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

QUESTION 1

You work for a mid-sized enterprise that needs to move its operational system transaction data from an on-premises database to GCP. The database is about 20 TB in size. Which database should you choose?

A. Cloud SQL

B. Cloud Bigtable

C. Cloud Spanner

D. Cloud Datastore

Correct Answer: A

QUESTION 2

You work for an economic consulting firm that helps companies identify economic trends as they happen. As part of your analysis, you use Google BigQuery to correlate customer data with the average prices of the 100 most common goods sold, including bread, gasoline, milk, and others. The average prices of these goods are updated every 30 minutes. You want to make sure this data stays up to date so you can combine it with other data in BigQuery as cheaply as possible. What should you do?

A. Load the data every 30 minutes into a new partitioned table in BigQuery.

B. Store and update the data in a regional Google Cloud Storage bucket and create a federated data source in BigQuery

C. Store the data in Google Cloud Datastore. Use Google Cloud Dataflow to query BigQuery and combine the data programmatically with the data stored in Cloud Datastore

D. Store the data in a file in a regional Google Cloud Storage bucket. Use Cloud Dataflow to query BigQuery and combine the data programmatically with the data stored in Google Cloud Storage.

Correct Answer: B

QUESTION 3

Your organization has been collecting and analyzing data in Google BigQuery for 6 months. The majority of the data analyzed is placed in a time-partitioned table named events_partitioned. To reduce the cost of queries, your organization created a view called events, which queries only the last 14 days of data. The view is described in legacy SQL. Next month, existing applications will be connecting to BigQuery to read the events data via an ODBC connection. You need to ensure the applications can connect. Which two actions should you take? (Choose two.)

A. Create a new view over events using standard SQL

B. Create a new partitioned table using a standard SQL query

C. Create a new view over events_partitioned using standard SQL

D. Create a service account for the ODBC connection to use for authentication

E. Create a Google Cloud Identity and Access Management (Cloud IAM) role for the ODBC connection and shared "events"

Correct Answer: CD

A standard SQL query cannot reference a view defined using legacy SQL syntax. In order to connect through ODBC connection, we need to use standard SQL. So, we need to create a new view over events_partitioned table using standard SQL which is C. Need service account to connect through ODBC which is option D. Check the links below. I am not sure about A whether we can create a view over another view which was built using legacy SQL https://cloud.google.com/bigquery/docs/views https://cloud.google.com/community/tutorials/bigquery-from-excel https://www.simba.com/products/BigQuery/doc/ODBC_InstallGuide/mac/content/odbc/bq/configuring/authenticating/serviceaccount.htm

---

QUESTION 4

In order to securely transfer web traffic data from your computer\\'s web browser to the Cloud Dataproc cluster you should use a(n) _____.

A. VPN connection

B. Special browser

C. SSH tunnel

D. FTP connection

Correct Answer: C

To connect to the web interfaces, it is recommended to use an SSH tunnel to create a secure connection to the master node.

Reference: https://cloud.google.com/dataproc/docs/concepts/cluster-web-interfaces#connecting_to_the_web_interfaces

---

QUESTION 5

You are building a model to predict whether or not it will rain on a given day. You have thousands of input features and want to see if you can improve training speed by removing some features while having a minimum effect on model accuracy. What can you do?

A. Eliminate features that are highly correlated to the output labels.

B. Combine highly co-dependent features into one representative feature.

C. Instead of feeding in each feature individually, average their values in batches of 3.

D. Remove the features that have null values for more than 50% of the training records.

Correct Answer: B

PROFESSIONAL-DATA-
ENGINEER PDF Dumps

PROFESSIONAL-DATA-
ENGINEER VCE Dumps

PROFESSIONAL-DATA-
ENGINEER Braindumps