

PRACTICE QUESTIONS

Google Cloud

Professional Data Engineer

SHARED BY [u/Silent-Cluster-7116](#) on aiexamcerts

EDITION	Free Set 1
QUESTIONS	10
SHARED	October 15, 2026
EXAM	Google Cloud Professional Data Engineer

Shared by [u/Silent-Cluster-7116](#) on aiexamcerts

10 practice questions for the Google Cloud Professional Data Engineer exam, each followed by its answer and a short explanation. More free sets and the full practice exam:

www.aiexamcerts.com/google-cloud/professional-data-engineer. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. A team must choose between batch and streaming for a fraud signal that must react within seconds. Which approach fits the requirement?

- A A nightly BigQuery scheduled query
- B A streaming pipeline in Dataflow reading from Pub/Sub
- C A weekly Dataproc job reading files from Cloud Storage
- D A Cloud Composer DAG that runs every hour

Answer: B

Seconds-level reaction needs continuous processing, which a streaming Dataflow pipeline on Pub/Sub provides. A scheduled query is the tempting distractor because it is managed and simple, but it runs on a fixed interval measured in hours. Weekly and hourly batch options cannot meet the latency target.

2. The managed service that runs Apache Beam pipelines in streaming and batch modes on Google Cloud is _____.

Answer: Dataflow (also accepted: Cloud Dataflow)

Dataflow executes Apache Beam pipelines with autoscaling and managed workers. Dataproc is a tempting alternative because it also runs data processing jobs, but it runs Spark and Hadoop. Accept either Dataflow or Cloud Dataflow.

3. Which Pub/Sub feature preserves the order of messages that share the same key, such as a device ID?

- A Ordering keys with message ordering enabled on the publisher and subscription
- B Dead-letter topics
- C Message retention snapshots
- D Push subscriptions with a higher ack deadline

Answer: A

Ordering keys guarantee order within a key when ordering is enabled on both sides. Dead-letter topics are a tempting distractor because they handle failed messages, but they do not preserve order. Retention snapshots and ack deadlines control replay and redelivery timing.

4. A BigQuery partitioned table always scans every partition when a query filters on the partition column.
- A True
 - B False

Answer: B (False)

Filters on the partition column allow BigQuery to prune partitions and skip the rest. The full-scan belief is the tempting misconception that appears when filters are missing or wrapped in functions. The answer is False.

5. Which two actions help a Dataflow pipeline handle a hot key that slows the whole job? (Choose two.)
- A Increase the Pub/Sub retention period to seven days
 - B Remove all windowing so every event is processed alone
 - C Add a salt or secondary key to spread the work across workers
 - D Switch the job from streaming to a single-threaded batch run
 - E Combine values with a combiner before the grouping step

Answer: C, E

Salting spreads a skewed key across workers, and combiners reduce the data sent for grouping. Increasing retention changes replay, not skew, which is a tempting distractor because it sounds like a buffer. Removing windowing and single-threaded batch runs make skew and throughput worse.

6. Which BigQuery table option forces every query to include a partition filter so an accidental full scan is rejected?
- A `partition_expiration_days`
 - B `cluster_by`
 - C `enable_refresh`
 - D `require_partition_filter`

Answer: D

Setting `require_partition_filter` makes queries without a partition predicate fail, which guards against unintended cost. Partition expiration is the tempting distractor because it is a partition setting, but it deletes old partitions rather than enforcing filters. `cluster_by` and `enable_refresh` do not enforce query filters.

7. A data scientist wants to train a logistic regression model directly with SQL, without exporting data from BigQuery. Which statement does this?
- A CREATE TABLE with a regression schema and a model_type column that selects the algorithm
 - B CREATE VIEW using the LOGISTIC function
 - C CREATE MODEL with model_type set to LOGISTIC_REG
 - D EXPORT DATA to a Vertex AI notebook

Answer: C

BigQuery ML trains models with CREATE MODEL, and LOGISTIC_REG is one supported model type. CREATE TABLE is a tempting distractor because the model is stored in a dataset, but it does not train anything. Views and exports do not train models inside BigQuery.

8. The BigQuery statement that inserts, updates, or deletes rows in a target table based on a source query is _____.

Answer: MERGE

MERGE applies conditional insert, update, and delete actions in one statement. INSERT is a tempting alternative because it writes rows too, but it cannot update matches.

9. A pipeline must run only after an upstream table has been loaded, and it must retry failed steps. Which orchestration approach fits a multi-step dependency graph?
- A A single cron entry on a VM that sleeps for an hour
 - B Cloud Composer DAG with task dependencies and retries
 - C A manual run from a laptop after checking email
 - D A Pub/Sub topic with no subscribers

Answer: B

Composer DAGs express task dependencies and retry policies, which suits multi-step pipelines. The cron option is a tempting distractor because scheduling is part of the job, but it does not manage dependencies or retries. Manual runs and unsubscribed topics do not automate the workload.

10. A team wants to know who ran a BigQuery query that scanned an unexpectedly large table. Which source provides this audit detail?
- A The Cloud Storage object lifecycle report for the bucket that backs the external table and its storage classes
 - B The Dataproc cluster initialization logs
 - C Cloud Audit Logs, data access logs for BigQuery
 - D The Pub/Sub subscription acknowledgment count

Answer: C

BigQuery data access audit logs record the principal, query, and bytes processed. Storage lifecycle reports are a tempting distractor because they describe object management, but they do not show query activity. Cluster and subscription logs cover other services.