



VCE & PDF

GeekCert.com

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

PROFESSIONAL-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.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



VCE & PDF

GeekCert.com

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

2024 Latest geekcert 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've migrated a Hadoop job from an on-premises cluster to Dataproc and Good Storage. Your Spark job is a complex analytical workload that consists of many shuffling operations, and initial data are parquet files (on average 200-400 MB size each). You see some degradation in performance after the migration to Dataproc so you'd like to optimize for it. Your organization is very cost-sensitive so you'd like to continue using Dataproc on preemptibles (with 2 non-preemptible workers only) for this workload. What should you do?

- A. Switch from HDDs to SSDs, override the preemptible VMs configuration to increase the boot disk size
- B. Increase the size of your parquet files to ensure them to be 1 GB minimum
- C. Switch to TFRecords format (approx 200 MB per file) instead of parquet files
- D. Switch from HDDs to SSDs, copy initial data from Cloud Storage to Hadoop Distributed File System (HDFS), run the Spark job and copy results back to Cloud Storage

Correct Answer: A

QUESTION 2

You're using Bigtable for a real-time application, and you have a heavy load that is a mix of read and writes. You've recently identified an additional use case and need to perform hourly an analytical job to calculate certain statistics across the whole database. You need to ensure both the reliability of your production application as well as the analytical workload.

What should you do?

- A. Export Bigtable dump to GCS and run your analytical job on top of the exported files.
- B. Add a second cluster to an existing instance with a multi-cluster routing, use live-traffic app profile for your regular workload and batch-analytics profile for the analytics workload.
- C. Add a second cluster to an existing instance with a single-cluster routing, use live-traffic app profile for your regular workload and batch-analytics profile for the analytics workload.
- D. Increase the size of your existing cluster twice and execute your analytics workload on your new resized cluster.

Correct Answer: B

QUESTION 3

You are building an application to share financial market data with consumers, who will receive data feeds. Data is collected from the markets in real time. Consumers will receive the data in the following ways:

1.

Real-time event stream

2.



ANSI SQL access to real-time stream and historical data Batch historical exports

Which solution should you use?

- A. Cloud Dataflow, Cloud SQL, Cloud Spanner
- B. Cloud Pub/Sub, Cloud Storage, BigQuery
- C. Cloud Dataproc, Cloud Dataflow, BigQuery
- D. Cloud Pub/Sub, Cloud Dataproc, Cloud SQL

Correct Answer: A

QUESTION 4

You are designing the database schema for a machine learning-based food ordering service that will predict what users want to eat. Here is some of the information you need to store:

The user profile: What the user likes and doesn't like to eat
The user account information: Name, address, preferred meal times
The order information: When orders are made, from where, to whom

The database will be used to store all the transactional data of the product. You want to optimize the data schema. Which Google Cloud Platform product should you use?

- A. BigQuery
- B. Cloud SQL
- C. Cloud Bigtable
- D. Cloud Datastore

Correct Answer: A

QUESTION 5

You are designing a pipeline that publishes application events to a Pub/Sub topic. You need to aggregate events across hourly intervals before loading the results to BigQuery for analysis. Your solution must be scalable so it can process and load large volumes of events to BigQuery. What should you do?

- A. Create a streaming Dataflow job to continually read from the Pub/Sub topic and perform the necessary aggregations using tumbling windows
- B. Schedule a batch Dataflow job to run hourly, pulling all available messages from the Pub-Sub topic and performing the necessary aggregations
- C. Schedule a Cloud Function to run hourly, pulling all available messages from the Pub/Sub topic and performing the necessary aggregations
- D. Create a Cloud Function to perform the necessary data processing that executes using the Pub/Sub trigger every



time a new message is published to the topic.

Correct Answer: A

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

[PROFESSIONAL-DATA-
ENGINEER Exam
Questions](#)

[PROFESSIONAL-DATA-
ENGINEER Braindumps](#)