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.pass4itsure.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

PROFESSIONAL-DATA-ENGINEER Study Guide | PROFESSIONAL-DATA-ENGINEER Exam Questions | PROFESSIONAL-DATA-ENGINEER Braindumps

1 / 5

QUESTION 1

You are migrating an application that tracks library books and information about each book, such as author or year published, from an on-premises data warehouse to BigQuery In your current relational database, the author information is kept in a separate table and joined to the book information on a common key Based on Google\\'s recommended practice for schema design, how would you structure the data to ensure optimal speed of queries about the author of each book that has been borrowed?

A. Keep the schema the same, maintain the different tables for the book and each of the attributes, and query as you are doing today

B. Create a table that is wide and includes a column for each attribute, including the author\\'s first name, last name, date of birth, etc

C. Create a table that includes information about the books and authors, but nest the author fields inside the author column

D. Keep the schema the same, create a view that joins all of the tables, and always query the view

Correct Answer: C

QUESTION 2

You are building a report-only data warehouse where the data is streamed into BigQuery via the streaming API Following Google\\'s best practices, you have both a staging and a production table for the data How should you design your data loading to ensure that there is only one master dataset without affecting performance on either the ingestion or reporting pieces?

A. Have a staging table that is an append-only model, and then update the production table every three hours with the changes written to staging

B. Have a staging table that is an append-only model, and then update the production table every ninety minutes with the changes written to staging

C. Have a staging table that moves the staged data over to the production table and deletes the contents of the staging table every three hours

D. Have a staging table that moves the staged data over to the production table and deletes the contents of the staging table every thirty minutes

Correct Answer: D

QUESTION 3

You are designing a data mesh on Google Cloud with multiple distinct data engineering teams building data products. The typical data curation design pattern consists of landing files in Cloud Storage, transforming raw data in Cloud Storage and BigQuery datasets. and storing the final curated data product in BigQuery datasets You need to configure Dataplex to ensure that each team can access only the assets needed to build their data products. You also need to ensure that teams can easily share the curated data product. What should you do?

A. 1 Create a single Dataplex virtual lake and create a single zone to contain landing, raw.and curated data. 2 Provide each data engineering team access to the virtual lake.

B. 1 Create a single Dataplex virtual lake and create a single zone to contain landing, raw.and curated data. 2 Build separate assets for each data product within the zone.

3. Assign permissions to the data engineering teams at the zone level.

C. 1 Create a Dataplex virtual lake for each data product, and create a single zone to contain landing, raw, and curated data.

2. Provide the data engineering teams with full access to the virtual lake assigned to their data product.

D. 1 Create a Dataplex virtual lake for each data product, and create multiple zones for landing, raw. and curated data.

2. Provide the data engineering teams with full access to the virtual lake assigned to their data product.

Correct Answer: D

This option is the best way to configure Dataplex for a data mesh architecture, as it allows each data engineering team to have full ownership and control over their data products, while also enabling easy discovery and sharing of the curated data across the organization12. By creating a Dataplex virtual lake for each data product, you can isolate the data assets and resources for each domain, and avoid conflicts and dependencies between different teams3. By creating multiple zones for landing, raw, and curated data, you can enforce different security and governance policies for each stage of the data curation process, and ensure that only authorized users can access the data assets45. By providing the data engineering teams with full access to the virtual lake assigned to their data product, you can empower them to manage and monitor their data products, and leverage the Dataplex features such as tagging, quality, and lineage. Option A is not suitable, as it creates a single point of failure and a bottleneck for the data mesh, and does not allow for fine-grained access control and governance for different data products2. Option B is also not suitable, as it does not isolate the data assets and resources for each data product, and assigns permissions at the zone level, which may not reflect the different roles and responsibilities of the data engineering teams34. Option C is better than option A and B, but it does not create multiple zones for landing, raw, and curated data, which may compromise the security and quality of the data products5. References:

1: Building a data mesh on Google Cloud using BigQuery and Dataplex | Google Cloud Blog

2: Data Mesh - 7 Effective Practices to Get Started - Confluent

3: Best practices | Dataplex | Google Cloud

4: Secure your lake | Dataplex | Google Cloud

5: Zones | Dataplex | Google Cloud [6]: Managing a Data Mesh with Dataplex ?ROI Training

QUESTION 4

Your company is in a highly regulated industry. One of your requirements is to ensure individual users have access only to the minimum amount of information required to do their jobs. You want to enforce this requirement with Google BigQuery. Which three approaches can you take? (Choose three.)

A. Disable writes to certain tables.

B. Restrict access to tables by role.

C. Ensure that the data is encrypted at all times.

D. Restrict BigQuery API access to approved users.

E. Segregate data across multiple tables or databases.

F. Use Google Stackdriver Audit Logging to determine policy violations.

Correct Answer: BDF

QUESTION 5

You have a petabyte of analytics data and need to design a storage and processing platform for it. You must be able to perform data warehouse-style analytics on the data in Google Cloud and expose the dataset as files for batch analysis tools in other cloud providers. What should you do?

A. Store and process the entire dataset in BigQuery.

B. Store and process the entire dataset in Cloud Bigtable.

C. Store the full dataset in BigQuery, and store a compressed copy of the data in a Cloud Storage bucket.

D. Store the warm data as files in Cloud Storage, and store the active data in BigQuery.Keep this ratio as 80% warm and 20% active.

Correct Answer: C

[PROFESSIONAL-DATA-ENGINEER Study Guide](#)

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

[PROFESSIONAL-DATA-ENGINEER Braindumps](#)