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 132 DISCUSSI...**

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 132

Topic #: 1

[All Professional Machine Learning Engineer Questions]

You work on a data science team at a bank and are creating an ML model to predict loan default risk. You have collected and cleaned hundreds of millions of records worth of training data in a BigQuery table, and you now want to develop and compare multiple models on this data using TensorFlow and Vertex AI. You want to minimize any bottlenecks during the data ingestion state while considering scalability. What should you do?

A. Use the BigQuery client library to load data into a dataframe, and use tf.data.Dataset.from_tensor_slices() to read it.

B. Export data to CSV files in Cloud Storage, and use tf.data.TextLineDataset() to read them.

C. Convert the data into TFRecords, and use tf.data.TFRecordDataset() to read them.

D. Use TensorFlow I/O's BigQuery Reader to directly read the data.

Show Suggested Answer

---

by 👤 mil_spyro at *Dec. 13, 2022, 7:43 p.m.*

Comments

Type your comment…

Submit

👤 **hiromi** `Highly Voted 👍` 1 year, 10 months ago

Selected Answer: D

D
- https://www.tensorflow.org/io/api_docs/python/tfio/bigquery

👍 ↩ 🚩 upvoted 6 times

👤 **mil_spyro** `Highly Voted 👍` 1 year, 10 months ago

Selected Answer: D

Vote on D. This will allow to directly access the data from BigQuery without having to first load it into a dataframe or export it to files in Cloud Storage.

👍 ↩ 🚩 upvoted 5 times

👤 **fitri001** `Most Recent ⊘` 6 months, 2 weeks ago

Selected Answer: D

Direct Data Access: TensorFlow I/O's BigQuery Reader allows you to directly access data from BigQuery tables within your TensorFlow script.expand_more This eliminates the need for intermediate data movement (e.g., to CSV files) and data manipulation steps (e.g., loading into DataFrames).exclamation
Scalability: BigQuery Reader is designed to handle large datasets efficiently. It leverages BigQuery's parallel processing capabilities to stream data into your TensorFlow training pipeline, minimizing processing bottlenecks and enabling scalability as your data volume grows.

👍 ↩ 🚩 upvoted 2 times

> 👤 **fitri001** 6 months, 2 weeks ago
>
> . BigQuery Client Library and Dataframe: While the BigQuery client library can access BigQuery data, loading it into a DataFrame and using tf.data.Dataset.from_tensor_slices() is inefficient for massive datasets due to memory limitations and potential processing bottlenecks.
> B. CSV Files and TextLineDataset: Exporting data to CSV and using tf.data.TextLineDataset() introduces unnecessary data movement and processing overhead, hindering both efficiency and scalability.
> C. TFRecords: TFRecords can be efficient for certain use cases, but converting hundreds of millions of records into TFRecords can be time-consuming and resource-intensive.
>
> pen_spark
> exclamation Additionally, reading them might require parsing logic within your TensorFlow script.
>
> 👍 ↩ 🚩 upvoted 1 times

👤 **guilhermebutzke** 9 months, 1 week ago

Selected Answer: D

D
https://cloud.google.com/blog/products/ai-machine-learning/tensorflow-enterprise-makes-accessing-data-on-google-cloud-faster-and-easier

👍 ↩ 🚩 upvoted 1 times

👤 **julliet** 1 year, 5 months ago

Selected Answer: D

D
BigQuery is more compact way to store the data than TFRecords

👍 ↩ 🚩 upvoted 2 times

👤 **M25** 1 year, 6 months ago

Selected Answer: D

Went with D

👍 ↩ 🚩 upvoted 1 times

👤 **TNT87** 1 year, 8 months ago

Selected Answer: D

D. Use TensorFlow I/O's BigQuery Reader to directly read the data.

The reason for this choice is that using TensorFlow I/O's BigQuery Reader is the most efficient and scalable option for reading data directly from BigQuery into TensorFlow models. It allows for distributed processing and avoids unnecessary data duplication, which can cause bottlenecks and consume large amounts of storage. Additionally, the BigQuery Reader is optimized for reading data in parallel from BigQuery tables and streaming them directly into TensorFlow. This eliminates the need for any intermediate file formats or data copies, reducing latency and increasing performance.

👍 ↩ 🚩 upvoted 2 times

Start Learning for free

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.