



Visit and Download Full Version Certificationtime Exam Dumps
<https://certificationtime.com/updated/professional-data-engineer-exam-dumps->



Google

Exam Questions Professional-Data-Engineer

Cloud Certified Professional Data Engineer

<https://certificationtime.com/>





QUESTION 1

Your company built a TensorFlow neural-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

Correct Answer: D

Explanation/Reference:

Reference <https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-pricepredictionusing-tensorflow-30505541d877>

QUESTION 2

You have Google Cloud Dataflow streaming pipeline running with a Google Cloud Pub/Sub subscription as the source. You need to make an update to the code that will make the new Cloud Dataflow pipeline incompatible with the current version. You do not want to lose any data when making this update. What should you do?

- A. . Update the current pipeline and use the drain flag.
- B. Update the current pipeline and provide the transform mapping JSON object.
- C. Create a new pipeline that has the same Cloud Pub/Sub subscription and cancel the old pipeline.
- D. Create a new pipeline that has a new Cloud Pub/Sub subscription and cancel the old pipeline.

Correct Answer: D

QUESTION 3

What are two of the benefits of using denormalized data structures in BigQuery?

- A. . Reduces the amount of data processed, reduces the amount of storage required
- B. Increases query speed, makes queries simpler
- C. Reduces the amount of storage required, increases query speed
- D. Reduces the amount of data processed, increases query speed

Correct Answer: B

Explanation/Reference:



Denormalization increases query speed for tables with billions of rows because BigQuery's performance degrades when doing JOINS on large tables, but with a denormalized data structure, you don't have to use JOINS, since all of the data has been combined into one table. Denormalization also makes queries simpler because you do not have to use JOIN clauses. Denormalization increases the amount of data processed and the amount of storage required because it creates redundant data.

Reference:

https://cloud.google.com/solutions/bigquery-data-warehouse#denormalizing_data

QUESTION 4

How can you get a neural network to learn about relationships between categories in a categorical feature?

- A. Create a multi-hot column
- B. Create a one-hot column
- C. Create a hash bucket
- D. Create an embedding column

Correct Answer: D

Explanation/Reference:

There are two problems with one-hot encoding. First, it has high dimensionality, meaning that instead of having just one value, like a continuous feature, it has many values, or dimensions. This makes computation more time-consuming, especially if a feature has a very large number of categories. The second problem is that it doesn't encode any relationships between the categories. They are completely independent from each other, so the network has no way of knowing which ones are similar to each other.

Both of these problems can be solved by representing a categorical feature with an embedding column. The idea is that each category has a smaller vector with, let's say, 5 values in it. But unlike a one-hot vector, the values are not usually 0. The values are weights, similar to the weights that are used for basic features in a neural network. The difference is that each category has a set of weights (5 of them in this case).

You can think of each value in the embedding vector as a feature of the category. So, if two categories are very similar to each other, then their embedding vectors should be very similar too.

Reference:

<https://cloudacademy.com/google/introduction-to-google-cloud-machine-learning-enginecourse/awide-and-deep-model.html>



QUESTION 5

When creating a new Cloud Dataproc cluster with the projects.regions.clusters.create operation, these four values are required: project, region, name, and

- A. zone
- B. node
- C. label
- D. type

Correct Answer: A

Explanation/Reference:

At a minimum, you must specify four values when creating a new cluster with the projects.regions.clusters.create operation:

The project in which the cluster will be created

The region to use

The name of the cluster

The zone in which the cluster will be created

You can specify many more details beyond these minimum requirements. For example, you can also specify the number of workers, whether preemptible compute should be used, and the network settings.

Reference:

https://cloud.google.com/dataproc/docs/tutorials/pythonlibraryexample#create_a_new_cloud_dataproc_cluste

Full Access

<https://certificationtime.com/updated/professional-data-engineer-exam-dumps-pdf/>