EXAMTOPICS

- Expert Verified, Online, **Free.**

≡ MENU   🔍

← Google Discussions

Exam Professional Machine Learning Engineer All Questions
View all questions & answers for the Professional Machine Learning Engineer exam

Go to Exam

📄 **EXAM PROFESSIONAL MACHINE LEARNING ENGINEER TOPIC 1 QUESTION 176 DISCUSSI...**

Actual exam question from Google's Professional Machine Learning Engineer
Question #: 176
Topic #: 1

[All Professional Machine Learning Engineer Questions]

You work for a food product company. Your company's historical sales data is stored in BigQuery.You need to use Vertex AI's custom training service to train multiple TensorFlow models that read the data from BigQuery and predict future sales. You plan to implement a data preprocessing algorithm that performs mm-max scaling and bucketing on a large number of features before you start experimenting with the models. You want to minimize preprocessing time, cost, and development effort. How should you configure this workflow?

   A. Write the transformations into Spark that uses the spark-bigquery-connector, and use Dataproc to preprocess the data.

   B. Write SQL queries to transform the data in-place in BigQuery.

   C. Add the transformations as a preprocessing layer in the TensorFlow models.

   D. Create a Dataflow pipeline that uses the BigQueryIO connector to ingest the data, process it, and write it back to BigQuery.

Show Suggested Answer

---

by 👤 **b1a8fae** at *Jan. 10, 2024, 4:43 p.m.*

Comments

Type your comment…

Submit

👤 **cert_pz** 2 weeks ago

Selected Answer: C

Since it is already given that we will be using a TF-Model and do experiments exclusevly there, I don't see why we wouldn't use TF-Layers to preprocess the data. We would minimize costs by not having to store additional data. Time would be around the same as the layer transforms the attribute during training time and development would also be simpler, since if you are using keras it would literally be 2 more lines of code. However I see the Argument for B as well but I would still go with C in this case.

Specifically in this case I would use Normalization layer for normalization and Discretization layer for binning/bucketing.

👍 ↩ ⚑ upvoted 1 times

👤 **fitri001** 6 months, 2 weeks ago

Selected Answer: B

In-place Transformation: BigQuery allows you to perform data transformations directly within the data warehouse using SQL queries. This eliminates the need for data movement and reduces processing time compared to other options that involve data transfer.
Minimized Development Effort: Since you're already familiar with SQL, writing queries for mm-max scaling and bucketing requires minimal additional development effort compared to learning and implementing new frameworks like Spark or Dataflow.
Cost-Effective: BigQuery's serverless architecture scales processing power based on your workload. This can be more cost-effective than managing separate processing clusters like Dataproc.

👍 ↩ ⚑ upvoted 2 times

👤 **shadz10** 9 months, 3 weeks ago

Selected Answer: B

B - Keeps the preprocessing algorithm seperate from the model

👍 ↩ ⚑ upvoted 2 times

👤 **36bdc1e** 9 months, 3 weeks ago

C
This option allows you to leverage the power and simplicity of TensorFlow to preprocess and transform the data with simple Python code

👍 ↩ ⚑ upvoted 2 times

👤 **BlehMaks** 9 months, 3 weeks ago

Selected Answer: B

BigQuery can do both transformations https://cloud.google.com/bigquery/docs/manual-preprocessing#numerical_functions

👍 ↩ ⚑ upvoted 1 times

👤 **b1a8fae** 10 months ago

Selected Answer: B

BigQuery (SQL) is the easiest, cheapest approach

👍 ↩ ⚑ upvoted 1 times

Start Learning for free

Social Media

Facebook , Twitter
YouTube , Reddit
Pinterest

EXAMTOPICS

We are the biggest and most updated IT certification exam material website.
Using our own resources, we strive to strengthen the IT professionals community for free.