

Google Discussions



Exam Professional Machine Learning Engineer All Questions

View all questions & answers for the Professional Machine Learning Engineer exam

Go to Exam

EXAM PROFESSIONAL MACHINE LEARNING ENGINEER TOPIC 1 QUESTION 148 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 148

Topic #: 1

[\[All Professional Machine Learning Engineer Questions\]](#)

You are training an ML model using data stored in BigQuery that contains several values that are considered Personally Identifiable Information (PII). You need to reduce the sensitivity of the dataset before training your model. Every column is critical to your model. How should you proceed?

- A. Using Dataflow, ingest the columns with sensitive data from BigQuery, and then randomize the values in each sensitive column.
- B. Use the Cloud Data Loss Prevention (DLP) API to scan for sensitive data, and use Dataflow with the DLP API to encrypt sensitive values with Format Preserving Encryption.
- C. Use the Cloud Data Loss Prevention (DLP) API to scan for sensitive data, and use Dataflow to replace all sensitive data by using the encryption algorithm AES-256 with a salt.
- D. Before training, use BigQuery to select only the columns that do not contain sensitive data. Create an authorized view of the data so that sensitive values cannot be accessed by unauthorized individuals.

Show Suggested Answer

by [imamapri](#) at Feb. 4, 2023, 5:14 a.m.

Comments

Type your comment...

Submit

🗄️ 👤 **AnnaR** 6 months, 1 week ago

Selected Answer: B

Not A: Randomizing values alters the data in a way that it can significantly degrade the utility of the data for machine learning purposes (does not preserve original distributions, ...)

Not C: Encryption with AES-256 secures the data, but does not preserve the format and would make the data unusable for ML models

Not D: This ignores columns with sensitive data, which is not viable here as every column is critical to the model. + Creating an authorized view does not alter the data itself but restricts access, which does not address the need to reduce data sensitivity for model training

👍 🔄 🚩 upvoted 3 times

🗄️ 👤 **M25** 1 year, 5 months ago

Selected Answer: B

https://cloud.google.com/dlp/docs/transformations-reference#types_of_de-identification_techniques

<https://cloud.google.com/dlp/docs/transformations-reference#crypto>

👍 🔄 🚩 upvoted 1 times

🗄️ 👤 **TNT87** 1 year, 8 months ago

Selected Answer: B

Answer B

👍 🔄 🚩 upvoted 1 times

🗄️ 👤 **TNT87** 1 year, 8 months ago

model. The Cloud Data Loss Prevention (DLP) API can scan for sensitive data in the dataset and can help to encrypt the sensitive data using Format Preserving Encryption. This approach will allow for the preservation of the data distribution and format, enabling the model to maintain its accuracy. Additionally, using Dataflow with the DLP API can help to efficiently process the data at scale.

👍 🔄 🚩 upvoted 1 times

🗄️ 👤 **Scipione_** 1 year, 8 months ago

Selected Answer: B

Format Preserving Encryption uses deidentify configuration in which you can specify the param wrapped_key (the encrypted ('wrapped') AES-256 key to use).

Answer is B according to me.

Ref: <https://cloud.google.com/dlp/docs/samples/dlp-deidentify-fpe>

👍 🔄 🚩 upvoted 3 times

🗄️ 👤 **TNT87** 1 year, 8 months ago

Selected Answer: D

This approach would allow you to keep the critical columns of data while reducing the sensitivity of the dataset by removing the personally identifiable information (PII) before training the model. By creating an authorized view of the data, you can ensure that sensitive values cannot be accessed by unauthorized individuals.

https://cloud.google.com/bigquery/docs/data-governance#data_loss_prevention

👍 🔄 🚩 upvoted 2 times

🗄️ 👤 **TNT87** 1 year, 8 months ago

Actually its B

👍 🔄 🚩 upvoted 1 times

🗄️ 👤 **alelamb** 1 year, 8 months ago

It says "every" column is critical to your model, why would select specific columns?

👍 🔄 🚩 upvoted 2 times

🗄️ 👤 **TNT87** 1 year, 8 months ago

Hence i provided a link, that should answer your flimsy question. it says "BigQuery that contains several values that are considered Personally Identifiable Information (PII)" i dnt know where you are getting it wrong. the "every" means you cant leave out sensitive data to train your model because every column is critical. its not difficult its easy bro....


👍 🔄 🚩 upvoted 1 times

🗄️ 👤 **RaghavAI** 1 year, 9 months ago

Selected Answer: B

<https://cloud.google.com/dlp/docs/samples/dlp-deidentify-fpe>

👍 🔄 🚩 upvoted 1 times

 **imamapri** 1 year, 9 months ago

Selected Answer: C

Vote C. <https://cloud.google.com/dlp/docs/samples/dlp-deidentify-fpe>

   upvoted 1 times

Start Learning for free



Social Media

[Facebook](#) , [Twitter](#)

[YouTube](#) , [Reddit](#)

[Pinterest](#)



We are the biggest and most updated IT certification exam material website.

Using our own resources, we strive to strengthen the IT professionals community for free.



ExamTopics doesn't offer Real Microsoft Exam Questions. ExamTopics doesn't offer Real Amazon Exam Questions. ExamTopics Materials do not contain actual questions and answers from Cisco's Certification Exams.

CFA Institute does not endorse, promote or warrant the accuracy or quality of ExamTopics. CFA® and Chartered Financial Analyst® are registered trademarks owned by CFA Institute.