. Top-level cluster by store ID first, then city then state.

Correct Answer: A

<https://cloud.google.com/bigquery/docs/partitioned-tables>

This page provides an overview of partitioned tables in BigQuery. A partitioned table is a special table that is divided into segments, called partitions, that make it easier to manage and query your data. By dividing a large table into smaller partitions, you can improve query performance, and you can control costs by reducing the number of bytes read by a query.

You can partition BigQuery tables by:

-Time-unit column: Tables are partitioned based on a `TIMESTAMP`, `DATE`, or `DATETIME` column in the table.

<https://cloud.google.com/bigquery/docs/clustered-tables>

Clustered tables in BigQuery are tables that have a user-defined column sort order using clustered columns. Clustered tables can improve query performance and reduce query costs.

#### QUESTION 5

You have enabled the free integration between Firebase Analytics and Google BigQuery. Firebase now automatically creates a new table daily in BigQuery in the format `app_events_YYYYMMDD`. You want to query all of the tables for the past 30 days in legacy SQL. What should you do?

- A. Use the `TABLE_DATE_RANGE` function
- B. Use the `WHERE_PARTITIONTIME` pseudo column



C. Use WHERE date BETWEEN YYYY-MM-DD AND YYYY-MM-DD

D. Use SELECT IF.(date >= YYYY-MM-DD AND date