



- 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 249 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 249

Topic #: 1

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

You are developing an ML model to identify your company's products in images. You have access to over one million images in a Cloud Storage bucket. You plan to experiment with different TensorFlow models by using Vertex AI Training. You need to read images at scale during training while minimizing data I/O bottlenecks. What should you do?

- A. Load the images directly into the Vertex AI compute nodes by using Cloud Storage FUSE. Read the images by using the `tf.data.Dataset.from_tensor_slices` function
- B. Create a Vertex AI managed dataset from your image data. Access the `AIP_TRAINING_DATA_URI` environment variable to read the images by using the `tf.data.Dataset.list_files` function.
- C. Convert the images to TFRecords and store them in a Cloud Storage bucket. Read the TFRecords by using the `tf.data.TFRecordDataset` function.
- D. Store the URLs of the images in a CSV file. Read the file by using the `tf.data.experimental.CsvDataset` function.

Show Suggested Answer

by [pikachu007](#) at Jan. 13, 2024, 9:23 a.m.

Comments

Type your comment...

🗄️ 👤 **tavva_prudhvi** 5 months, 3 weeks ago

Selected Answer: C

TFRRecords is a binary storage format optimized for TensorFlow. By storing images as TFRRecords, you can improve the I/O efficiency as the data is serialized and can be efficiently loaded off-disk in a batched manner. TFRecordDataset is specifically designed for reading these files efficiently, which helps in minimizing I/O bottlenecks. This approach is typically recommended for large-scale image datasets as it ensures data is read efficiently in a manner suitable for distributed training.

👍 ↩ 🚩 upvoted 2 times

🗄️ 👤 **gscharly** 6 months, 2 weeks ago

Selected Answer: C

agree with pikachu007

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **fitri001** 6 months, 3 weeks ago

Selected Answer: A

Read the images by using the `tf.data.Dataset.from_tensor_slices` function.

Here's why this option is most efficient:

Cloud Storage FUSE: This mounts your Cloud Storage bucket directly to the training VM, allowing on-demand access to image data as local files. It minimizes network overhead and data transfer compared to downloading the entire dataset beforehand.

`tf.data.Dataset.from_tensor_slices`: This function is suitable for reading data directly from memory. Since Cloud Storage FUSE presents the images as local files, you can leverage this function for efficient data access within your training script.

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **fitri001** 6 months, 3 weeks ago

B. Vertex AI Managed Dataset: While managed datasets offer convenience, accessing them might involve additional network overhead compared to Cloud Storage FUSE.

C. TFRRecords: Converting images to TFRRecords can be an additional processing step, potentially introducing I/O overhead. While TFRecord format might be efficient for some models, it's not strictly necessary for minimizing I/O during data access.

D. CSV with Image URLs: Reading image URLs from a CSV and fetching each image individually creates significant network traffic, leading to I/O bottlenecks. It's less efficient than directly accessing the images through Cloud Storage FUSE.

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **fitri001** 6 months, 3 weeks ago

TensorFlow Datasets (TFDs): Consider implementing TFDs within your training script. They offer functionalities like parallelized data loading and on-the-fly data augmentation to further optimize training efficiency.

Preprocessing and Caching: Preprocess data (resizing, normalization) within your TFD pipeline or training script.

Cache preprocessed data locally on the VM to avoid redundant processing during training iterations.

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **felipecin** 8 months, 2 weeks ago

Selected Answer: C

The TFRecord format is a simple format for storing a sequence of binary records.

Protocol buffers are a cross-platform, cross-language library for efficient serialization of structured data.

👍 ↩ 🚩 upvoted 2 times

🗄️ 👤 **pikachu007** 9 months, 3 weeks ago

Selected Answer: C

Option A: Cloud Storage FUSE can be slower for large datasets and adds complexity.

Option B: Vertex AI managed datasets offer convenience but might not match TFRecord performance for large-scale image training.

Option D: CSV files require manual loading and parsing, increasing overhead.

👍 ↩ 🚩 upvoted 4 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.