≡ MENU 🔍

⊙ Google Discussions

Exam Professional Data Engineer All Questions

View all questions & answers for the Professional Data Engineer exam

Go to Exam

📄 **EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 306 DISCUSSION**

Actual exam question from Google's Professional Data Engineer

Question #: 306

Topic #: 1

[All Professional Data Engineer Questions]

You are preparing an organization-wide dataset. You need to preprocess customer data stored in a restricted bucket in Cloud Storage. The data will be used to create consumer analyses. You need to comply with data privacy requirements.

What should you do?

A. Use Dataflow and the Cloud Data Loss Prevention API to mask sensitive data. Write the processed data in BigQuery.

B. Use customer-managed encryption keys (CMEK) to directly encrypt the data in Cloud Storage. Use federated queries from BigQuery. Share the encryption key by following the principle of least privilege.

C. Use the Cloud Data Loss Prevention API and Dataflow to detect and remove sensitive fields from the data in Cloud Storage. Write the filtered data in BigQuery.

D. Use Dataflow and Cloud KMS to encrypt sensitive fields and write the encrypted data in BigQuery. Share the encryption key by following the principle of least privilege.

Show Suggested Answer

by 👤 scaenruy at *Jan. 4, 2024, 1:21 p.m.*

Comments

Type your comment...

⊟ 👤 **raaad** `Highly Voted 👍` 10 months ago

`Selected Answer: A`

- Prioritizes Data Privacy: It protects sensitive information by masking it, reducing the risk of exposure in case of unauthorized access or accidental leaks.
- Reduces Data Sensitivity: Masking renders sensitive data unusable for attackers, even if they gain access to it.
- Preserves Data Utility: Masked data can still be used for consumer analyses, as patterns and relationships are often preserved, allowing meaningful insights to be derived.

👍 ↩ 🏳 upvoted 13 times

⊟ 👤 **dungct** 8 months, 3 weeks ago

why not d ?

👍 ↩ 🏳 upvoted 2 times

⊟ 👤 **ML6** 8 months, 2 weeks ago

Data in Cloud Storage is encrypted by default.

👍 ↩ 🏳 upvoted 2 times

⊟ 👤 **desertlotus1211** `Most Recent ⊘` 1 month, 1 week ago

`Selected Answer: C`

If I had to choose...

I choose C or A... A can still leave partial sensitive data available.

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **desertlotus1211** 1 month, 1 week ago

For data privacy, removing data through a DLP (Data Loss Prevention) system is generally considered better than masking, as it permanently eliminates sensitive information, whereas masking only conceals it, potentially leaving traces or vulnerabilities

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **AlizCert** 5 months ago

`Selected Answer: A`

What made me decide on A instead of C was the "The data will be used to create consumer analyses" sentence. Having all the PIIs completely redacted from the records, we were unable to distinguish between the individual customers.

👍 ↩ 🏳 upvoted 3 times

⊟ 👤 **Matt_108** 9 months, 3 weeks ago

`Selected Answer: A`

Option A, agree with raaad explanation

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **scaenruy** 10 months ago

`Selected Answer: A`

A. Use Dataflow and the Cloud Data Loss Prevention API to mask sensitive data. Write the processed data in BigQuery.

👍 ↩ 🏳 upvoted 2 times