

PR OFESSIONAL-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 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 are designing a cloud-native historical data processing system to meet the following conditions:

1.

The data being analyzed is in CSV, Avro, and PDF formats and will be accessed by multiple analysis tools including Cloud Dataproc, BigQuery, and Compute Engine.

2.

A streaming data pipeline stores new data daily.

3.

Peformance is not a factor in the solution.

4.

The solution design should maximize availability. How should you design data storage for this solution?

A. Create a Cloud Dataproc cluster with high availability. Store the data in HDFS, and peform analysis as needed.

B. Store the data in BigQuery. Access the data using the BigQuery Connector or Cloud Dataproc and Compute Engine.

C. Store the data in a regional Cloud Storage bucket. Access the bucket directly using Cloud Dataproc, BigQuery, and Compute Engine.

D. Store the data in a multi-regional Cloud Storage bucket. Access the data directly using Cloud Dataproc, BigQuery, and Compute Engine.

Correct Answer: D

QUESTION 2

All Google Cloud Bigtable client requests go through a front-end server ______ they are sent to a Cloud Bigtable node.

A. before

B. after

C. only if

D. once

Correct Answer: A

In a Cloud Bigtable architecture all client requests go through a front-end server before they are sent to a Cloud Bigtable node. The nodes are organized into a Cloud Bigtable cluster, which belongs to a Cloud Bigtable instance, which is a container for the cluster. Each node in the cluster handles a subset of the requests to the cluster. When additional nodes are added to a cluster, you can increase the number of simultaneous requests that the cluster can handle, as well



as the maximum throughput for the entire cluster. Reference: https://cloud.google.com/bigtable/docs/overview

QUESTION 3

You need to create a new transaction table in Cloud Spanner that stores product sales data. You are deciding what to use as a primary key. From a performance perspective, which strategy should you choose?

- A. The current epoch time
- B. A concatenation of the product name and the current epoch time
- C. A random universally unique identifier number (version 4 UUID)
- D. The original order identification number from the sales system, which is a monotonically increasing integer

Correct Answer: C

QUESTION 4

You work for a large ecommerce company. You store your customers order data in Bigtable. You have a garbage collection policy set to delete the data after 30 days and the number of versions is set to 1. When the data analysts run a query to report total customer spending, the analysts sometimes see customer data that is older than 30 days. You need to ensure that the analysts do not see customer data older than 30 days while minimizing cost and overhead. What should you do?

A. Set the expiring values of the column families to 30 days and set the number of versions to 2.

- B. Use a timestamp range filter in the query to fetch the customer\\'s data for a specific range.
- C. Set the expiring values of the column families to 29 days and keep the number of versions to 1.
- D. Schedule a job daily to scan the data in the table and delete data older than 30 days.

Correct Answer: B

By using a timestamp range filter in the query, you can ensure that the analysts only see the customer data that is within the desired time range, regardless of the garbage collection policy1. This option is the most cost-effective and simple way to avoid fetching data that is marked for deletion by garbage collection, as it does not require changing the existing policy or creating additional jobs. You can use the Bigtable client libraries or the cbt CLI to apply a timestamp range filter to your read requests2. Option A is not effective, as it increases the number of versions to 2, which may cause more data to be retained and increase the storage costs. Option C is not reliable, as it reduces the expiring values to 29 days, which may not match the actual data arrival and usage patterns. Option D is not efficient, as it requires scheduling a job daily to scan and delete the data, which may incur additional overhead and complexity. Moreover, none of these options guarantee that the data older than 30 days will be immediately deleted, as garbage collection is an asynchronous process that can take up to a week to remove the data3. References:

- 1: Filters | Cloud Bigtable Documentation | Google Cloud
- 2: Read data | Cloud Bigtable Documentation | Google Cloud
- 3: Garbage collection overview | Cloud Bigtable Documentation | Google Cloud



QUESTION 5

The Development and External teams nave the project viewer Identity and Access Management (1AM) role m a folder named Visualization. You want the Development Team to be able to read data from both Cloud Storage and BigQuery, but the External Team should only be able to read data from BigQuery. What should you do?

| | Folder Visualization |
|---------------------|----------------------------|
| on-premises | Google Cloud Platform |
| Development Team | Project: acme-raw-data |
| | Raw Data Cloud Storage |
| on-premises | Google Cloud Platform |
| External Team | Project: acme-presentation |
| | Presentation BigQuery |
| | |

A. Remove Cloud Storage IAM permissions to the External Team on the acme-raw-data project

B. Create Virtual Private Cloud (VPC) firewall rules on the acme-raw-data protect that deny all Ingress traffic from the External Team CIDR range

C. Create a VPC Service Controls perimeter containing both protects and BigQuery as a restricted API Add the External Team users to the perimeter s Access Level

D. Create a VPC Service Controls perimeter containing both protects and Cloud Storage as a restricted API. Add the Development Team users to the perimeter\\'s Access Level

Correct Answer: C

PROFESSIONAL-DATA-ENGINEER PDF Dumps PROFESSIONAL-DATA-ENGINEER VCE Dumps PROFESSIONAL-DATA-ENGINEER Practice Test