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

QUESTION 1

You are designing a basket abandonment system for an ecommerce company. The system will send a message to a user based on these rules:

No interaction by the user on the site for 1 hour Has added more than $30 worth of products to the basket Has not completed a transaction

You use Google Cloud Dataflow to process the data and decide if a message should be sent. How should you design the pipeline?

A. Use a fixed-time window with a duration of 60 minutes.

B. Use a sliding time window with a duration of 60 minutes.

C. Use a session window with a gap time duration of 60 minutes.

D. Use a global window with a time based trigger with a delay of 60 minutes.

Correct Answer: C

QUESTION 2

You orchestrate ETL pipelines by using Cloud Composer One of the tasks in the Apache Airflow directed acyclic graph (DAG) relies on a third-party service. You want to be notified when the task does not succeed. What should you do?

A. Configure a Cloud Monitoring alert on the sla_missed metric associated with the task at risk to trigger a notification.

B. Assign a function with notification logic to the sla_miss_callback parameter for the operator responsible for the task at risk.

C. Assign a function with notification logic to the on_retry_callback parameter for the operator responsible for the task at risk.

D. Assign a function with notification logic to the on_failure_callback parameter for the operator responsible for the task at risk.

Correct Answer: D

By assigning a function with notification logic to the on_failure_callback parameter, you can customize the action that is taken when a task fails in your DAG1. For example, you can send an email, a Slack message, or a PagerDuty alert to notify yourself or your team about the task failure2. This option is more flexible and reliable than configuring a Cloud Monitoring alert on the sla_missed metric, which only triggers when a task misses its scheduled deadline3. The sla_miss_callback parameter is also related to the sla_missed metric, and it is executed when the task instance has not succeeded and the time is past the task\\'s scheduled execution date plus its sla4. The on_retry_callback parameter is executed before a task is retried4. These options are not suitable for notifying when a task does not succeed, as they depend on the task\\'s schedule and retry settings, which may not reflect the actual task completion status. References:

1: Callbacks | Cloud Composer | Google Cloud

2: How to Send an Email on Task Failure in Airflow - Astronomer

3: Monitoring SLA misses | Cloud Composer | Google Cloud

4: BaseOperator | Apache Airflow Documentation

QUESTION 3

You need to migrate a Redis database from an on-premises data center to a Memorystore for Redis instance. You want to follow Google-recommended practices and perform the migration for minimal cost. time, and effort. What should you do?

A. Make a secondary instance of the Redis database on a Compute Engine instance, and then perform a live cutover.

B. Write a shell script to migrate the Redis data, and create a new Memorystore for Redis instance.

C. Create a Dataflow job to road the Redis database from the on-premises data center. and write the data to a Memorystore for Redis instance

D. Make an RDB backup of the Redis database, use the gsutil utility to copy the RDB file into a Cloud Storage bucket, and then import the RDB tile into the Memorystore for Redis instance.

Correct Answer: D

The import and export feature uses the native RDB snapshot feature of Redis to import data into or export data out of a Memorystore for Redis instance. The use of the native RDB format prevents lock-in and makes it very easy to move data within Google Cloud or outside of Google Cloud. Import and export uses Cloud Storage buckets to store RDB files. Reference: https://cloud.google.com/memorystore/docs/redis/import-export-overview

QUESTION 4

Which of the following is NOT one of the three main types of triggers that Dataflow supports?

A. Trigger based on element size in bytes

B. Trigger that is a combination of other triggers

C. Trigger based on element count

D. Trigger based on time

Correct Answer: A

There are three major kinds of triggers that Dataflow supports: 1. Time-based triggers 2. Data-driven triggers. You can set a trigger to emit results from a window when that window has received a certain number of data elements. 3. Composite triggers. These triggers combine multiple time-based or data-driven triggers in some logical way Reference: https://cloud.google.com/dataflow/model/triggers

QUESTION 5

Which of these are examples of a value in a sparse vector? (Select 2 answers.)

A. [0, 5, 0, 0, 0, 0]

B. [0, 0, 0, 1, 0, 0, 1]

C. [0, 1]

D. [1, 0, 0, 0, 0, 0, 0]

Correct Answer: CD

Categorical features in linear models are typically translated into a sparse vector in which each possible value has a corresponding index or id. For example, if there are only three possible eye colors you can represent \\'eye_color\\' as a length

3 vector: \\'brown\\' would become [1, 0, 0], \\'blue\\' would become [0, 1, 0] and \\'green\\' would become [0, 0, 1]. These vectors are called "sparse" because they may be very long, with many zeros, when the set of possible values is very large (such

as all English words). [0, 0, 0, 1, 0, 0, 1] is not a sparse vector because it has two 1s in it. A sparse vector contains only a single 1.

[0, 5, 0, 0, 0, 0] is not a sparse vector because it has a 5 in it. Sparse vectors only contain 0s and 1s.

Reference:

https://www.tensorflow.org/tutorials/linear#feature_columns_and_transformations