Congratulation!! You have selected the best success partner to become

Exam overview

Level: Professional

Length: 2 hours

Registration fee: $200 (plus tax where applicable)

Exam format: Multiple choice and multiple select taken remotely or in person at a test center. Locate a test center near you.

Exam delivery method:

a. Take the online-proctored exam from a remote location

b. Take the onsite-proctored exam at a testing center

https://cloud.google.com/certification/data-engineer

QUESTION NO: 1

Your company built a TensorFlow neutral-network model with a large number of neurons and layers. The model fits well for the training data. However, when tested against new data, it performs poorly. What method can you employ to address this?

A.
Threading

B.
Serialization

C.
Dropout Methods

D.
Dimensionality Reduction

Answer: C
Reference: https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877

QUESTION NO: 2

You are building a model to make clothing recommendations. You know a user's fashion preference is likely to change over time, so you build a data pipeline to stream new data back to the model as it becomes available. How should you use this data to train the model?

A.
Continuously retrain the model on just the new data.

B.
Continuously retrain the model on a combination of existing data and the new data.

C.
Train on the existing data while using the new data as your test set.

D.
Train on the new data while using the existing data as your test set.

Answer: B

Explanation:

QUESTION NO: 3

You designed a database for patient records as a pilot project to cover a few hundred patients in three clinics. Your design used a single database table to represent all patients and their visits, and you used self-joins to generate reports. The server resource utilization was at 50%. Since then, the scope of the project has expanded. The database must now store 100 times more patient records. You can no longer run the reports, because they either take too long or they encounter errors with insufficient compute resources. How should you adjust the database design?

A.
Add capacity (memory and disk space) to the database server by the order of 200.

B.
Shard the tables into smaller ones based on date ranges, and only generate reports with prespecified date ranges.

C.
Normalize the master patient-record table into the patient table and the visits table, and create other necessary tables to avoid self-join.

D.
Partition the table into smaller tables, with one for each clinic. Run queries against the smaller table pairs, and use unions for consolidated reports.

Answer: C
Explanation:

QUESTION NO: 4

You create an important report for your large team in Google Data Studio 360. The report uses Google BigQuery as its data source. You notice that visualizations are not showing data that is less than 1 hour old. What should you do?

A.
Disable caching by editing the report settings.

B.

Disable caching in BigQuery by editing table details.

C.
Refresh your browser tab showing the visualizations.

D.
Clear your browser history for the past hour then reload the tab showing the virtualizations.

Answer: A
Reference: https://support.google.com/datastudio/answer/7020039?hl=en

QUESTION NO: 5

An external customer provides you with a daily dump of data from their database. The data flows into Google Cloud Storage GCS as comma-separated values (CSV) files. You want to analyze this data in Google BigQuery, but the data could have rows that are formatted incorrectly or corrupted. How should you build this pipeline?

A.
Use federated data sources, and check data in the SQL query.

B.
Enable BigQuery monitoring in Google Stackdriver and create an alert.

C.
Import the data into BigQuery using the gcloud CLI and set max_bad_records to 0.

D.
Run a Google Cloud Dataflow batch pipeline to import the data into BigQuery, and push errors to another dead-letter table for analysis.

Answer: D
Explanation:

QUESTION NO: 6

Your weather app queries a database every 15 minutes to get the current temperature. The frontend is powered by Google App Engine and server millions of users. How should you design the frontend to respond to a database failure?

A.
Issue a command to restart the database servers.

B.
Retry the query with exponential backoff, up to a cap of 15 minutes.

C.
Retry the query every second until it comes back online to minimize staleness of data.

D.
Reduce the query frequency to once every hour until the database comes back online.

Answer: B
Explanation:

QUESTION NO: 7

You are creating a model to predict housing prices. Due to budget constraints, you must run it on a single resource-constrained virtual machine. Which learning algorithm should you use?

A.
Linear regression

B.
Logistic classification

C.
Recurrent neural network

D.
Feedforward neural network

Answer: A
Explanation:

QUESTION NO: 8

You are building new real-time data warehouse for your company and will use Google BigQuery streaming inserts. There is no guarantee that data will only be sent in once but you do have a unique ID for each row of data and an event timestamp. You want to ensure that duplicates are

not included while interactively querying data. Which query type should you use?

A.
Include ORDER BY DESK on timestamp column and LIMIT to 1.

B.
Use GROUP BY on the unique ID column and timestamp column and SUM on the values.

C.
Use the LAG window function with PARTITION by unique ID along with WHERE LAG IS NOT NULL.

D.
Use the ROW_NUMBER window function with PARTITION by unique ID along with WHERE row equals 1.

Answer: D
Explanation:

QUESTION NO: 9

Your company is using WILDCARD tables to query data across multiple tables with similar names. The SQL statement is currently failing with the following error:

```
# Syntax error : Expected end of statement but got "-" at [4:11]
SELECT age
FROM
    bigquery-public-data.noaa_gsod.gsod
WHERE
    age != 99
    AND_TABLE_SUFFIX = '1929'
ORDER BY
    age DESC
```

Which table name will make the SQL statement work correctly?

A.
'bigquery-public-data.noaa_gsod.gsod'

B.

bigquery-public-data.noaa_gsod.gsod*

C.
'bigquery-public-data.noaa_gsod.gsod'*

D.
'bigquery-public-data.noaa_gsod.gsod*`

Answer: D
Reference: https://cloud.google.com/bigquery/docs/wildcard-tables

QUESTION NO: 10

Your company is in a highly regulated industry. One of your requirements is to ensure individual users have access only to the minimum amount of information required to do their jobs. You want to enforce this requirement with Google BigQuery. Which three approaches can you take? (Choose three.)

A.
Disable writes to certain tables.

B.
Restrict access to tables by role.

C.
Ensure that the data is encrypted at all times.

D.
Restrict BigQuery API access to approved users.

E.
Segregate data across multiple tables or databases.

F.
Use Google Stackdriver Audit Logging to determine policy violations.

Answer: B,D,F
Explanation: