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 have developed three data processing jobs. One executes a Cloud Dataflow pipeline that transforms data uploaded to Cloud Storage and writes results to BigQuery. The second ingests data from on-premises servers and uploads it to Cloud Storage. The third is a Cloud Dataflow pipeline that gets information from third-party data providers and uploads the information to Cloud Storage. You need to be able to schedule and monitor the execution of these three workflows and manually execute them when needed. What should you do?

A. Create a Direct Acyclic Graph in Cloud Composer to schedule and monitor the jobs.

B. Use Stackdriver Monitoring and set up an alert with a Webhook notification to trigger the jobs.

C. Develop an App Engine application to schedule and request the status of the jobs using GCP API calls.

D. Set up cron jobs in a Compute Engine instance to schedule and monitor the pipelines using GCP API calls.

Correct Answer: A

QUESTION 2

You have a data stored in BigQuery. The data in the BigQuery dataset must be highly available. You need to define a storage, backup, and recovery strategy of this data that minimizes cost. How should you configure the BigQuery table?

A. Set the BigQuery dataset to be regional. In the event of an emergency, use a point-in-time snapshot to recover the data.

B. Set the BigQuery dataset to be regional. Create a scheduled query to make copies of the data to tables suffixed with the time of the backup. In the event of an emergency, use the backup copy of the table.

C. Set the BigQuery dataset to be multi-regional. In the event of an emergency, use a point-in-time snapshot to recover the data.

D. Set the BigQuery dataset to be multi-regional. Create a scheduled query to make copies of the data to tables suffixed with the time of the backup. In the event of an emergency, use the backup copy of the table.

Correct Answer: C

QUESTION 3

You want to optimize your queries for cost and performance. How should you structure your data?

A. Partition table data by create_date, location_id and device_version

B. Partition table data by create_date cluster table data by location_Id and device_version

C. Cluster table data by create_date location_id and device_version

D. Cluster table data by create_date partition by locationed and device_version

Correct Answer: B

QUESTION 4

Which software libraries are supported by Cloud Machine Learning Engine?

A. Theano and TensorFlow

B. Theano and Torch

C. TensorFlow

D. TensorFlow and Torch

Correct Answer: C

Cloud ML Engine mainly does two things:

1.

Enables you to train machine learning models at scale by running TensorFlow training applications in the cloud.

2.

Hosts those trained models for you in the cloud so that you can use them to get predictions about new data. Reference: https://cloud.google.com/ml-engine/docs/technical-overview#what_it_does

QUESTION 5

The marketing team at your organization provides regular updates of a segment of your customer dataset. The marketing team has given you a CSV with 1 million records that must be updated in BigQuery. When you use the UPDATE statement in BigQuery, you receive a quotaExceeded error. What should you do?

A. Reduce the number of records updated each day to stay within the BigQuery UPDATE DML statement limit.

B. Increase the BigQuery UPDATE DML statement limit in the Quota management section of the Google Cloud Platform Console.

C. Split the source CSV file into smaller CSV files in Cloud Storage to reduce the number of BigQuery UPDATE DML statements per BigQuery job.

D. Import the new records from the CSV file into a new BigQuery table. Create a BigQuery job that merges the new records with the existing records and writes the results to a new BigQuery table.

Correct Answer: D

BigQuery DML statements have no quota limits. https://cloud.google.com/bigquery/quotas#data-manipulation-language-statements

However, DML statements are counted toward the maximum number of table operations per day and partition modifications per day. DML statements will not fail due to these limits.

In addition, DML statements are subject to the maximum rate of table metadata update operations. If you exceed this limit, retry the operation using exponential backoff between retries.