

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 263 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 263

Topic #: 1

[\[All Professional Machine Learning Engineer Questions\]](#)

You are developing a custom TensorFlow classification model based on tabular data. Your raw data is stored in BigQuery. contains hundreds of millions of rows, and includes both categorical and numerical features. You need to use a MaxMin scaler on some numerical features, and apply a one-hot encoding to some categorical features such as SKU names. Your model will be trained over multiple epochs. You want to minimize the effort and cost of your solution. What should you do?



- A. 1. Write a SQL query to create a separate lookup table to scale the numerical features.
2. Deploy a TensorFlow-based model from Hugging Face to BigQuery to encode the text features.
3. Feed the resulting BigQuery view into Vertex AI Training.
- B. 1. Use BigQuery to scale the numerical features.
2. Feed the features into Vertex AI Training.
3. Allow TensorFlow to perform the one-hot text encoding.
- C. 1. Use TFX components with Dataflow to encode the text features and scale the numerical features.
2. Export results to Cloud Storage as TFRecords.
3. Feed the data into Vertex AI Training.
- D. 1. Write a SQL query to create a separate lookup table to scale the numerical features.
2. Perform the one-hot text encoding in BigQuery.
3. Feed the resulting BigQuery view into Vertex AI Training.

Show Suggested Answer

Comments

Type your comment...

Submit

  **b2aaace** Highly Voted  5 months, 4 weeks ago

Selected Answer: C

"Full-pass stateful transformations aren't suitable for implementation in BigQuery. If you use BigQuery for full-pass transformations, you need auxiliary tables to store quantities needed by stateful transformations, such as means and variances to scale numerical features. Further, implementation of full-pass transformations using SQL on BigQuery creates increased complexity in the SQL scripts, and creates intricate dependency between training and the scoring SQL scripts." https://www.tensorflow.org/tfx/guide/tft_bestpractices#where_to_do_preprocessing

   upvoted 6 times

  **Prakzz** 3 months, 3 weeks ago

Isn't Dataflow includes a lot of effort as the question asking to minimize the effort here?

   upvoted 1 times

  **YangG** Most Recent  1 week, 2 days ago

Selected Answer: C

multiple epochs --> need to persist data after preprocessing

   upvoted 1 times

  **wences** 1 month ago

Selected Answer: D

Option D since it says minimize effort and cost following that adding something rather than BQ will increase complexity.

   upvoted 1 times

  **AzureDP900** 3 months, 2 weeks ago

Option C uses TFX (TensorFlow Extended) components with Dataflow, which is a great way to perform complex data preprocessing tasks like one-hot encoding and scaling.

This approach allows you to process your data in a scalable and efficient manner, using Cloud Storage as the output location.

By exporting the results as TFRecords, you can easily feed this preprocessed data into Vertex AI Training for model development.

   upvoted 1 times

  **dija123** 3 months, 3 weeks ago

Selected Answer: C

agree with TFX components with Dataflow

   upvoted 1 times

  **bobjr** 4 months, 2 weeks ago

Selected Answer: D

GPT says D, Gemini says B, Perplexity says C....

I say D : stay in one tool, BQ, which is cheap and natively scalable.

B has a risk of out of memory error.

   upvoted 2 times

  **fitri001** 5 months, 4 weeks ago

Selected Answer: B

BigQuery for Preprocessing:

BigQuery is a serverless data warehouse optimized for large datasets. It can handle scaling numerical features using built-in functions like SCALE or QUANTILE_SCALE, reducing the need for complex custom logic or separate lookup tables.

TensorFlow for One-Hot Encoding:

TensorFlow excels at in-memory processing. One-hot encoding of categorical features, especially text features like SKU names, can be efficiently performed within your TensorFlow model during training. This avoids unnecessary data movement or transformations in BigQuery.

Vertex AI Training:

By feeding the preprocessed data (scaled numerical features) directly into Vertex AI Training, you leverage its managed

By feeding the preprocessed data (scaled numerical features) directly into TensorFlow training, you leverage its managed infrastructure for training your custom TensorFlow model.

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **fitri001** 5 months, 4 weeks ago

Option A: Creates unnecessary complexity and data movement. BigQuery is better suited for scaling numerical features, and TensorFlow is efficient for one-hot encoding.

Option C: TFX is a powerful framework for complex pipelines, but for a simpler scenario like this, it might be an overkill. Additionally, exporting data as TFRecords adds an extra step, potentially increasing cost and complexity.

Option D: One-hot encoding in BigQuery might be cumbersome for textual features like SKU names.

pen_spark

exclamation It can be computationally expensive and result in data explosion. TensorFlow handles this efficiently within the model.

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **cruise93** 6 months ago

Selected Answer: C

Agree with b1a8fae

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **gscharly** 6 months ago

Selected Answer: C

agree with daidai75

👍 ↩ 🚩 upvoted 2 times

🗄️ 👤 **pinimichele01** 5 months, 4 weeks ago

Option B is not suitable for the big volume of data processing?????

BQ is not suitable for big volume??..

for me is B

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **guilhermebutzke** 8 months ago

Selected Answer: B

My Answer: B

1. Use BigQuery to scale the numerical features.: Simpler and cheaper then use TFX components with Dataflow to scale the numerical features

2. Feed the features into Vertex AI Training.

3. Allow TensorFlow to perform the one-hot text encoding: TensorFlow handles the one-hot text encoding better than BQ.

👍 ↩ 🚩 upvoted 3 times

🗄️ 👤 **daidai75** 9 months ago

Selected Answer: C

key messages: "contains hundreds of millions of rows, and includes both categorical and numerical features. You need to use a MaxMin scaler on some numerical features, and apply a one-hot encoding to some categorical features such as SKU names".

Option B is not suitable for the big volume of data processing. Option C is better.

👍 ↩ 🚩 upvoted 2 times

🗄️ 👤 **b1a8fae** 9 months ago

Selected Answer: C

Inclined to choose C over B. By using TFX components with Dataflow, you can perform feature engineering on large-scale tabular data in a distributed and efficient way. You can use the Transform component to apply the MaxMin scaler and the one-hot encoding to the numerical and categorical features, respectively. You can also use the ExampleGen component to read data from BigQuery and the Trainer component to train your TensorFlow model.

👍 ↩ 🚩 upvoted 2 times

🗄️ 👤 **pikachu007** 9 months, 1 week ago

Selected Answer: B

Option A: Involves creating a separate lookup table and deploying a Hugging Face model in BigQuery, increasing complexity and cost.

Option C: While TFX offers robust preprocessing capabilities, it adds overhead for this use case and requires knowledge of Dataflow.

Option D: Performing one-hot encoding in BigQuery can be less efficient than TensorFlow's optimized implementation.

👍 ↩ 🚩 upvoted 3 times

Start Learning for free



Social Media

[Facebook](#) , [Twitter](#)

[YouTube](#) , [Reddit](#)

[Pinterest](#)



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.



© 2024 ExamTopics

ExamTopics doesn't offer Real Microsoft Exam Questions. ExamTopics doesn't offer Real Amazon Exam Questions. ExamTopics Materials do not contain actual questions and answers from Cisco's Certification Exams.

CFA Institute does not endorse, promote or warrant the accuracy or quality of ExamTopics. CFA® and Chartered Financial Analyst® are registered trademarks owned by CFA Institute.