



Navigation


Read the [Linux Academy Oath \(https://go.acloudguru.com/faq-la-customer-oath\)](https://go.acloudguru.com/faq-la-customer-oath) from Sam Kroonenburg, our CEO.

[See What the LA Product Team Has Been Up To! \(/updates\)](#)

Google Certified Professional Data Engineer - Data Analytics with BigQuery Quiz

 30
minutes

 10
Questions

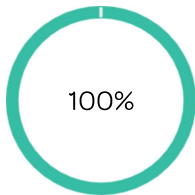
 3
Minutes
per
Question

[Advanced \(/search?type=Practice Exam Challenge&difficulty=Advanced&categories=Cloud\).](#)

Go Back

Start Challenge

Question List [Show All Answers](#)



[← Go Back](#)

Congratulations!

You passed this challenge on this attempt.

Expectations Report Card

Google Certified Professional Data Engineer - Data Analytics with BigQuery Quiz

100%

Exam Breakdown

Google Certified Professional Data Engineer - Data Analytics with BigQuery Quiz



1. What could you consider to improve BigQuery performance for queries that use filters or aggregation against particular columns?



A Use partitioned tables

B Use clustered partitioned tables

C Normalize data based on commonly used filters.

D Use clustering on non-partitioned tables.

Correct Answer: B

Why is this correct?

Clustering can improve the performance of certain types of queries such as queries that use filter clauses and queries that aggregate data. When you submit a query that contains a clause that filters data based on the clustering columns, BigQuery uses the sorted blocks to eliminate scans of unnecessary data.

2. How can you ensure that table modifications in BigQuery are ACID compliant?



A Add a nominal wait time to any application queries following an update to allow for eventual consistency.

B Maintain a separate table that records transactions themselves so changes can be re-applied if any are lost.

C No special accommodations are required, BigQuery table modifications are ACID compliant by design.

D Enforce ordering of SQL statements to ensure the operation completes atomically.

Correct Answer: C

Why is this correct?

All table modifications in BigQuery are ACID compliant. This applies to DML operations, queries with destination tables, and load jobs. A table that goes through inserts, updates, and deletes while serving user queries handles the concurrency gracefully and transitions from one state to the next in an atomic fashion.

3. What steps do you need to take to set up BigQuery before use?



A You must create a BigQuery cluster.

B You must create processing nodes in Compute Engine.

C BigQuery is a serverless product and all compute and storage resources are managed for you.

D You must create storage buckets for BigQuery to use.

Correct Answer: C

Why is this correct?

BigQuery is a serverless product for storing and querying massive data sets without configuring or managing hardware or software.



4. What are some ways to help control costs in BigQuery?

A There are no ways to control the cost of BigQuery.

B Split data across multiple projects to benefit from multiple free tier allowances.

C Only query specific columns, avoiding SELECT *.

D Use LIMIT clauses in SQL queries.

E Preview queries before running them.

Correct Answer: C

Why is this correct?

Querying only the data you need and using the query validator will help you to predict and control costs.

Correct Answer: E

Why is this correct?

Querying only the data you need and using the query validator will help you to predict and control costs.

5. When you run a query in BigQuery, what happens to the results?



A Query results are either written to a destination table specified by the user, or to a temporary cached results table.

B Query results exist only in BigQuery memory and are not stored in a table.

C Query results exist only within the BigQuery UI and are not stored in a table.

D Query results are stored in a results table, however it can not be used for querying cached results.

Correct Answer: A

Why is this correct?

BigQuery writes all query results to a table. The table is either explicitly identified by the user (a destination table), or it is a temporary, cached results table.

6. Your company stores data in BigQuery for long-term and short-term analytics queries. Most of the jobs only need to study the last 7 days of data. Over time, the cost of queries keeps going up. How can you redesign the database to lower the cost of the most frequent queries?



A Use Integer range partitioned tables.

B Create a new table every 7 days. Maintain a separate table that duplicates the weekly tables to contain all data.

C Use DATE partitioned tables.

D Create a new table every 7 days. Use JOIN statements to conduct long-term analytics queries.

Correct Answer: C

Why is this correct?

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.

7. You are required to share a subset of data from a BigQuery data set with a 3rd party analytics team. The data may change over time, and you should not grant unnecessary projects permissions to this team if you can avoid it. How should you proceed?



A Add the team to your GCP project and assign them the BigQuery Data Viewer IAM role for the data set.

B Create an export of the data subset to a Cloud Storage bucket. Provide a signed URL for the team to download the data from the bucket.

C Add the team to your GCP project and assign them the BigQuery Data Editor IAM role for the data set.

D Create an authorized view based on a specific query for the subset of data, and provide access for the team only to that view.

Correct Answer: D

Why is this correct?

An authorized view allows you to share query results with particular users and groups without giving them access to the underlying tables.

8. You are required to load a large volume of data into BigQuery that does contain some duplication. What should you do to ensure the best query performance once the data is loaded?



A Normalize the data to eliminate duplicates from being stored, and reformat the data as Avro.

B Normalize the data to eliminate duplicates from being stored, and reformat the data as JSON.

C Denormalize the data by using nested or repeated fields.

D Break up the data into multiple CSV files.

Correct Answer: C

Why is this correct?

BigQuery performs best when your data is denormalized. Rather than preserving a relational schema, such as a star or snowflake schema, you can improve performance by denormalizing your data and taking advantage of nested and

repeated fields.

9. When do BigQuery jobs run?



- A Jobs run for large batch operations, but not for interactive queries in the BigQuery UI.
- B Jobs run for most actions, including those in the BigQuery UI but only if they are submitted with DML.
- C Jobs run for all scheduled actions, but not for interactive queries in the BigQuery UI.
- D Jobs run for all BigQuery actions including loading, exporting, querying or copying data.**

Correct Answer: D

Why is this correct?

When you use the Cloud Console, the classic BigQuery web UI, or the CLI to load, export, query, or copy data, a job resource is automatically created, scheduled, and run.

10. What is a federated data source?



- A A public BigQuery data set.
- B An external data source that can be queried directly even though the data is not stored in BigQuery.**
- C A BigQuery data set that belongs to a different GCP project, used in an SQL statement.
- D A BigQuery table that belongs to a different data set, used in an SQL statement.

Correct Answer: B

Why is this correct?

An external data source (also known as a federated data source) is a data source that you can query directly even though the data is not stored in BigQuery. Instead of loading or streaming the data, you create a table that references the external data source.
