



# 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

Flowlogistic's management has determined that the current Apache Kafka servers cannot handle the data volume for their real-time inventory tracking system. You need to build a new system on Google Cloud Platform (GCP) that will feed the proprietary tracking software. The system must be able to ingest data from a variety of global sources, process and query in real-time, and store the data reliably.

Which combination of GCP products should you choose?

- A. Cloud Pub/Sub, Cloud Dataflow, and Cloud Storage
- B. Cloud Pub/Sub, Cloud Dataflow, and Local SSD
- C. Cloud Pub/Sub, Cloud SQL, and Cloud Storage
- D. Cloud Load Balancing, Cloud Dataflow, and Cloud Storage

Correct Answer: A

---

### QUESTION 2

Which of these operations can you perform from the BigQuery Web UI?

- A. Upload a file in SQL format.
- B. Load data with nested and repeated fields.
- C. Upload a 20 MB file.
- D. Upload multiple files using a wildcard.

Correct Answer: B

You can load data with nested and repeated fields using the Web UI. You cannot use the Web UI to:

1.  
Upload a file greater than 10 MB in size
2.  
Upload multiple files at the same time
3.  
Upload a file in SQL format

All three of the above operations can be performed using the "bq" command.

Reference: <https://cloud.google.com/bigquery/loading-data>

---



### QUESTION 3

Which of the following statements is NOT true regarding Bigtable access roles?

- A. Using IAM roles, you cannot give a user access to only one table in a project, rather than all tables in a project.
- B. To give a user access to only one table in a project, grant the user the Bigtable Editor role for that table.
- C. You can configure access control only at the project level.
- D. To give a user access to only one table in a project, you must configure access through your application.

Correct Answer: B

For Cloud Bigtable, you can configure access control at the project level. For example, you can grant the ability to:

1.

Read from, but not write to, any table within the project.

2.

Read from and write to any table within the project, but not manage instances.

3.

Read from and write to any table within the project, and manage instances. Reference:

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

---

### QUESTION 4

You plan to deploy Cloud SQL using MySQL. You need to ensure high availability in the event of a zone failure. What should you do?

- A. Create a Cloud SQL instance in one zone, and create a failover replica in another zone within the same region.
- B. Create a Cloud SQL instance in one zone, and create a read replica in another zone within the same region.
- C. Create a Cloud SQL instance in one zone, and configure an external read replica in a zone in a different region.
- D. Create a Cloud SQL instance in a region, and configure automatic backup to a Cloud Storage bucket in the same region.

Correct Answer: A

---

### QUESTION 5

You've migrated a Hadoop job from an on-prem cluster to dataproc and GCS. Your Spark job is a complicated 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. You need to keep in mind that 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. Increase the size of your parquet files to ensure them to be 1 GB minimum.
- B. Switch to TFRecords formats (appr. 200MB per file) instead of parquet files.
- C. Switch from HDDs to SSDs, copy initial data from GCS to HDFS, run the Spark job and copy results back to GCS.
- D. Switch from HDDs to SSDs, override the preemptible VMs configuration to increase the boot disk size.

Correct Answer: D

It's the recommended best practice for this scenario. [https://cloud.google.com/architecture/hadoop/migrating-apache-spark-jobs-to-cloud-dataproc#optimize\\_performance](https://cloud.google.com/architecture/hadoop/migrating-apache-spark-jobs-to-cloud-dataproc#optimize_performance)

[https://cloud.google.com/architecture/hadoop/migrating-apache-spark-jobs-to-cloud-dataproc#switch\\_to\\_ssd\\_disks](https://cloud.google.com/architecture/hadoop/migrating-apache-spark-jobs-to-cloud-dataproc#switch_to_ssd_disks) If you perform many shuffling operations or partitioned writes, switch to SSDs to boost performance.

[https://cloud.google.com/architecture/hadoop/migrating-apache-spark-jobs-to-cloud-dataproc#use\\_preemptible\\_vms](https://cloud.google.com/architecture/hadoop/migrating-apache-spark-jobs-to-cloud-dataproc#use_preemptible_vms) As a default, preemptible VMs are created with a smaller boot disk size, and you might want to override this configuration if you are running shuffle-heavy workloads. For details, see the page on preemptible VMs in the Dataproc documentation.

Elimination Strategy:

- A. Increase the size of your parquet files to ensure them to be 1 GB minimum (doesn't make sense as the file size are fit for migration to proceed with given scenario, recommended size is between 128 MB to 1 GB.)
- B. Switch to TFRecords formats (appr. 200MB per file) instead of parquet files(doesn't make sense to make changes to file format )
- C. Switch from HDDs to SSDs, copy initial data from GCS to HDFS, run the Spark job and copy results back to GCS(doesn't make sense to copy the file from GCS to HDFS as the workload that consists of many shuffling operations)
- D. Switch from HDDs to SSDs, override the preemptible VMs configuration to increase the boot disk size(perfect fit as the workload that consists of many shuffling operations which requires attention to increase the performance reference doc:- [https://cloud.google.com/architecture/hadoop/migrating-apache-spark-jobs-to-cloud-dataproc#optimize\\_performance](https://cloud.google.com/architecture/hadoop/migrating-apache-spark-jobs-to-cloud-dataproc#optimize_performance) )

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

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

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