

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.passapply.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.passapply.com/professional-data-engineer.html 2024 Latest passapply PROFESSIONAL-DATA-ENGINEER PDF and VCE dumps Download

- Instant Download After Purchase
- 100% Money Back Guarantee
- 😳 365 Days Free Update

VCE & PDF

PassApply.com

800,000+ Satisfied Customers





QUESTION 1

You are developing an application on Google Cloud that will automatically generate subject labels for users\\' blog posts. You are under competitive pressure to add this feature quickly, and you have no additional developer resources. No one on your team has experience with machine learning. What should you do?

A. Call the Cloud Natural Language API from your application. Process the generated Entity Analysis as labels.

B. Call the Cloud Natural Language API from your application. Process the generated Sentiment Analysis as labels.

C. Build and train a text classification model using TensorFlow. Deploy the model using Cloud Machine Learning Engine. Call the model from your application and process the results as labels.

D. Build and train a text classification model using TensorFlow. Deploy the model using a Kubernetes Engine cluster. Call the model from your application and process the results as labels.

Correct Answer: B

QUESTION 2

Your company\\'s data platform ingests CSV file dumps of booking and user profile data from upstream sources into Cloud Storage. The data analyst team wants to join these datasets on the email field available in both the datasets to perform analysis. However, personally identifiable information (PII) should not be accessible to the analysts. You need to de-identify the email field in both the datasets before loading them into BigQuery for analysts. What should you do?

A. 1. Create a pipeline to de-identify the email field by using recordTransformations in Cloud Data Loss Prevention (Cloud DLP) with masking as the de-identification transformations type.

2. Load the booking and user profile data into a BigQuery table.

B. 1. Create a pipeline to de-identify the email field by using recordTransformations in Cloud DLP with format-preserving encryption with FFX as the de-identification transformation type.

2. Load the booking and user profile data into a BigQuery table.

C. 1. Load the CSV files from Cloud Storage into a BigQuery table, and enable dynamic data masking.

2.

Create a policy tag with the email mask as the data masking rule.

3.

Assign the policy to the email field in both tables. A

4.

Assign the Identity and Access Management bigquerydatapolicy.maskedReader role for the BigQuery tables to the analysts.

D. 1. Load the CSV files from Cloud Storage into a BigQuery table, and enable dynamic data masking.



2.

Create a policy tag with the default masking value as the data masking rule.

3.

Assign the policy to the email field in both tables.

4.

Assign the Identity and Access Management bigquerydatapolicy.maskedReader role for the BigQuery tables to the analysts

Correct Answer: B

Cloud DLP is a service that helps you discover, classify, and protect your sensitive data. It supports various deidentification techniques, such as masking, redaction, tokenization, and encryption. Format-preserving encryption (FPE) with FFX is a technique that encrypts sensitive data while preserving its original format and length. This allows you to join the encrypted data on the same field without revealing the actual values. FPE with FFX also supports partial encryption, which means you can encrypt only a portion of the data, such as the domain name of an email address. By using Cloud DLP to de-identify the email field with FPE with FFX, you can ensure that the analysts can join the booking and user profile data on the email field without accessing the PII. You can create a pipeline to de-identify the email field by using recordTransformations in Cloud DLP, which allows you to specify the fields and the de-identification transformations to apply to them. You can then load the de-identified data into a BigQuery table for analysis. References: De-identify sensitive data | Cloud Data Loss Prevention Documentation Format-preserving encryption with FFX | Cloud Data Loss Prevention Documentation De-identify and re-identify data with the Cloud DLP API De-identify data in a pipeline

QUESTION 3

Your analytics team wants to build a simple statistical model to determine which customers are most likely to work with your company again, based on a few different metrics. They want to run the model on Apache Spark, using data housed in Google Cloud Storage, and you have recommended using Google Cloud Dataproc to execute this job. Testing has shown that this workload can run in approximately 30 minutes on a 15-node cluster, outputting the results into Google BigQuery. The plan is to run this workload weekly. How should you optimize the cluster for cost?

- A. Migrate the workload to Google Cloud Dataflow
- B. Use pre-emptible virtual machines (VMs) for the cluster
- C. Use a higher-memory node so that the job runs faster
- D. Use SSDs on the worker nodes so that the job can run faster

Correct Answer: A

QUESTION 4

You are designing a real-time system for a ride hailing app that identifies areas with high demand for rides to effectively reroute available drivers to meet the demand. The system ingests data from multiple sources to Pub/Sub. processes the

data, and stores the results for visualization and analysis in real-time dashboards. The data sources include driver



location updates every 5 seconds and app-based booking events from riders. The data processing involves real-time

aggregation of supply and demand data for the last 30 seconds, every 2 seconds, and storing the results in a lowlatency system for visualization.

What should you do?

A. Group the data by using a tumbling window in a Dataflow pipeline, and write the aggregated data to Memorystore

B. Group the data by using a hopping window in a Dataflow pipeline, and write the aggregated data to Memorystore

C. Group the data by using a session window in a Dataflow pipeline, and write the aggregated data to BigQuery.

D. Group the data by using a hopping window in a Dataflow pipeline, and write the aggregated data to BigQuery.

Correct Answer: B

A hopping window is a type of sliding window that advances by a fixed period of time, producing overlapping windows. This is suitable for the scenario where the system needs to aggregate data for the last 30 seconds, every 2 seconds, and provide real-time updates. A Dataflow pipeline can implement the hopping window logic using Apache Beam, and process both streaming and batch data sources. Memorystore is a low-latency, in- memory data store that can serve the aggregated data to the visualization layer. BigQuery is not a good choice for this scenario, as it is not optimized for low-latency queries and frequent updates.

QUESTION 5

You currently have a single on-premises Kafka cluster in a data center in the us-east region that is responsible for ingesting messages from IoT devices globally. Because large parts of globe have poor internet connectivity, messages sometimes batch at the edge, come in all at once, and cause a spike in load on your Kafka cluster. This is becoming difficult to manage and prohibitively expensive. What is the Google-recommended cloud native architecture for this scenario?

A. Edge TPUs as sensor devices for storing and transmitting the messages.

B. Cloud Dataflow connected to the Kafka cluster to scale the processing of incoming messages.

C. An IoT gateway connected to Cloud Pub/Sub, with Cloud Dataflow to read and process the messages from Cloud Pub/Sub.

D. A Kafka cluster virtualized on Compute Engine in us-east with Cloud Load Balancing to connect to the devices around the world.

Correct Answer: C

PROFESSIONAL-DATA-ENGINEER VCE Dumps PROFESSIONAL-DATA-ENGINEER Exam Questions PROFESSIONAL-DATA-ENGINEER Braindumps